Algebraic branching programs, border complexity, and tangent spaces

Markus Bläser
Department of Computer Science, Saarland University
Saarland Informatics Campus, Saarbrücken, Germany

Christian Ikenmeyer
University of Liverpool, United Kingdom

Meena Mahajan
The Institute of Mathematical Sciences, HBNI, Chennai, India

Anurag Pandey
Max Planck Institute for Informatics
Saarland Informatics Campus, Saarbrücken, Germany

Nitin Saurabh
Technion-IIT, Haifa, Israel

—— **Abstract** ——

Nisan showed in 1991 that the width of a smallest noncommutative single-(source,sink) algebraic branching program (ABP) to compute a noncommutative polynomial is given by the ranks of specific matrices. This means that the set of noncommutative polynomials with ABP width complexity at most k is Zariski-closed, an important property in geometric complexity theory. It follows that approximations cannot help to reduce the required ABP width.

It was mentioned by Forbes that this result would probably break when going from single-(source,sink) ABPs to trace ABPs. We prove that this is correct. Moreover, we study the commutative monotone setting and prove a result similar to Nisan, but concerning the analytic closure. We observe the same behavior here: The set of polynomials with ABP width complexity at most k is closed for single-(source,sink) ABPs and not closed for trace ABPs. The proofs reveal an intriguing connection between tangent spaces and the vector space of flows on the ABP. We close with additional observations on VQP and the closure of VNP which allows us to establish a separation between the two classes.

1   Introduction and Results

Algebraic branching programs (ABPs) are an elegant model of computation that is widely studied in algebraic complexity theory (see e.g. [4, 40, 30, 31, 1, 3, 25, 27, 15, 17]) and is a focus of study in geometric complexity theory [28, 18, 19]. An ABP is a layered directed graph with $d+1$ layers of vertices (edges only go from layers i to $i+1$) such that the first and last layer have exactly the same number of vertices. Each vertex in the first layer has exactly one so-called *corresponding* vertex in the last layer. One interesting classical case is when the first and last layer have exactly one vertex, which is usually studied in theoretical computer science. We call this the *single-(source,sink) model*. Among algebraic geometers working on ABPs it is common to not impose restrictions on the number of vertices in the first and last layer [28, 18, 29]. We call this the *trace model*. Every edge in an ABP is labeled by a homogeneous linear form. If we denote by $\ell(e)$ the homogeneous linear form of edge e, then we say that the ABP computes $\sum_p \prod_{e \in p} \ell(e)$, where the sum is over all paths that start in the first layer and end in the last layer at the vertex corresponding to the start vertex.

The *width* of an ABP is the number of vertices in its largest layer. We denote by $\mathsf{w}(f)$ the minimal width required to compute f in the trace model and we call $\mathsf{w}(f)$ the *trace ABP width complexity* of f. We denote by $\mathsf{w}_1(f)$ the minimal width required to compute f in the single-(source,sink) model and we call $\mathsf{w}_1(f)$ the *single-(source,sink) ABP width complexity* of f.

The complexity class VBP is defined as the set of sequences of polynomials (f_m) for which the sequence $\mathsf{w}(f_m)$ is polynomially bounded. Let $\mathrm{per}_m := \sum_{\pi \in \mathfrak{S}_m} \prod_{i=1}^{m} x_{i,\pi(i)}$ be the permanent polynomial. Valiant's famous VBP $\neq$ VNP conjecture can concisely be stated as "The sequence of natural numbers $\big(\mathsf{w}(\mathrm{per}_m)\big)_m$ is not polynomially bounded." Alternatively, this is phrased with w_1 or other polynomially related complexity measures in a completely analogous way. In geometric complexity theory (GCT), one searches for lower bounds on algebraic complexity measures over $\mathbb{C}$ such as w and w_1 for explicit polynomials such as the permanent. All lower bounds methods in GCT and most lower bounds methods in algebraic complexity theory are *continuous*, which means that if f_ε is a curve of polynomials with $\lim_{\varepsilon \to 0} f_\varepsilon = f$ (coefficient-wise limit) and $\mathsf{w}(f_\varepsilon) \leq w$, then these methods cannot be used to prove $\mathsf{w}(f) > w$. This is usually phrased in terms of so-called *border complexity* (see e.g. [14, 28]): The *border trace ABP width complexity* $\underline{\mathsf{w}}(f)$ is the smallest w such that f can be approximated arbitrarily closely by polynomials f_ε with $\mathsf{w}(f_\varepsilon) \leq w$. Analogously, we define the *border single-(source,sink) ABP width complexity* $\underline{\mathsf{w}_1}(f)$ as the smallest w such that f can be approximated arbitrarily closely by polynomials f_ε with $\mathsf{w}_1(f_\varepsilon) \leq w$. We define $\overline{\text{VBP}}$ as the set of sequences of polynomials such that $(\underline{\mathsf{w}}(f_m))$ is polynomially bounded. Clearly VBP $\subseteq \overline{\text{VBP}}$. Mulmuley and Sohoni [32, 33, 14] (see also [12, 10] for a related conjecture) conjectured a strengthening of Valiant's conjecture, namely that VNP $\not\subseteq \overline{\text{VBP}}$. In principle it could be that $\underline{\mathsf{w}}(f) < \mathsf{w}(f)$; the gap could even be superpolynomial, which would mean that VBP $\subsetneq \overline{\text{VBP}}$. If VBP $= \overline{\text{VBP}}$, then Valiant's conjecture is the same as the Mulmuley-Sohoni conjecture, which would mean that if VBP $\neq$ VNP, then continuous lower bounds methods exist that show this separation.

Border complexity is an old area of study in algebraic geometry. In theoretical computer science it was introduced by Bini et al. [6], where [5] proves that in the study of fast matrix multiplication, the gap between complexity and border complexity is not too large. The study of the gap between complexity and border complexity of algebraic complexity measures in general has started recently [21, 9, 26] as an approach to understand if strong algebraic complexity lower bounds can be obtained from continuous methods.

In this paper we study two very different settings of ABPs: The noncommutative and the monotone setting. To capture commutative, noncommutative, and monotone computation, let R be a graded semiring with homogeneous components R_d. In our case the settings for R_d are

- $R_d = \mathbb{F}[x_1, \ldots, x_m]_d$ the set of homogeneous degree d polynomials in m variables over a field $\mathbb{F}$,
- $R_d = \mathbb{F}\langle x_1, \ldots, x_m \rangle_d$ the set of homogeneous degree d polynomials in m noncommuting variables over a field $\mathbb{F}$,
- $R_d = \mathbb{R}_+[x_1, \ldots, x_m]_d$ the set of homogeneous degree d polynomials in m variables with nonnegative coefficients.

As it is common in the theoretical computer science literature, we call elements of R_d *polynomials*. Note that $\mathbb{F}\langle x_1, \ldots, x_m \rangle_d$ is naturally isomorphic to the d-th tensor power of $\mathbb{F}^m$, so *tensor* would be the better name. We hope that no confusion arises when in the later sections (where we use concepts from multilinear algebra) we use the tensor language. In the *homogeneous setting*, all ABP edge labels are in R_1, and hence the polynomial that is computed is in R_d. In the *affine setting*, all ABP edge labels are in $R_0 + R_1$, and hence the polynomial that is computed is in $\bigoplus_{d' \leq d} R_{d'}$.

Noncommutative ABPs

Let $R_d = \mathbb{F}\langle x_1, \ldots, x_m \rangle_d$ and consider the homogeneous setting. We write ncw instead of w and ncw_1 instead of w_1 to highlight that we are in the noncommutative setting. Nisan [35] proved:

▶ **Theorem 1.** *Let M_i denote the $m^i \times m^{d-i}$ matrix whose entry at position $((k_1, \ldots, k_i), (k_{i+1}, \ldots, k_d))$ is the coefficient of the monomial $x_{k_1} x_{k_2} \cdots x_{k_d}$ in f. Then every single-(source,sink) ABP computing f has at least $\mathsf{rk}(M_i)$ many vertices in layer i. Conversely, there exists a single-(source,sink) ABP computing f with exactly $\mathsf{rk}(M_i)$ many vertices in layer i.*

Nisan used this formulation to prove strong complexity lower bounds for the noncommutative determinant and permanent. Forbes [16] remarked that Theorem 1 implies that for fixed w

$$\text{the set } \{f \mid \mathsf{ncw}_1(f) \leq w\} \text{ is Zariski-closed}^1 \tag{1}$$

and hence that

$$\underline{\mathsf{ncw}_1}(f) = \mathsf{ncw}_1(f) \text{ for all } f. \tag{2}$$

Proving a similar result (even up to polynomial blowups) in the commutative setting would be spectacular: It would imply $\mathrm{VBP} = \overline{\mathrm{VBP}}$ and hence that Valiant's conjecture is the same as the Mulmuley-Sohoni conjecture. By a general principle, for all standard algebraic complexity measures, over $\mathbb{C}$ we have that the Zariski-closure of a set of polynomials of complexity at most w equals the Euclidean closure [34, §2.C].

---

[1] We identify each m-variate homogeneous degree d polynomial with its coefficient vector. There is a standard topology on the vector space of these coefficient vectors that we call the Euclidean topology. The Zariski-closure of a set X of vectors is the smallest set of vectors that contains X and that is the common zero set of a finite set of polynomials in the coordinate variables, see e.g. [7, Ch. 4] for the commutative case.

Forbes mentioned that he believes that Nisan's proof cannot be lifted to the trace model. In this paper we prove that Forbes is correct, by constructing a polynomial f_0 with

$$\underline{\mathsf{ncw}}(f_0) < \mathsf{ncw}(f_0). \tag{3}$$

The proof is given in Sections 5–8. It is a surprisingly subtle application of differential geometry (inspired by [24]) and interprets tangent spaces to certain varieties as vector spaces of flows on an ABP digraph.

The gap between $\underline{\mathsf{ncw}}(f)$ and $\mathsf{ncw}(f)$ can never be very large though:

$$\underline{\mathsf{ncw}}(f) \le \mathsf{ncw}(f) \le \mathsf{ncw}_1(f) \overset{(2)}{=} \underline{\mathsf{ncw}_1}(f) \le \left(\underline{\mathsf{ncw}}(f)\right)^2 \text{ for all } f. \tag{4}$$

It is worth noting that for our separating polynomial f_0, the gap is even less; $\underline{\mathsf{ncw}}(f_0) < \mathsf{ncw}(f_0) \le 2\underline{\mathsf{ncw}}(f_0)$. This is the first algebraic model of computation where complexity and border complexity differ, but their gap is known to be polynomially bounded! For most models of computation almost nothing is known about the gap between complexity and border complexity. For commutative width 2 affine ABPs the gap is even as large as between computable and non-computable [9]!

Monotone ABPs

Let $R_d = \mathbb{R}_+[x_1, \ldots, x_m]_d$ and consider the affine or homogeneous setting.

Since $\mathbb{R}$ is not algebraically closed, we switch to a more algebraic definition of approximation. Let $\mathbb{R}[\varepsilon, \varepsilon^{-1}]_+$ denote the ring of Laurent polynomials that are nonnegative for all sufficiently small $\varepsilon > 0$. Clearly, elements from $\mathbb{R}[\varepsilon, \varepsilon^{-1}]_+$ can have a pole at $\varepsilon = 0$ of arbitrarily high order. We define $\underline{\mathsf{mw}}(f)$ to be the smallest w such that there exists a polynomial f' over the ring $\mathbb{R}[\varepsilon, \varepsilon^{-1}]_+$ such that

- there exists a width w ABP over $\mathbb{R}[\varepsilon, \varepsilon^{-1}]_+$ that computes f',
- no coefficient in f' contains an ε with negative exponent, and setting ε to 0 in f' yields f, i.e., $f'_{\varepsilon=0} = f$.

We prove a result that is comparable to (2), but uses a very different proof technique:

$$\underline{\mathsf{mw}_1}(f) = \mathsf{mw}_1(f) \text{ for all } f. \tag{5}$$

In terms of complexity classes, this implies

$$\mathrm{MVBP} = \overline{\mathrm{MVBP}}^{\mathbb{R}}.$$

Our proof also works if the ABP is not layered and the labels are affine.

Intuitively, in this monotone setting, one would think that approximations do not help, because there cannot be cancellations. But quite surprisingly the same construction as in (3) can be used to find f_0 such that

$$\underline{\mathsf{mw}}(f_0) < \mathsf{mw}(f_0). \tag{6}$$

---

[2] Given a trace ABP Γ computing f and a pair of corresponding start and end vertices, we can extract a single-(source,sink) ABP by deleting all other start and end vertices. If we do this for each pair of start and end vertices, and if we then idenfity all start vertices to a single start vertex, and also all end vertex to a single end vertex, then we obtain a single-(source,sink) ABP computing f. The width has grown by a factor of w, where w is the number of start vertices in Γ.

By the same reasoning as in (4), we obtain

$$\underline{\mathsf{mw}}(f) \leq \mathsf{mw}(f) \leq \big(\underline{\mathsf{mw}}(f)\big)^2 \text{ for all } f. \tag{7}$$

This gives a natural monotone model of computation where approximations speed up the computation. Again, the gap is polynomially bounded!

Separating VQP from $\overline{\text{VNP}}$

Bürgisser in his monograph [11] defined the complexity class VQP as the class of polynomials with quasi-polynomially bounded straight-line programs, and established its relation to the classes VP and VNP (see Section 9 for definitions). He showed that the determinant polynomial is VQP-complete with respect to the so-called *qp*-projections (see [11], Corollary 2.29). He strengthened Valiant's hypothesis of VNP $\not\subseteq$ VP to VNP $\not\subseteq$ VQP and called it *Valiant's extended hypothesis* (see [11], section 2.5). He further showed that VP is strictly contained in VQP as one would intuitively expect (see [11], section 8.2). Finally, he also showed that VQP is not contained in VNP (see [11], Proposition 8.5 and Corollary 8.9). In this article, we observe that his proof is stronger and actually shows that VQP is not contained in $\overline{\text{VNP}}$ either, where $\overline{\text{VNP}}$ is the closure of the complexity class VNP in the sense as mentioned above.

Structure of the paper

In Section 4 we prove (5). Sections 5 to 8 are dedicated to proving (3) and (6) via a new connection between tangent spaces and flow vector spaces. In Section 9, we discuss the separation between VQP and $\overline{\text{VNP}}$.

2   Related work

Grenet [20] showed that $\mathsf{mw}(\text{per}_m) \leq \binom{m}{\lceil m/2 \rceil}$ by an explicit construction of a monotone single-(source,sink) ABP. Even though the construction is monotone, its size is optimal for $m = 3$ [2] (for 4 this is already unknown). The noncommutative version of this setting has been studied in [17]. [42] recently showed that the monotone circuit classes MVP and MVNP are different. We refer the reader to [42] and [38] and the references therein to get more information about monotone algebraic models of computation and their long history.

Hüttenhain and Lairez [24] present a method that can be used to show that a complexity measure and its border variant are not the same. They used it to prove that an explicit polynomial has border determinantal complexity 3, but higher determinantal complexity. We use their ideas as a starting point in Section 5 and the later sections.

3   Preliminaries

For a homogeneous degree d ABP Γ, we denote by V the set of vertices of Γ and by V^i the set of vertices in layer i, $1 \leq i \leq d+1$. We choose an explicit bijection between the sets V^1 and V^{d+1}, so that each vertex v in V^1 has exactly one *corresponding* vertex $\mathsf{corr}(v)$ in V^{d+1}. We denote by E^i the set of edges from V^i to V^{i+1}. Let E denote the union of all E^i.

There is a classical interpretation in terms of iterated matrix multiplication: Fix some arbitrary ordering of the vertices within each layer, such that the i-th vertex in V^1 corresponds to the i-th vertex in V^{d+1}. For $1 \leq k \leq d$ let M_k be the $|V^k| \times |V^{k+1}|$ matrix whose entry

at position (i, j) in M_k is the label from the i-th vertex in V^k to the j-th vertex in V^{k+1}. Then Γ computes the trace

$$\sum_{\substack{1 \leq k_1 \leq |V^1| \\ 1 \leq k_2 \leq |V^2| \\ \vdots \\ 1 \leq k_d \leq |V^d|}} (M_1)_{k_1, k_2} (M_2)_{k_2, k_3} \cdots (M_{d-1})_{k_{d-1}, k_d} (M_d)_{k_d, k_1} = \mathsf{tr}\big(M_1 M_2 \cdots M_d\big). \tag{8}$$

Hence the name *trace model*. In the single-(source,sink) model, the trace is taken of a 1×1 matrix.

4 Monotone commutative single-(source,sink) ABPs are closed

For fixed $w \in \mathbb{N}$ we study

the set $\{f \in \mathbb{R}_+[x_1, \ldots, x_n]_d \mid \mathsf{mw}_1(f) \leq w\}$. $\tag{9}$

We first start with the simple observation that it is *not* Zariski-closed.

▶ **Proposition 2.** $\{f \in \mathbb{R}_+[x_1, \ldots, x_n]_d \mid \mathsf{mw}_1(f) \leq w\}$ *is not Zariski-closed.*

Proof. Note that a homogeneous degree d single-(source,sink) width w ABP has $2w + w^2(d-2)$ many edges. The label on each edge is a linear form in n variables, so such an ABP is determined by $N := n(2w + w^2(d - 2))$ many parameters. Let $F : \mathbb{C}^N \to \mathbb{C}[x_1, \ldots, x_n]_d$ be the map that maps these parameters to the polynomial computed by the ABP. Every coordinate function of F is given by polynomials in N variables, so F is Zariski-continuous. Therefore

$$\overline{F((\mathbb{R}_+)^N)} = \overline{F(\overline{(\mathbb{R}_+)^N})} = \overline{F(\mathbb{C}^N)} \supseteq F(\mathbb{C}^N) \supsetneq F((\mathbb{R}_+)^N),$$

where the overline means the Zariski-closure. We remark that we did not use any special feature of the model of computation other than the fact that it is defined over $\mathbb{R}$.   ◀

Recall that an ABP has $d + 1$ layers of vertices. If an ABP has w_i many vertices in layer i, $1 \leq i \leq d$, we say the ABP has *format* $w = (w_1, w_2, \ldots, w_d)$. We further recall that $w_{d+1} = w_1$. The following theorem is our closure result, which proves (5) and hence $\mathrm{MVBP} = \overline{\mathrm{MVBP}}^{\mathbb{R}}$.

▶ **Theorem 3.** *Given a polynomial f over $\mathbb{R}$ and given a format w single-(source,sink) ABP with affine linear labels over $\mathbb{R}[\varepsilon, \varepsilon^{-1}]_+$ computing f_ε such that $\lim_{\varepsilon \to 0} f_\varepsilon = f$. Then there exists a format w monotone single-(source,sink) ABP that computes f.*

Proof. The proof is constructive and done by a two-step process. In the first step (which is fairly standard and works in many computational models) we move all the ε with negative exponents to edges adjacent to the source. The second step then uses the monotonicity.

Given Γ with affine linear labels over $\mathbb{R}[\varepsilon, \varepsilon^{-1}]_+$ we repeat the following process until all labels that contain an ε with a negative exponent are incident to the source vertex.

▪ Let e be an edge whose label contains ε with a negative exponent $-i < 0$. Moreover, assume that e is not incident to the source vertex. Let v be the start vertex of e. We rescale all edges outgoing of v with ε^i and we rescale all edges incoming to v with ε^{-i}.

If we always choose the edge with the highest layer, then it is easy to see that this process terminates. Since every path from the source to the sink that goes through a vertex v must use exactly one edge that goes into v and exactly one edge that comes out of v, throughout the process the value of Γ does not change. We finish this first phase by taking the highest negative power i among all labels of edges that are incident to the source and then rescale all these edges with ε^i. The resulting ABP Γ^i computes $\varepsilon^i f_\epsilon$ and no label contains an ε with negative exponent. We now start phase 2 that transforms Γ^i into Γ^{i-1} that computes $\varepsilon^{i-1} f_\epsilon$ without introducing negative exponents of ε. We repeat phase 2 until we reach Γ^0 in which we safely set ε to 0. Throughout the whole process we do not change the structure of the ABP and only rescale edge labels with powers of ε, which preserves monotonicity, so the proof is finished. It remains to show how Γ^i can be transformed into Γ^{i-1}. An edge whose label is divisible by ε is called an ε-*edge*. Consider the set Δ of vertices that are reachable from the source using only non ε-edges in Γ^i. The crucial insight is that since Γ^i is monotone and computes a polynomial that is divisible by ε, we know that every path in Γ^i from the source to the sink uses an ε-edge. Therefore Δ cannot contain the sink. We call a vertex in Δ whose outdegree is zero a *leaf* vertex. We repeat the following procedure until the source is the only leaf vertex:

- Let v be a non-source leaf vertex in Δ. We rescale all edges outgoing of v with ε^{-1} and we rescale all edges incoming to v with ε.

It is easy to see that this process terminates with the source being the only leaf vertex. Since the source is a leaf vertex, all edges incident to the source are ε-edges. We divide all their labels by ε to obtain Γ^{i-1}.   ◀

5 Explicit construction of f_0 with higher complexity than border complexity

Fix some $d \geq 3$. In this section for every $m \geq 2$ we construct f_0 such that

$$m = \underline{\mathsf{ncw}}(f_0) < \mathsf{ncw}(f_0). \tag{10}$$

A completely analogous construction can be used to find f_0 with $\underline{\mathsf{w}}(f_0) < \mathsf{w}(f_0)$ and with $\underline{\mathsf{mw}}(f_0) < \mathsf{mw}(f_0)$. For the sake of simplicity, we carry out only the proof for (10).

Recall that in a format w ABP we have $w_{d+1} = w_1$. In each layer i we enumerate the vertices $V^i = \{v_1^i, \ldots, v_{w_i}^i\}$ and we assume without loss of generality that the correspondence bijection between V^{d+1} and V^1 is the identity on the indices j of v_j^1, i.e., the jth vertex in V^1 corresponds to the jth vertex in V^{d+1}.

Fix an ABP format $w = (w_1, w_2, \ldots, w_d)$ such that for all i, $w_i \geq 2$. Let Γ_{com} denote the directed acyclic graph underlying an ABP of format w. An edge can be described by the triple (a, b, i), where $1 \leq i \leq d$, $1 \leq a \leq w_i$ and $1 \leq b \leq w_{i+1}$. Consider the following labeling of the edges with triple-indexed variables: $\ell_{\mathrm{com}}((a, b, i)) = x_{(a,b)}^{(i)}$. Define f_{com} to be the polynomial computed by Γ_{com} with edge labels ℓ_{com}.

We now construct f_0 as follows. Let d be odd (the case when d is even works analogously). Since in each layer we enumerated the vertices, we can now assign to each vertex its parity: even or odd. We call an edge between two even or two odd vertices *parity preserving*, while we call the other edges *parity changing*. Let us consider the following labeling of Γ_{com}: We set $\ell_0((a, b, i)) := x_{(a,b)}^{(i)}$ if (a, b, i) is parity changing (i.e., $a \not\equiv b \pmod 2$) and set the label $\ell_0((a, b, i)) := \varepsilon x_{(a,b)}^{(i)}$ otherwise, where $\varepsilon \in \mathbb{C}$. Let f_ε' be the polynomial computed by Γ_{com} with edge labels ℓ_0 and set $f_\varepsilon := \frac{1}{\varepsilon} f_\varepsilon'$ for $\varepsilon \neq 0$. We define $f_0 := \lim_{\varepsilon \to 0} f_\varepsilon$ (convergence follows from the construction, because d is odd). By definition, for all $\varepsilon \neq 0$, f_ε can be

computed by a format w ABP. However, we will now prove that this property fails for the limit point f_0.

▶ **Theorem 4.** *Fix an ABP format $w = (w_1, w_2, \ldots, w_d)$ such that for all i, $w_i \geq 2$. Let f_0 be defined as above. Then, f_0 cannot be computed by an ABP of format w.*

Note that for a format where $m = w_1 = \cdots = w_d$, this gives the f_0 which was desired in (10). (Note, however, that f_0 can be computed by an ABP of width $2m$ as follows. Construct an ABP Γ' that has, for each vertex $v \in \Gamma_{\mathrm{com}}$, vertices v' and v''. For each parity changing edge $(a, b) \in \Gamma_{\mathrm{com}}$ with label ℓ_0, add edges (a', b') and (a'', b'') with the same label ℓ_0. For each parity preserving edge $(a, b) \in \Gamma_{\mathrm{com}}$ with label ℓ_0, add edge (a', b'') with label $(\frac{1}{\varepsilon})\ell_0$. For corresponding vertices u, v in Γ_{com}, let v'' be the corresponding vertex for u' and v' be the corresponding vertex for u'' in Γ'. All paths between corresponding vertices in this ABP use exactly one parity preserving edge of Γ_{com}, and so this ABP computes f_0.)

The proof of Theorem 4 works as follows. Let $\mathsf{G} := \mathsf{GL}_{w_1 w_2} \times \mathsf{GL}_{w_2 w_3} \times \cdots \times \mathsf{GL}_{w_d w_{d+1}}$. Let $\mathsf{End} := \overline{G}$ denote its Euclidean closure, i.e., tuples of matrices in which one or several matrices can be singular.

We consider noncommutative homogeneous polynomials in the variables $x_{(a,b)}^{(i)}$ such that the i-th variable in each monomial is $x_{(a,b)}^{(i)}$ for some $a \in [w_i]$ and $b \in [w_{i+1}]$. The vector space of these polynomials is isomorphic to $W := \mathbb{C}^{w_1 w_2} \otimes \mathbb{C}^{w_2 w_3} \otimes \cdots \otimes \mathbb{C}^{w_d w_{d+1}}$ and the monoid End (and thus also the group G) acts on this space in the canonical way. The set

$$\{f \in W \mid f \text{ can be computed by a format } w \text{ ABP}\}$$

is precisely the orbit $\mathsf{End} f_{\mathrm{com}}$. We follow the overall proof strategy in [24]. The monoid orbit $\mathsf{End} f_{\mathrm{com}}$ decomposes into two disjoint orbits:

$$\mathsf{End} f_{\mathrm{com}} = \mathsf{G} f_{\mathrm{com}} \cup (\mathsf{End} \setminus \mathsf{G}) f_{\mathrm{com}}.$$

Our goal is to show two things independently:

1. $f_0 \notin (\mathsf{End} \setminus \mathsf{G}) f_{\mathrm{com}}$, and
2. $f_0 \notin \mathsf{G} f_{\mathrm{com}}$,

which finishes the proof of Theorem 4.

All elements in $(\mathsf{End} \setminus \mathsf{G}) f_{\mathrm{com}}$ are *not concise*, a term that we define in Section 6, where we also prove that f_0 is concise. Therefore $f_0 \notin (\mathsf{End} \setminus \mathsf{G}) f_{\mathrm{com}}$.

All elements in $\mathsf{G} f_{\mathrm{com}}$ have *full orbit dimension*, a term that we define in Section 7 and we prove that f_0 does *not* have full orbit dimension in Section 8. This finishes the proof of Theorem 4.

6   Conciseness

In this section we show that $f_0 \notin (\mathsf{End} \setminus \mathsf{G}) f_{\mathrm{com}}$. To do so we use a notion called *conciseness*. Informally, it captures whether a polynomial depends on all variables independent of a change of basis, or a tensor cannot be embedded into a tensor product of smaller spaces.

Given a tensor f in $\mathbb{C}^{m_1} \otimes \mathbb{C}^{m_2} \otimes \cdots \otimes \mathbb{C}^{m_d}$, we associate the following matrices with f. For $j \in [d]$, define a matrix M_f^j of dimension $m_j \times (\prod_{i \in [d] \setminus \{j\}} m_i)$ with rows labeled by the standard basis of $\mathbb{C}^{m_j}$, and columns by elements in the Cartesian product $\{$standard basis of $\mathbb{C}^{m_1}\} \times \cdots \times \{$standard basis of $\mathbb{C}^{m_{j-1}}\} \times \{$standard basis of $\mathbb{C}^{m_{j+1}}\} \times$

$\cdots \times \{$standard basis of $\mathbb{C}^{m_d}\}$. We write the tensor f in the standard basis

$$f = \sum_{\substack{1 \le i_1 \le m_1 \\ 1 \le i_2 \le m_2 \\ \vdots \\ 1 \le i_d \le m_d}} \alpha_{i_1,\ldots,i_d} e_{i_1} \otimes \cdots \otimes e_{i_d}$$

and associate to it the matrix M_f^j whose entry at position $((i_j), (i_1, i_2, \ldots, i_{j-1}, i_{j+1}, \ldots, i_d))$ is $\alpha_{i_1,\ldots,i_d}$.

▶ **Definition 5.** *We say that a tensor f in $\mathbb{C}^{m_1} \otimes \mathbb{C}^{m_2} \otimes \cdots \otimes \mathbb{C}^{m_d}$ is* concise *if and only if for all $j \in [d]$, M_f^j has full rank.* [3]

As a warm-up exercise we now show that f_{com} is concise.

▶ **Proposition 6.** f_{com} *is concise.*

Proof. We know that $f_{\text{com}} \in W$. Let us consider the matrix $M_{f_{\text{com}}}^j$ for some $j \in [d]$. To establish that $M_{f_{\text{com}}}^j$ has full rank, it suffices to show that rows are linearly independent. In order to show that, we argue that every row is non-zero and every column has at most one non-zero entry. In other words, rows are supported on disjoint sets of columns.

A row of $M_{f_{\text{com}}}^j$ is labeled by an edge in the j-th layer of the ABP Γ_{com}. Recall that only paths that start at a vertex in V^1 and end at the corresponding vertex in V^{d+1} contribute to the computation in Γ_{com}. We call such paths *valid paths*. An entry in $M_{f_{\text{com}}}^j$ is non-zero iff the corresponding row and column labels form a valid path in Γ_{com}. Thus, it is easily seen that a row is non-zero iff there is a valid path in Γ_{com} that *passes through* the edge given by the row label. By the structure of Γ_{com}, in particular that every layer is a complete bipartite graph, we observe that passing through every edge there is some valid path. Hence, we obtain that every row is non-zero.

The second claim now follows from the observation that fixing $d-1$ edges either defines a unique dth edge so that these d edges form a valid path, or for these $d-1$ edges there is no such dth edge. ◀

As mentioned in Section 5, to establish $f_0 \notin (\mathsf{End} \setminus \mathsf{G})f_{\text{com}}$ we will show that f_0 is concise while any element in $(\mathsf{End} \setminus \mathsf{G})f_{\text{com}}$ is not.

▶ **Lemma 7.** f_0 *is concise.*

Proof. Analogous to the proof of Proposition 6, we again show that every row of $M_{f_0}^j$ is non-zero and every column of it has at most one non-zero entry. That is, rows of $M_{f_0}^j$ are supported on disjoint sets of columns.

From the construction of f_0 it is seen that a path in Γ_{com} contributes to the computation of f_0 iff it is a valid path that comprises of *exactly one* parity preserving edge. The second claim of every column having at most one non-zero entry now follows for the same reason as in the proof of Proposition 6.

---

[3] When f is viewed as a set-multilinear polynomial (see [36, Section 1.4]), this condition translates to the linear independence of the partial derivatives of f. In particular, M_f^j is testing if the partial derivatives of f with respect to the j-th block of variables are all linearly independent. This partial derivatives based criterion for testing if a polynomial depends on all the variables, independent of a change of basis, is pretty standard: see, for instance, [23, Corollary 5.1.4].

Before proving the first claim, we recall two assumptions in the construction of f_0. The first is that the format $w = (w_1, w_2, \ldots, w_d)$ is such that $w_i \geq 2$ for all $i \in [d]$ and the second is that d is odd. To argue that a row is non-zero it suffices to show that a valid path comprising of only one parity preserving edge passes through the edge given by the row level. Let us consider an arbitrary edge e in Γ_{com}. We have two cases to consider depending on whether it is parity *preserving* or *changing*.

Case 1. Suppose e is parity preserving and it belongs to a layer $j \in [d]$. The number of layers on the left of e is $j - 1$ and on the right is $d - j$. Since d is odd, these numbers are either both even or both odd. We now argue for the case when they are even (the odd case is analogous). Choose a vertex v in V^1 that has the same parity (different in the odd case) as one of the end points of e. (Such a choice exists because $w_1 \geq 2$.) We now claim that there exists a valid path starting at v that passes through e and contains exactly one parity preserving edge. Since e is parity preserving, all edges in the claimed path must be parity changing. We observe that e can be easily extended in both directions using parity changing edges such that the path ends at $\mathrm{corr}(v)$. The existence of parity changing edges at each layer uses the assumption that $w_i \geq 2$.

Case 2. Otherwise $e = (a, b)$ is parity changing. Again as before there are two cases based on whether both $j - 1$ and $d - j$ are even or odd. Consider the case when they are even (the odd case being analogous). We first assume that $j \neq d$. Choose a vertex v in V^1 that has the same parity as a. We now construct a valid path from v to $\mathrm{corr}(v)$ that passes through e and contains exactly one parity preserving edge. It is easily seen that there exists a path from v to a using only parity changing edges. We choose a parity preserving outgoing edge incident to b. We call its endpoint v_1. Since v_1 and v have different parities, we can connect v_1 to $\mathrm{corr}(v)$ in V^{d+1} using only parity changing edges. Thus we obtain the following valid path $v \rightarrow \cdots \rightarrow a \rightarrow b \rightarrow v_1 \rightarrow \cdots \rightarrow \mathrm{corr}(v)$ passing through exactly one parity preserving edge (b, v_1). In the case that $j = d$, choose an incoming parity preserving edge incident on a instead of an outgoing edge on b. ◄

▶ **Remark 8.** We note that if the format $w = (w_1, \ldots, w_d)$ defining f_0 is such that for some $j \in [d]$, $w_j = 1$, then f_0 is *not* concise. This can be seen as follows.

Let $w_j = 1$, and let v denote the unique vertex in V^j. Let e be the edge $e = (1, 1, j)$. If $j < d$, let e' be the edge $e' = (1, 1, j + 1)$, otherwise let e' be the edge $e' = (1, 1, j - 1)$. Both e, e' are parity preserving edges. By construction, every valid path using e' must also use e. Hence the corresponding row in the matrix $M_{f_0}^{j+1}$ if $j < d$, and in $M_{f_0}^{j-1}$ otherwise, is zero. Therefore f_0 is not concise.

This is an interesting observation, because this is the point where our proof fails for single-(source,sink) ABPs, and this is expected, because Nisan [35] had shown that the set of polynomials computed by such ABPs of format w is a closed set.

▶ **Lemma 9.** *Let $f \in (\mathsf{End} \setminus \mathsf{G})f_{\mathrm{com}}$. Then f is not concise.*

Proof. This statement is true in very high generality. In our specific case a proof goes as follows. If $f \in (\mathsf{End} \setminus \mathsf{G})f_{\mathrm{com}}$, then $f = g f_{\mathrm{com}}$ for some $g \in \mathsf{End} \setminus \mathsf{G}$. Let $g = (g_1, \ldots, g_d)$, where $g_i \in \mathbb{C}^{w_i w_{i+1} \times w_i w_{i+1}}$. Since $g \notin \mathsf{G}$, at least one of the g_i must be singular. The crucial property is $M_{g f_{\mathrm{com}}}^i = g_i M_{f_{\mathrm{com}}}^i$, which finishes the proof. ◄

7   Orbit dimension, tangent spaces, and flows

In this section we introduce tangent spaces and study their dimensions. We especially study them in the context of $\mathsf{G}f_{\mathrm{com}}$, and $\mathsf{G}f_0$.

The *orbit dimension* of a tensor $f \in \mathbb{C}^{w_1 w_2} \otimes \mathbb{C}^{w_2 w_3} \otimes \cdots \otimes \mathbb{C}^{w_d w_{d+1}}$ is the dimension of the orbit $\mathsf{G}f$ as an affine variety. It can be determined as the dimension of the tangent space T_f of the action of G at f, which is a vector space defined as follows. Let $\mathfrak{g} := \mathbb{C}^{w_1 w_2 \times w_1 w_2} \times \cdots \times \mathbb{C}^{w_d w_{d+1} \times w_d w_{d+1}}$. For $A \in \mathfrak{g}$ we define the *Lie algebra action* $Af := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left((\mathrm{id} + \varepsilon A)f - f \right)$, where $\mathrm{id} \in \mathsf{G}$ is the identity element. We define the vector space

$$T_f := \mathfrak{g}f = \{ Af \mid A \in \mathfrak{g} \}.$$

$\triangleright$ **Claim 10.** The dimension $\dim T_h$ is the same for all $h \in \mathsf{G}f$.

Proof. Since the action of G is linear, for all $g \in \mathsf{G}$ and $A \in \mathfrak{g}$ we have

$$
\begin{aligned}
A(gf) &= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \left((\mathrm{id} + \varepsilon A)(gf) - gf \right) = \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \left(gg^{-1}(\mathrm{id} + \varepsilon A)gf - gf \right) \\
&= g \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \left((\mathrm{id} + \varepsilon(g^{-1}Ag))f - f \right) = g((g^{-1}Ag)f)
\end{aligned}
$$

Since $A \mapsto g^{-1}Ag$ is a bijection on $\mathfrak{g}$, it follows that $T_{gf} = gT_f$. Hence the claim follows. $\blacktriangleleft$

In the following we will use Claim 10 to argue $f_0 \notin \mathsf{G}f_{\mathrm{com}}$ by showing that $\dim T_{f_{\mathrm{com}}}$ and $\dim T_{f_0}$ are different.

Let $e, e' \in E^i$ and let $A_{e,e'}^{(i)} \in \mathfrak{g}$ denote the matrix tuple where the i-th matrix has a 1 at position (e, e') and all other entries (also in all other matrices) are 0. Since these matrices form a basis of $\mathfrak{g}$, it follows that

$$\mathfrak{g}f = \mathrm{linspan}\{A_{e,e'}^{(i)}f\}.$$

For a tensor f we define the *support* of f as the set of monomials (i.e., standard basis tensors) for which f has nonzero coefficient. For a linear subspace $V \subseteq \mathbb{C}^{w_1 w_2} \otimes \mathbb{C}^{w_2 w_3} \otimes \cdots \otimes \mathbb{C}^{w_d w_{d+1}}$ we define the *support* of V as the union of the supports of all $f \in V$.

We write $e \cap e' = \emptyset$ to indicate that two edges e and e' do not share any vertex. We write $|e \cap e'| = 1$ if they share exactly one vertex. We observe that for $f \in \{f_{\mathrm{com}}, f_0\}$ the vector space T_f decomposes into a direct sum of three vector spaces,

$$
\begin{aligned}
\mathfrak{g}_2 &:= \mathrm{linspan}\{A_{e,e'}^{(i)} \mid 1 \le i \le d, 1 \le e, e' \le w_i w_{i+1}, e \cap e' = \emptyset\} \\
\mathfrak{g}_1 &:= \mathrm{linspan}\{A_{e,e'}^{(i)} \mid 1 \le i \le d, 1 \le e, e' \le w_i w_{i+1}, |e \cap e'| = 1\} \\
\mathfrak{g}_0 &:= \mathrm{linspan}\{A_{e,e}^{(i)} \mid 1 \le i \le d, 1 \le e \le w_i w_{i+1}\}. \\
\mathfrak{g} &= \mathfrak{g}_0 \oplus \mathfrak{g}_1 \oplus \mathfrak{g}_2 \\
T_f &= \mathfrak{g}_0 f \oplus \mathfrak{g}_1 f \oplus \mathfrak{g}_2 f
\end{aligned}
$$

The last direct sum decomposition follows from the fact that $\mathfrak{g}_0 f$, $\mathfrak{g}_1 f$, and $\mathfrak{g}_2 f$ have pairwise disjoint supports.

We show in this section that $\dim \mathfrak{g}_2 f_{\mathrm{com}} = \dim \mathfrak{g}_2 f_0$, and that $\dim \mathfrak{g}_1 f_{\mathrm{com}} = \dim \mathfrak{g}_1 f_0$. In Section 8 we show that $\dim \mathfrak{g}_0 f_{\mathrm{com}} > \dim \mathfrak{g}_0 f_0$, which then implies $f_0 \notin \mathsf{G}f_{\mathrm{com}}$ by Claim 10. In fact, Theorem 13 gives the exact dimension of $\mathfrak{g}_0 f_{\mathrm{com}}$ by proving that $\mathfrak{g}_0 f_{\mathrm{com}}$ is isomorphic to the vector space of flows on the ABP digraph when identifying vertices in V^1 with their corresponding vertices in V^{d+1}. Theorem 14 establishes an additional equation based on the vertex parities that shows that $\mathfrak{g}_0 f_0$ is strictly lower dimensional than $\mathfrak{g}_0 f_{\mathrm{com}}$.

We start with Lemma 11, which shows that $\dim \mathfrak{g}_2 f_{\mathrm{com}}$ and $\dim \mathfrak{g}_2 f_0$ have *full* dimension.

▶ **Lemma 11.** *Let $f \in \{f_{\mathrm{com}}, f_0\}$. The space $\mathfrak{g}_2 f$ has full dimension. That is, its dimension equals $\sum_{i=1}^{d} w_i w_{i+1} (w_i - 1)(w_{i+1} - 1)$.*

Proof. Suppose $f = f_{\mathrm{com}}$. The other case being analogous, we only argue this case.

We analyze the monomials that appear in the different $A_{e,e'}^{(i)} f_{\mathrm{com}}$ and argue that a monomial that appears in some $A_{e,e'}^{(i)} f_{\mathrm{com}}$ can only appear in that specific $A_{e,e'}^{(i)} f_{\mathrm{com}}$. Indeed, each monomial corresponds to a valid path in which one edge e in layer i is changed to e'. Since e and e' share no vertex, from this edge sequence we can reconstruct i, e, and e' uniquely: e' is the edge that does not have any vertex in common with the rest of the edge sequence, i is its layer, and e is the unique edge that we can replace e' by in order to form a valid path. We conclude that the $A_{e,e'}^{(i)} f_{\mathrm{com}}$ have disjoint support and the lemma follows.   ◀

To establish that $\dim \mathfrak{g}_1 f_{\mathrm{com}} = \dim \mathfrak{g}_1 f_0$, we introduce some notation.

For a connected directed graph $G = (V, E)$ we define a *flow* to be a labeling of the edge set E by complex numbers such that at every vertex the sum of the labels of the incoming edges equals the sum of the labels of the outgoing edges. It is easily seen that the set of flows forms a vector space F. We have

$$\dim F = |E| - |V| + 1, \tag{11}$$

see e.g. Theorem 20.7 in [8].

Recall that E^i denotes the set of edges from V^i to V^{i+1}. Let $\mathscr{X} := E^1 \times \cdots \times E^d$ denote the direct product of the sets of edge lists. Each directed path of length d from layer 1 to $d+1$ is an element of $\mathscr{X}$, but $\mathscr{X}$ contains other edge sets as well. Define $E_i := \mathbb{C}^{E^i}$. Consider the following map φ from $\mathscr{X}$ to $E_1 \otimes \cdots \otimes E_d$,

$$\varphi(e_1, \ldots, e_d) = x_{e_1} \otimes \cdots \otimes x_{e_d} \in E_1 \otimes \cdots \otimes E_d$$

where (x_j) is the standard basis of E_i. Note φ is a bijection between $\mathscr{X}$ and the standard basis of $E_1 \otimes \cdots \otimes E_d$.

An edge set in $\mathscr{X}$ is called a *valid path* if it forms a path that starts and ends at corresponding vertices (see Sec. 1). Let $\mathscr{P} \subseteq \mathscr{X}$ denote the set of valid paths.

▶ **Proposition 12.** $\dim \mathfrak{g}_1 f_{\mathrm{com}} = \dim \mathfrak{g}_1 f_0 = \sum_{i=1}^{d} (w_{i-1} + w_{i+1} - 1)(w_i - 1)w_i$, *where* $w_0 := w_d$.

Proof. The proof works almost analogously for f_{com} and f_0, so we treat only the more natural case f_{com}. We show that $\mathfrak{g}_1 f_{\mathrm{com}}$ is isomorphic to a direct sum of vector spaces of flows on very simple digraphs. Fix $1 \le i \le d$. Fix distinct $1 \le a, b \le w_i$. For distinct edges $e, e' \in E^i$, let $\mathscr{P}_{e,e'} \subseteq \mathscr{X}$ be the set of edge sets containing e' that are not valid paths, but that become valid paths by removing e' and adding e. Let $\mathscr{P}_{a,b}^i \subseteq \mathscr{X}$ be the set of edge sets that are not valid paths, but that become valid paths by switching the end point of the $(i-1)$-th edge to v_b^i and that also become valid paths by switching the start point of the i-th edge to v_a^i (if $i - 1 = 0$, then interpret $i - 1 := d$). Pictorially, this means that elements in $\mathscr{P}_{a,b}^i$ are almost valid paths, but there is a discontinuity at layer i, where the path jumps from vertex v_a^i to vertex v_b^i. We have

$$A_{e,e'}^{(i)} f_{\mathrm{com}} = \sum_{p \in \mathscr{P}_{e,e'}} \varphi(p).$$

The vectors $\{A^{(i)}_{e,e'} f_{\mathrm{com}} \mid 1 \leq i \leq d, e, e' \in E^i, |e \cap e'| = 1\}$ are not linearly independent, because for $a \neq b$ we have

$$\sum_{\substack{e \text{ and } e' \text{ have the same start point} \\ e' \text{ ends at the } a\text{-th vertex} \\ e \text{ ends at the } b\text{-th vertex}}} A^{(i-1)}_{e,e'} f_{\mathrm{com}} = \sum_{p \in \mathscr{P}^i_{a,b}} \varphi(p) = \sum_{\substack{h \text{ and } h' \text{ have the same end point} \\ h \text{ starts at the } a\text{-th vertex} \\ h' \text{ starts at the } b\text{-th vertex}}} A^{(i)}_{h,h'} f_{\mathrm{com}}.$$

$$(12)$$

Define

$$T_{a,b,i} \;\; := \;\; \mathrm{linspan}\left\{ A^{(i-1)}_{e,e'} f_{\mathrm{com}} \;\middle|\; \begin{array}{c} e \text{ and } e' \text{ have the same start point} \\ e' \text{ ends at the } a\text{-th vertex} \\ e \text{ ends at the } b\text{-th vertex} \end{array} \right\}$$
$$+ \;\; \mathrm{linspan}\left\{ A^{(i)}_{h,h'} f_{\mathrm{com}} \;\middle|\; \begin{array}{c} h \text{ and } h' \text{ have the same end point} \\ h \text{ starts at the } a\text{-th vertex} \\ h' \text{ starts at the } b\text{-th vertex} \end{array} \right\}.$$

The support of $T_{a,b,i}$ and $T_{\tilde{a},\tilde{b},\tilde{i}}$ are disjoint, provided $(a,b,i) \neq (\tilde{a}, \tilde{b}, \tilde{i})$. Hence

$$\mathfrak{g}_1 f_{\mathrm{com}} = \bigoplus_{\substack{1 \leq i \leq d \\ 1 \leq a,b \leq w_i \\ a \neq b}} T_{a,b,i}$$

It remains to prove that the dimension of $T_{a,b,i}$ is $w_{i-1} + w_{i+1} - 1$, because then

$$\dim \mathfrak{g}_1 f_{\mathrm{com}} = \sum_{\substack{1 \leq i \leq d \\ 1 \leq a,b \leq w_i \\ a \neq b}} (w_{i-1} + w_{i+1} - 1) = \sum_{i=1}^{d} (w_{i-1} + w_{i+1} - 1)(w_i - 1) w_i.$$

Note that $T_{a,b,i}$ is defined as the linear span of $w_{i-1} + w_{i+1}$ many vectors, but (12) shows that these are not linearly independent. We prove that (12) is the only equality by showing that $T_{a,b,i}$ is isomorphic to a flow vector space. We define a multigraph with two vertices: $\odot$ and $\circledast$. We have w_{i+1} many edges from $\odot$ to $\circledast$, and we have w_{i-1} many edges from $\circledast$ to $\odot$. We denote by $\circledast \xrightarrow{k} \odot$ the k-th edge from $\circledast$ to $\odot$. Let $F_{a,b,i}$ denote the vector space of flows on this graph. Its dimension is $w_{i-1} + w_{i+1} - 1$, see (11). We define $\varrho : E^1 \otimes \cdots \otimes E^d \to F_{a,b,i}$ on rank 1 tensors via

$$\varrho(x_{e_1} \otimes \cdots \otimes x_{e_d})(\circledast \xrightarrow{k} \odot) \;\; = \;\; \begin{cases} 1 & \text{if } e_{i-1} \text{ starts at } k \text{ in layer } i-1 \text{ and ends at } a \text{ in layer } i, \\ 0 & \text{otherwise.} \end{cases}$$

$$\varrho(x_{e_1} \otimes \cdots \otimes x_{e_d})(\odot \xrightarrow{l} \circledast) \;\; = \;\; \begin{cases} 1 & \text{if } e_i \text{ starts at } b \text{ in layer } i \text{ and ends at } l \text{ in layer } i+1, \\ 0 & \text{otherwise.} \end{cases}$$

Using (12) it is readily verified that ϱ maps $T_{a,b,i}$ to $F_{a,b,i}$. It remains to show that $\varrho : T_{a,b,i} \to F_{a,b,i}$ is surjective. Let $\alpha := |\mathscr{P}^i_{a,b}|$. We observe that

$$\varrho(A^{(i-1)}_{e,e'} f_{\mathrm{com}})(\circledast \xrightarrow{k} \odot) \;\; = \;\; \begin{cases} \alpha/w_{i-1} & \text{if } e \text{ and } e' \text{ both start at the } k\text{-th vertex} \\ 0 & \text{if } e \text{ and } e' \text{ both start at the same vertex, but not at the } k\text{-th} \end{cases}$$

$$\varrho(A^{(i-1)}_{e,e'} f_{\mathrm{com}})(\odot \xrightarrow{l} \circledast) \;\; = \;\; \alpha/(w_{i-1} w_{i+1})$$

$$\varrho(A^{(i)}_{h,h'} f_{\mathrm{com}})(\odot \xrightarrow{l} \circledast) \;\; = \;\; \begin{cases} \alpha/w_{i+1} & \text{if } h \text{ and } h' \text{ both end at the } l\text{-th vertex} \\ 0 & \text{if } h \text{ and } h' \text{ both end at the same vertex, but not at the } l\text{-th} \end{cases}$$

$$\varrho(A^{(i)}_{h,h'} f_{\mathrm{com}})(\circledast \xrightarrow{k} \odot) \;\; = \;\; \alpha/(w_{i-1} w_{i+1})$$

Let $\Xi := \sum A_{e,e'}^{(i-1)} f_{\text{com}}$. Then $\forall k : \varrho(\Xi)(\circledast \xrightarrow{k} \odot) = \alpha/w_{i-1}$ and $\forall l : \varrho(\Xi)(\odot \xrightarrow{l} \circledast) = \alpha$. Therefore, for e, e' starting at the k_0-th vertex and h, h' ending at the l_0-th vertex we have that

$$\varrho\left(w_{i-1}w_{i+1}\varrho(A_{e,e'}^{(i-1)} f_{\text{com}}) + w_{i-1}w_{i+1}\varrho(A_{h,h'}^{i} f_{\text{com}}) - \Xi\right)$$

is nonzero only on exactly two edges: $\circledast \xrightarrow{k_0} \odot$ and $\odot \xrightarrow{l_0} \circledast$. Cycles form a generating set of the vector space $F_{a,b,i}$, which finishes the proof of the surjectivity of ϱ. ◀

8   Flows on ABPs

We now proceed to the analysis of $\mathfrak{g}_0 f_{\text{com}}$ and $\mathfrak{g}_0 f_0$. The connection to flow vector spaces will be even more prevalent than in Proposition 12. The main result of this section is $\dim \mathfrak{g}_0 f_{\text{com}} > \dim \mathfrak{g}_0 f_0$ (Theorems 13 and 14), which implies that f_{com} and f_0 have different orbit dimensions. We thereby conclude that $f_0 \notin \mathsf{G} f_{\text{com}}$.

To each edge e we assign its *path tensor* $\psi(e)$ by summing tensors over all valid paths passing through e,

$$\psi(e) := \sum_{p \in \mathscr{P} \text{ with } e \in p} \varphi(p) \in E_1 \otimes \cdots \otimes E_d.$$

By linear continuation this gives a linear map $\psi : \mathbb{C}^E \to E_1 \otimes \cdots \otimes E_d$.

Observe that $\psi(e) = A_{e,e}^{(i)} f_{\text{com}}$. Let $\mathscr{T}$ denote the linear span of all $\psi(e)$, $e \in E$. In other words, $\mathscr{T} = \mathfrak{g}_0 f_{\text{com}}$.

Let $\mathscr{P}' \subseteq \mathscr{P} \subseteq \mathscr{X}$ be the set of valid paths that contain exactly one parity preserving edge. To each edge e we assign its *parity path tensor* $\psi'(e)$ by summing tensors over paths in $\mathscr{P}'$,

$$\psi'(e) := \sum_{p \in \mathscr{P}' \text{ with } e \in p} \varphi(p) \in E_1 \otimes \cdots \otimes E_d.$$

By linear continuation this gives a linear map $\psi' : \mathbb{C}^E \to E_1 \otimes \cdots \otimes E_d$. Observe that $\psi'(e) = A_{e,e}^{(i)} f_0$. Let $\mathscr{T}'$ denote the linear span of all $\psi'(e)$, $e \in E$. In other words, $\mathscr{T}' = \mathfrak{g}_0 f_0$.

We will establish the following bounds on the dimensions of $\mathscr{T}$ and $\mathscr{T}'$.

▶ **Theorem 13.** $\dim \mathscr{T} = |E| - \sum_{i=1}^{d} w_i + 1$.

▶ **Theorem 14.** $\dim \mathscr{T}' \leq |E| - \sum_{i=1}^{d} w_i$.

The rest of this section is dedicated to the proofs of Theorem 13 and Theorem 14 by showing that $\mathscr{T}$ is isomorphic to the vector space of flows "on the ABP", while the parity constraints lead to a smaller dimension of $\mathscr{T}'$.

From an ABP Γ we construct a digraph $\tilde{\Gamma}$ by identifying corresponding vertices from the first and the last layer in V and calling the resulting vertex set $\tilde{V}$. Note $|\tilde{V}| = \sum_{i=1}^{d} w_i$. The directed graphs Γ and $\tilde{\Gamma}$ have the same edge set. The resulting directed graph is called $\tilde{\Gamma} = (\tilde{V}, E)$. Let F denote the vector space of flows on $\tilde{\Gamma}$. Note that by (11) we have $\dim F = |E| - |\tilde{V}| + 1$. All directed cycles in $\tilde{\Gamma}$ have a length that is a multiple of d. In particular, all cycles of length exactly d are in one-to-one correspondence with valid paths in Γ_{com}. For an edge $e \in E$, let $\chi(e) \in \mathbb{C}^E$ denote the characteristic function of e, i.e., the function whose value is 1 on e and 0 everywhere else.

We now prove Theorem 13 by establishing a matching upper (Lemma 15) and lower bound (Lemma 16) of $|E| - |\tilde{V}| + 1 = \dim F$ on $\dim \mathscr{T}$.

The upper bound

▶ **Lemma 15.** $\dim \mathcal{T} \leq |E| - |\tilde{V}| + 1$.

Proof. For $v \in \tilde{V}$, let $\mathsf{in}(v) \subseteq E$ denote the set of incoming edges incident to v and $\mathsf{out}(v) \subseteq E$ denote the set of outgoing edges incident to v. For each $v \in \tilde{V}$, define the row vector

$$r_v = \sum_{e \in \mathsf{in}(v)} \chi(e) - \sum_{e \in \mathsf{out}(v)} \chi(e).$$

These vectors are the rows of the signed incidence matrix of $\tilde{\Gamma}$, and since $\tilde{\Gamma}$ is connected, they span a space of dimension $|\tilde{V}| - 1$ ([8, Ex. 1.5.6]). Now observe that for all $v \in \tilde{V}$,

$$\sum_{e \in \mathsf{in}(v)} \psi(e) = \sum_{e \in \mathsf{out}(v)} \psi(e).$$

Since ψ is linear, this is equivalent to

$$\psi \left(\sum_{e \in \mathsf{in}(v)} \chi(e) - \sum_{e \in \mathsf{out}(v)} \chi(e) \right) = 0.$$

Hence each r_v is in the kernel of ψ, and hence $\dim \ker \psi \geq |\tilde{V}| - 1$. Using (11), we obtain $\dim \mathcal{T} = \dim \mathrm{im}\, \psi = |E| - \dim \ker \psi \leq |E| - |\tilde{V}| + 1 = \dim F$. ◀

The lower bound

To obtain the lower bound, we define a linear map $\varrho : E_1 \otimes \cdots \otimes E_d \to \mathbb{C}^E$ such that the image of ϱ restricted to $\mathcal{T}$ equals F. This will imply that $\dim \mathcal{T} \geq \dim F$, thereby achieving the required lower bound.

We define the linear map ϱ on standard basis elements $x_{e_1} \otimes \cdots \otimes x_{e_d}$ as follows,

$$\varrho(x_{e_1} \otimes \cdots \otimes x_{e_d}) := \chi(e_1) + \cdots + \chi(e_d),$$

and then extend it to the domain $E_1 \otimes \cdots \otimes E_d$ via linear continuation.

▶ **Lemma 16.** *Let $\varrho|_{\mathcal{T}}$ denote the restriction of ϱ to the linear subspace $\mathcal{T}$. Then, $\mathrm{im}\, \varrho|_{\mathcal{T}} = F$. In particular, $\dim \mathcal{T} \geq \dim F = |E| - |\tilde{V}| + 1$.*

Proof. To prove equality it suffices to show $\mathrm{im}\, \varrho|_{\mathcal{T}} \subseteq F$ and $F \subseteq \mathrm{im}\, \varrho|_{\mathcal{T}}$.

The first containment is easy to see. For an edge e, consider the image of $\psi(e)$ under the map ϱ,

$$\varrho(\psi(e)) = \sum_{e \in p \in \mathscr{P}} \sum_{e' \in p} \chi(e').$$

Observe that for a path $p \in \mathscr{P}$, $\sum_{e' \in p} \chi(e')$ is a flow on $\tilde{\Gamma}$ and hence it belongs to F. Thus, we have $\varrho(\psi(e)) \in F$. Since $\mathcal{T}$ is spanned by $\psi(e)$, for $e \in E$, we obtain that $\mathrm{im}\, \varrho|_{\mathcal{T}} \subseteq F$.

To establish the second containment it suffices to show that the image of $\mathcal{T}$ under the map ϱ contains a basis of F. We identify a specific basis for F in Claim 17 and prove that it is contained in $\mathrm{im}\, \varrho|_{\mathcal{T}}$ in Claim 18 to complete the argument. ◀

We identify directed cycles with their characteristic flows, i.e., flows that have value 1 on the cycle's edges and 0 everywhere else. We also identify directed cycles that use edges
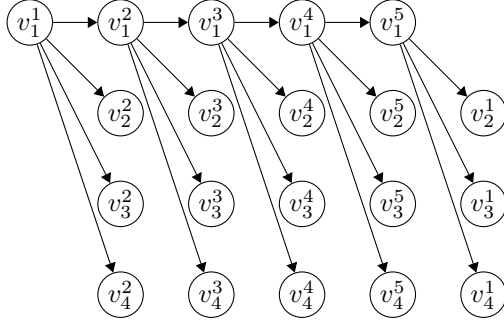
■ **Figure 1** The spanning tree construction for width 4 and $d = 5$.

in any direction with their characteristic flow: the characteristic flow is defined to take the value 1 on an edge e if e is traversed in the direction of e, and value -1 on e if e is traversed against its direction.

From the theory of flows we know that for every (undirected) spanning tree T of $\tilde{\Gamma}$, the vector space $F \in \mathbb{C}^E$ has a basis given by the characteristic flows of cycles that only use edges from T and exactly one additional edge (for example, see Theorem 20.8 in [8]). Thus, the cycle flows corresponding to the elements not in the spanning tree form a basis of F.

▷ Claim 17. F is spanned by the set of directed cycles in $\tilde{\Gamma}$ of length exactly d.

Proof. We construct a spanning tree τ as follows, which will be a tree whose edges are all directed away from its root. Informally, the tree is given by the following subgraph, we make the first vertex in V^1 as root, and include all the outgoing edges incident to it. We then move to the first vertex in V^2 and include all the outgoing edges incident to it. We continue in this way until we reach V^d. Upon reaching the first vertex in V^d we include all but one outgoing edges incident to it. The one that is an incoming edge to the root is not included. Figure 1 illustrates the construction. We now formally define this.

Let $v_1^i \in V^i$ denote the first vertex in the layer i, $1 \le i \le d$. Further recall $\mathsf{in}(v) \subseteq E$ and $\mathsf{out}(v) \subseteq E$ denote the set of incoming and outgoing edges, respectively, incident to v. Define the edge set

$$\tau := \left(\bigcup_{i=1}^{d} \mathsf{out}(v_1^i) \right) \setminus \{(v_1^d, v_1^1)\},$$

which is a spanning tree in $\tilde{\Gamma}$. We know that every edge not in the tree when added to the tree gives a unique undirected cycle. We now show that the characteristic flows of these undirected cycles can be expressed as a linear combination of the characteristic flows of directed cycles of length d. For $e \in E \setminus \tau$, let c_e denote the characteristic flow of the unique undirected cycle that uses e in its correct direction and only edges of τ. We argue depending on which layer the edge e belongs to.

▬ Suppose $e \in E^1 \setminus \tau$.
   ▪ If e is incident to v_1^2, the first vertex in V^2, then the inclusion of e creates a directed cycle of length d. Hence, c_e equals the characteristic flow of this directed cycle.
   ▪ Otherwise, the inclusion of e creates an undirected cycle of length $d+2$. If $e = (v_{j_1}^1, v_{j_2}^2)$ for some $j_1 \in [2, w_1]$ and $j_2 \in [2, w_2]$, then the cycle c_e is given as follows:

   $$v_1^d - v_{j_1}^1 - v_{j_2}^2 - v_1^1 - v_1^2 - \cdots - v_1^{d-1} - v_1^d.$$
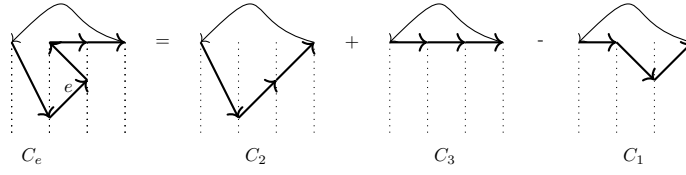
Figure 2 Decomposing a cycle of length $d + 2$ as a linear combination of cycles of length d. The figure is an illustration when $d = 3$. The dotted layers in each cycle from the left are V^3, V^1, V^2, and V^3 again.

Consider the following two directed cycles:

$$C_1 \; : \; v_1^1 - v_{j_2}^2 - \cdots - v_1^d - v_1^1 \text{ and}$$
$$C_2 \; : \; v_{j_1}^1 - v_{j_2}^2 - \cdots - v_1^d - v_{j_1}^1,$$

such that the part $v_{j_2}^2 - \cdots - v_1^d$ between $v_{j_2}^2$ and v_1^d in the two cycles is the same. Let us denote the characteristic flow of a cycle C by $\chi(C)$. We now observe that $\chi(C_2) - \chi(C_1)$ equals the characteristic flow of the undirected cycle $v_{j_1}^1 - v_{j_2}^2 - v_1^1 - v_1^d - v_{j_1}^1$. This is because the common part in C_1 and C_2 cancels out. To $\chi(C_2) - \chi(C_1)$ we add the characteristic flow of the directed cycle,

$$C_3 \; : \; v_1^1 - v_1^2 - v_1^3 - \cdots - v_1^{d-1} - v_1^d - v_1^1.$$

It is now easily seen that $\chi(C_2) - \chi(C_1) + \chi(C_3)$ equals the characteristic flow of the cycle c_e (see Figure 2 for an illustration).

- Suppose $e \in E^d \setminus \tau$.
  - If e is incident to v_1^1, the first vertex in V^1, then as before the inclusion of e creates a directed cycle of length d. Hence, c_e equals the characteristic flow of this directed cycle.
  - Otherwise, the inclusion of e creates an undirected cycle of length 4. If $e = (v_{j_1}^d, v_{j_2}^1)$ for some $j_1 \in [2, w_d]$ and $j_2 \in [2, w_1]$, then the cycle c_e is given as follows:

    $$v_{j_1}^d - v_{j_2}^1 - v_1^d - v_1^{d-1} - v_{j_1}^d.$$

   Consider the following two directed cycles:

    $$C_4 \; : \; v_{j_2}^1 - \cdots - v_1^{d-1} - v_1^d - v_{j_2}^1 \text{ and}$$
    $$C_5 \; : \; v_{j_2}^1 - \cdots - v_1^{d-1} - v_{j_1}^d - v_{j_2}^1,$$

    such that the part $v_{j_2}^1 - \cdots - v_1^{d-1}$ between $v_{j_2}^1$ and v_1^{d-1} in the two cycles is the same. We now claim that $\chi(C_5) - \chi(C_4)$ equals the characteristic flow of c_e. This is because the common part in C_4 and C_5 cancels out.

- Otherwise $e \in E^i \setminus \tau$ for some $i \in \{2, \ldots, d-1\}$. In such a case inclusion of e creates an undirected cycle of length 4. We can again argue exactly like in the previous case, and so we omit the argument here. ◄

We now prove that the generating set given by the directed cycles of length d is contained in the image of $\mathcal{T}$ under the map ϱ.
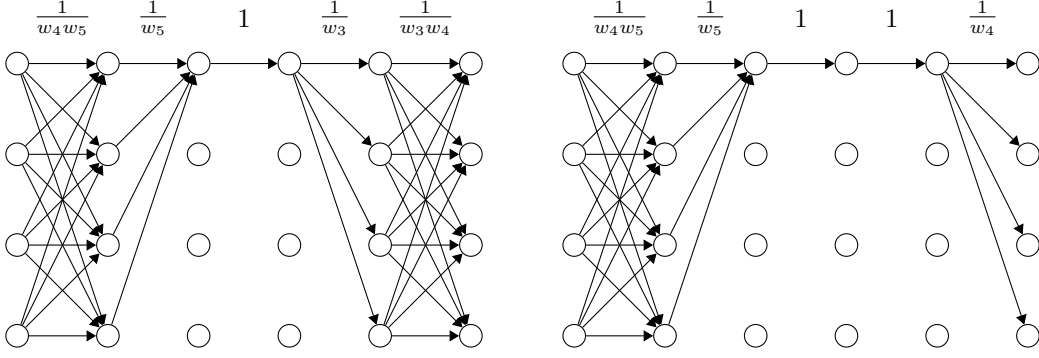
Figure 3 On the left: $\varrho(\bar{\psi}(e_1))$. On the right: $\varrho(\bar{\psi}(e_1)) - \frac{1}{w_3} \sum_{j=1}^{w_3-1} \varrho(\bar{\psi}(e_2^{(j)})) + \frac{w_3-1}{w_3} \varrho(\bar{\psi}(e_2))$. This is the case $d = 5$ and format $(4, 4, 4, 4, 4)$. Edges that are not drawn carry 0 flow. All edges in the same layer carry either 0 flow or the value that is depicted above the edge layer. For the purposes of illustation, e_1 is the top edge in the *center*. Here we assume that each e_i points from the first vertex in V^i to the first vertex in V^{i+1}.

▷ Claim 18. $\mathrm{im}(\varrho|_{\mathscr{T}})$ contains the characteristic flow of each directed cycle of length d.

Proof. Let $\{e_1, e_2, \ldots, e_d\} \subseteq E$ be a directed cycle of length d, where each e_i points from a vertex in V^i to a vertex in V^{i+1}. Let $\{e_i^{(j)}\}$ denote the set of edges that start at the same vertex as e_i, but for which $e_i^{(j)} \neq e_i$. Thus $|\{e_i^{(j)}\}| = |V^{i+1}| - 1$. Let

$$\bar{\psi}(e) := \frac{1}{|\{p \in \mathscr{P} \text{ with } e \in p\}|} \psi(e),$$

so that $\varrho(\bar{\psi}(e))$ is a flow with value 1 on the edge e. It is instructive to have a look at the left side of Figure 3, where $\varrho(\bar{\psi}(e_1))$ is depicted. Subtracting $\frac{1}{w_3} \sum_{j=1}^{w_3-1} \varrho(\bar{\psi}(e_2^{(j)}))$ and adding $\frac{w_3-1}{w_3} \varrho(\bar{\psi}(e_2))$ reduces the support significantly and brings us one step closer to the cycle, see the right side of Figure 3. We iterate this process until only the cycle is left. Formally:

$$
\begin{aligned}
\chi(e_1, \ldots, e_d) \quad = \quad & \varrho(\bar{\psi}(e_1)) \\
+ \quad & \tfrac{w_3-1}{w_3} \varrho(\bar{\psi}(e_2)) - \tfrac{1}{w_3} \sum_{j=1}^{w_3-1} \varrho(\bar{\psi}(e_2^{(j)})) \\
+ \quad & \cdots \\
+ \quad & \tfrac{w_d-1}{w_d} \varrho(\bar{\psi}(e_{d-1})) - \tfrac{1}{w_d} \sum_{j=1}^{w_d-1} \varrho(\bar{\psi}(e_{d-1}^{(j)})).
\end{aligned}
$$

◀

The stronger upper bound via parities

We now proceed to upper bound $\dim \mathscr{T}'$ (Theorem 14). The proof is analogous to the proof of Lemma 15.

▶ **Theorem 19** (Restatement of Theorem 14). $\dim \mathscr{T}' \leq |E| - |\tilde{V}|$.

Proof. As in the proof of Lemma 15, for $v \in \tilde{V}$, we have

$$\sum_{e \in \mathsf{in}(v)} \psi'(e) = \sum_{e \in \mathsf{out}(v)} \psi'(e).$$

Furthermore, we have the following additional constraint on ψ',

$$(d-1) \sum_{e \text{ parity preserving}} \psi'(e) = \sum_{e \text{ parity changing}} \psi'(e).$$

By the linearity of ψ', we have

$$\psi'\left((d-1) \sum_{e \text{ parity preserving}} \chi(e) - \sum_{e \text{ parity changing}} \chi(e)\right) = 0.$$

Therefore, the kernel of ψ' is spanned by the vectors $(\sum_{e \in \text{in}(v)} \chi(e) - \sum_{e \in \text{out}(v)} \chi(e))$, for $v \in \tilde{V}$, and an additional vector $((d-1) \sum_{e \text{ parity preserving}} \chi(e) - \sum_{e \text{ parity changing}} \chi(e))$.

We now claim that the new vector is linearly independent from the earlier set of vectors. We prove the claim by constructing a vector in $\mathbb{C}^E$ that is orthogonal to the earlier set of vectors but is non-orthogonal to the additional vector. One such vector is given by the characteristic flow of the directed cycle $v_1^1 - v_1^2 - v_1^3 - \cdots - v_1^{d-1} - v_1^d - v_1^1$.

Thus, it follows that $\dim \ker \psi' \geq |\tilde{V}|$, and hence $\dim \mathscr{T}' \leq |E| - |\tilde{V}|$.   ◀

9   VQP versus $\overline{\text{VNP}}$

In this section, we compare the complexity classes VQP and $\overline{\text{VNP}}$. Valiant in his seminal paper [41] defined the complexity classes that are now called as VP and VNP, and the central question of algebraic complexity is to understand whether the two complexity classes are indeed different as sets (Valiant's hypothesis). Bürgisser [11] defined the complexity class VQP and related it to the complexity classes VP and VNP. We proceed to define the above three classes for establishing the context. For an exhaustive treatment of the classes, we refer the readers to Bürgisser's monograph [11] from where we are lifting the definitions. We first need to define so-called p-families.

▶ **Definition 20.** *A sequence $f = (f_n)$ of multivariate polynomials over a field k is called a p-family (over k) iff the number of variables as well as the degree of f_n are bounded by polynomial functions in n.*

We now need to define the model of computation and the notion of complexity in order to define the complexity classes of interest.

▶ **Definition 21.** *A* straight-line program Γ *(expecting m inputs) represents a sequence $(\Gamma_1, \ldots, \Gamma_r)$ of instructions $\Gamma_\rho = (\omega_\rho; i_\rho, j_\rho)$ with operation symbols $\omega_\rho \in \{+, -, *\}$ and the address i_ρ, j_ρ which are integers satisfying $-m < i_\rho, j_\rho < \rho$. We call r the size of Γ.*

So, essentially, in a straight-line program, we either perform addition or subtraction or multiplication on the inputs or the previously computed elements. The size of the straight-line program naturally induces a size complexity measure on polynomials as follows:

▶ **Definition 22.** *The* complexity $L(f)$ of a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is the minimal size of a straight-line program computing f from variables x_i and constants in $\mathbb{F}$.

We are now all set to define the above discussed complexity classes.

▶ **Definition 23.** *A p-family $f = (f_n)$ is said to be p-computable iff the complexity $L(f_n)$ is a polynomially bounded function of n. $\text{VP}_{\mathbb{F}}$ consists of all p-computable families over the field $\mathbb{F}$.*

▶ **Definition 24.** *A p-family $f = (f_n)$ is said to be p-*definable *iff there exists a p*-computable *family $g = (g_n)$, $g_n \in \mathbb{F}[x_1, \ldots, x_{u(n)}]$, such that for all n*

$$f_n(x_1, \ldots, x_{v(n)}) = \sum_{e \in \{0,1\}^{u(n)-v(n)}} g_n(x_1, \ldots, x_{v(n)}, e_{v(n)+1}, \ldots, e_{u(n)}).$$

The set of p-definable families over $\mathbb{F}$ forms the complexity class $\mathrm{VNP}_\mathbb{F}$.

▶ **Definition 25.** *A p-family $f = (f_n)$ is said to be qp*-computable *iff the complexity $L(f_n)$ is a quasi-polynomially bounded function of n. The complexity class* $\mathrm{VQP}_\mathbb{F}$ *consists of all qp-computable families over $\mathbb{F}$.*

In the above three definitions, if the underlying field is clear from the context, we can drop the subscript $\mathbb{F}$ and simply represent the classes as VP, VNP and VQP respectively. In what follows, the underlying field is always assumed to be $\mathbb{Q}$, the field of rational numbers.

In [11], Bürgisser showed the completeness of the determinant polynomial for VQP under *qp*-projections and strengthened Valiant's hypothesis of VNP $\not\subseteq$ VP to VNP $\not\subseteq$ VQP and called it *Valiant's extended hypothesis* (see [11], Section 2.5). He also established that VP $\subsetneq$ VQP and went on to show that VQP $\not\subseteq$ VNP (see [11], Proposition 8.5 and Corollary 8.9). The main observation of this section is that his proof is stronger and is sufficient to conclude that VQP is not contained in the closure of VNP either, where the closure is in the sense as mentioned in Section 1.

In fact, Bürgisser in his monograph [11] also gives a set of conditions which if the coefficients of a polynomial sequence satisfies, then that polynomial sequence cannot be in VNP [11, Theorem 8.1]. His theorem and the proof is inspired by Heintz and Sieveking [22]. The second observation of this section is that this proof is even stronger and actually those conditions are sufficient to show that the given polynomial sequence is not contained in $\overline{\mathrm{VNP}}$ either.

We now discuss both the observations.

9.1   VQP $\not\subseteq$ $\overline{\mathrm{VNP}}$

We first show that there is a $\log n$ variate polynomial of degree $(n-1)\log n$ which is in VQP but not in $\overline{\mathrm{VNP}}$. In this exposition, for the sake of better readability, we do not present the Bürgisser's statements in full generality since it is not essential for the theorem that we want to show here. Moreover, the less general version that we present here contains all the ideas for the theorem statements and their proofs.

▶ **Theorem 26.** *Let $N_n := \{0, \ldots, n-1\}^{\log n}$ and $f_n := \sum_{\mu \in N_n} 2^{2^{j(\mu)}} X_1^{\mu_1} \cdots X_{\log n}^{\mu_{\log n}}$, where $j(\mu) := \sum_{j=1}^{\log n} \mu_j n^{j-1}$. Then $f_n \in \mathrm{VQP}$, but $f_n \notin \overline{\mathrm{VNP}}$, and hence $\mathrm{VQP} \not\subseteq \overline{\mathrm{VNP}}$.*

The theorem consists of two parts. The containment in VQP follows immediately from the fact that the total number of monomials in f_n is $n^{\log n}$. For the other part, we closely follow Bürgisser's lower bound proof [11, Proposition 8.5] against VNP here, making transparent the fact that the proof works also against $\overline{\mathrm{VNP}}$. His proof techniques were borrowed from Strassen ([39]). The idea is to use the universal representation for polynomial sequences in VNP, so that we get a hold on how the coefficients of the polynomials look like. Using that, we establish polynomials H_n that vanish on all the polynomial sequences in VNP (in other words, H_n is in the vanishing ideal of sequences in VNP), but do not vanish on f_n (because the growth rate of its coefficients is too high), hence giving the separation. Since the vanishing ideal of a set characterizes its closure, we get the stronger separation, i.e., f_n does not belong to the closure of VNP, namely, $\overline{\mathrm{VNP}}$.

Proof of Theorem 26. As stated above, the proof works in three stages: first, assuming the contrary and writing f_n using the universal representation for the polynomial sequences in VNP, then giving polynomials H_n of special forms in the vanishing ideal of polynomial sequences in VNP, and finally showing that H_n cannot vanish on our sequence f_n, hence arriving at a contradiction.

Assuming $(f_n) \in$ VNP implies the existence of a family $(g_n) \in$ VP, with $L(g_n)$ bounded by a polynomial $r(n)$, and a polynomial $u(n)$ such that

$$f_n(X_1, \ldots, X_{\log n}) = \sum_{e \in \{0,1\}^{u(n)-\log n}} g_n(X_1, \ldots, X_{\log n}, e_{\log n+1}, \ldots, e_{u(n)}).$$

Next, we use the universal representation theorem (see [39], [37]) as stated in Bürgisser's monograph ([11], Proposition 8.3; for a proof see [13], Proposition 9.11) for size $r(n)$ straight-line program to get that there exist polynomials $G_\nu^{(n)} \in \mathbb{Z}[Y_1, \ldots, Y_{q(n)}]$, with $q(n)$ being a polynomial in n (more precisely, it is a polynomial in $r(n)$ and $u(n)$) which for $|\nu| \leq \deg g_n = n^{O(1)}$, guarantee that $\deg G_\nu = n^{O(1)}, \log \mathsf{wt}(G_\nu)^{(n)} = 2^{n^{O(1)}}$, and also guarantee the existence of some $\zeta \in \overline{\mathbb{Q}}^{q(n)}$, such that

$$g_n = \sum_\nu G_\nu^{(n)}(\zeta) X_1^{\nu_1} \cdots X_{u(n)}^{\nu_{u(n)}},$$

where weight of a polynomial f, $\mathsf{wt}(f)$ refers to the sum of the absolute values of its coefficients.

Now, taking exponential sum yields that

$$f_n = \sum_{\mu \in N_n} F_\mu^{(n)}(\zeta) X_1^{\mu_1} \cdots X_{\log n}^{\mu_{\log n}},$$

where the polynomials $F_\mu^{(n)}$ are obtained as a sum of at most $2^{u(n)}$ polynomials $G_\nu^{(n)}$. Thus, we now have a good hold on $F_\mu^{(n)}$, i.e. $\deg F_\mu^{(n)} \leq \alpha(n)$ and $\log \mathsf{wt}(F_\mu^{(n)}) \leq 2^{\beta(n)}$, where both $\alpha(n)$ and $\beta(n)$ are polynomially bounded functions of n.

Thus, for f_n to be in VNP, the coefficients of f_n should be in the image of the polynomial map $F_\mu^n : \overline{\mathbb{Q}}^{q(n)} \to \overline{\mathbb{Q}}^{n^{\log n}}$. In other words, we must have some $\zeta \in \overline{\mathbb{Q}}^{q(n)}$, such that for all $\mu \in N_n$, we have $F_\mu^n(\zeta) = 2^{2^{j(\mu)}}$, where $j(\mu) := \sum_{j=1}^{\log n} \mu_j n^{j-1}$. Since F_μ^n takes all the values from 2^{2^0} to $2^{2^{n^{\log n}-1}}$, we have a subset of indices $\tilde{N}_n \subseteq N_n$ of size $s(n) := \lfloor |N_n|/n \rfloor = \lfloor n^{\log n}/n \rfloor$, such that for $\sigma \in \{0, 1, \ldots, s(n)-1\}$ and a bijection $\delta : \{0, 1, \ldots, s(n)-1\} \to \tilde{N}_n$ with $\sigma \mapsto \delta(\sigma)$, we have $F_{\delta(\sigma)}^n = 2^{2^{\sigma n+1}}$.

Now we can apply Lemma 9.28 from [13] which asserts that there will be polynomials of low height (ht) (the maximum of the absolute value of the coefficients) on which these coefficients shall vanish. More precisely, there exists non-zero forms $H_n \in \mathbb{Z}[Y_\mu \mid \mu \in \tilde{N}_n]$ with $\mathsf{ht}(H_n) \leq 3$, $\deg H_n \leq D(n)$, and such that $H_n(F_\mu^n \mid \mu \in N_n) = 0$, given that $D(n)^{s(n)-q(n)-2} > \alpha(n)^{q(n)} s(n)^{s(n)} 2^{\beta(n)}$.

It can be seen that $D(n) = 2^n - 1$ satisfies the above inequality, since $\alpha(n), \beta(n)$ and $q(n)$ are polynomially bounded and 2^n grows much faster than $s(n) = \lfloor n^{\log n}/n \rfloor$. This allows us to write $H_n = \sum_e \lambda_e \prod_{\mu \in \tilde{N}_n} Y_\mu^{e_\mu}$, where the absolute values of λ_e are bounded by 3. Since H_n vanishes on the subset of coefficients of f_n, i.e it vanishes on $F_{\delta(\sigma)}^n = 2^{2^{\sigma n+1}}$ with $\sigma \in \{0, 1, \ldots, s(n)-1\}$, we have

$$0 = H_n(F_\mu^n \mid \mu \in \tilde{N}_n) = \sum_e \lambda_e \prod_{\sigma=0}^{s(n)-1} 2^{e_{\delta(\sigma)} 2^{\sigma n+1}} = \sum_e \lambda_e \cdot 4^{\sum_\sigma e_{\delta(\sigma)}(2^n)^\sigma}.$$

The last sum is essentially a 4-adic integer, since firstly, $|\lambda_e| \leq 3$, and secondly, all the exponents of 4, that is, $\sum_\sigma e_{\delta(\sigma)}(2^n)^\sigma$ are all distinct, as they can be seen as 2^n-adic representation since $e_{\delta(\sigma)} < 2^n$. Thus λ_e has to be zero for all e. Hence H_n must be identically zero, which is a contradiction.   ◄

9.2   A criterion for non-membership in $\overline{\text{VNP}}$

In this section, we discuss a criterion Bürgisser presented in his monograph [11] based on a proof due to Heintz and Sieveking which gives a set of conditions that puts a p-family out of VNP. We observe that those conditions if satisfied, in fact, put a given p-family out of $\overline{\text{VNP}}$ as well.

▶ **Theorem 27.** *Let (p_n) be a sequence of polynomials over $\overline{\mathbb{Q}}$ and let $N(n)$ denote the degree of the field extension generated by the coefficients of p_n over $\mathbb{Q}$. Further suppose the following holds:*

1. *The map $n \mapsto \lceil \log N(n) \rceil$ is not p-bounded.*
2. *For all n, there is a system G_n of rational polynomials of degree at most $D(n)$ with finite zeroset, containing the coefficient system of f_n, and such that $n \mapsto \lceil \log D(n) \rceil$ is p-bounded.*

Then the family $(p_n) \notin \overline{\text{VNP}}$.

Thus the above theorem shows that certain p-families with algebraic coefficients of high degree are not contained in $\overline{\text{VNP}}$. We now give a simple example from [11] to illustrate the theorem.

▶ **Example 28.** Consider the following multivariate family defined as

$$p_n = \sum_{e \in \{0,1\}^n \setminus 0} \sqrt{p_{j(e)}} X^e,$$

where $j(e) = \sum_{s=1}^n e_s 2^{s-1}$ and p_j refers to the j-th prime number. Then using the above Theorem 27, we can conclude that $p_n \notin \overline{\text{VNP}}$. This is because the degree of field extension $N(n) = [\mathbb{Q}(\sqrt{p_j} \mid 1 \leq j \leq 2^n) : \mathbb{Q}] = 2^{2^n - 1}$ (see for example [13], Lemma 9.20), hence condition 1 above is satisfied. Condition 2 is also satisfied because the coefficients are the roots of the system $G_n = \{Z_j^2 - p_j \mid 1 \leq j < 2^n\}$, with $D(n) = 2$.

For a proof of the theorem, we refer the readers to [11, Theorem 8.1]. We point out that the proof in its original form already works. In his proof, he wanted to conclude that $f_n \notin \text{VNP}$. However, along the way, he arrives at a contradiction to the assertion that f_n is contained in the Zariski-closure of VNP, which is exactly what is now known as $\overline{\text{VNP}}$. During the time of the original proof, the complexity class $\overline{\text{VNP}}$ was not defined.

──── **References** ────

1   Eric Allender and Fengming Wang. On the power of algebraic branching programs of width two. *Comput. Complex.*, 25(1):217–253, March 2016. URL: http://dx.doi.org/10.1007/s00037-015-0114-7, doi:10.1007/s00037-015-0114-7.

2   Jarod Alper, Tristram Bogart, and Mauricio Velasco. A lower bound for the determinantal complexity of a hypersurface. *Foundations of Computational Mathematics*, pages 1–8, 2015.

3   Matthew Anderson, Michael A. Forbes, Ramprasad Saptharishi, Amir Shpilka, and Ben Lee Volk. Identity testing and lower bounds for read-k oblivious algebraic branching programs. In *Proceedings of the 31st Conference on Computational Complexity*, CCC '16, Dagstuhl, DEU, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

4   Michael Ben-Or and Richard Cleve. Computing algebraic formulas using a constant number of registers. *SIAM J. Comput.*, 21(1):54–58, 1992. URL: `https://doi.org/10.1137/0221006`, `doi:10.1137/0221006`.

5   D. Bini. Relations between exact and approximate bilinear algorithms. applications. *CAL-COLO*, 17(1):87–97, Jan 1980. URL: `https://doi.org/10.1007/BF02575865`, `doi:10.1007/BF02575865`.

6   Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti. O($n^{2.7799}$) complexity for $n \times n$ approximate matrix multiplication. *Inf. Process. Lett.*, 8(5):234–235, 1979.

7   Markus Bläser and Christian Ikenmeyer. Introduction to geometric complexity theory. `http://pcwww.liv.ac.uk/~iken/teaching_sb/summer17/introtogct/gct.pdf`, version from July 25, 2018.

8   J.A. Bondy and U.S.R Murty. *Graph Theory*. Springer Publishing Company, Incorporated, 2008.

9   Karl Bringmann, Christian Ikenmeyer, and Jeroen Zuiddam. On algebraic branching programs of small width. *J. ACM*, 65(5), August 2018. URL: `https://doi.org/10.1145/3209663`, `doi:10.1145/3209663`.

10   Peter Bürgisser. Erratum to the complexity of factors of multivariate polynomials. `https://www.math.tu-berlin.de/fileadmin/i26_fg-buergisser/erratum_factors.pdf`.

11   Peter Bürgisser. *Completeness and Reduction in Algebraic Complexity Theory*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

12   Peter Bürgisser. The complexity of factors of multivariate polynomials. *Found. Comput. Math.*, 4(4):369–396, 2004. URL: `https://doi.org/10.1007/s10208-002-0059-5`, `doi:10.1007/s10208-002-0059-5`.

13   Peter Bürgisser, Michael Clausen, and Mohammad A Shokrollahi. *Algebraic complexity theory*, volume 315. Springer Science & Business Media, 2013.

14   Peter Bürgisser, J.M. Landsberg, Laurent Manivel, and Jerzy Weyman. An overview of mathematical issues arising in the Geometric complexity theory approach to VP v.s. VNP. *SIAM J. Comput.*, 40(4):1179–1209, 2011.

15   Prerona Chatterjee, Mrinal Kumar, Adrian She, and Ben Lee Volk. A quadratic lower bound for algebraic branching programs. *CoRR*, abs/1911.11793, 2019. To appear in the proceedings of Conference on Computational Complexity, CCC'20. URL: `http://arxiv.org/abs/1911.11793`.

16   Michael Forbes. Some concrete questions on the Border Complexity of polynomials, 2016. Talk at the Workshop on Algebraic Complexity Theory (WACT) 2016 in Tel Aviv. URL: `https://www.youtube.com/watch?v=1HMogQIHT6Q`.

17   Hervé Fournier, Guillaume Malod, Maud Szusterman, and Sébastien Tavenas. Nonnegative rank measures and monotone algebraic branching programs. In *39th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 150 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 15, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019.

18   Fulvio Gesmundo. Geometric aspects of iterated matrix multiplication. *Journal of Algebra*, 461:42 – 64, 2016.

19   Fulvio Gesmundo, Christian Ikenmeyer, and Greta Panova. Geometric complexity theory and matrix powering. *Differential Geom. Appl.*, 55:106–127, 2017. URL: `https://doi.org/10.1016/j.difgeo.2017.07.001`, `doi:10.1016/j.difgeo.2017.07.001`.

20   Bruno Grenet. An upper bound for the permanent versus determinant problem. Manuscript, 2011. URL: `http://www.lirmm.fr/~grenet/publis/Gre11.pdf`.

21   Joshua A. Grochow, Ketan D. Mulmuley, and Youming Qiao. Boundaries of VP and VNP. In *43rd International Colloquium on Automata, Languages, and Programming*, volume 55 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 34, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.

22 Joos Heintz and Malte Sieveking. Lower bounds for polynomials with algebraic coefficients. *Theoretical Computer Science*, 11(3):321–330, 1980.

23 Jesko Hüttenhain. *Geometric Complexity Theory and Orbit Closures of Homogeneous Forms*. PhD thesis, Technische Universität Berlin, July 2017.

24 Jesko Hüttenhain and Pierre Lairez. The boundary of the orbit of the 3-by-3 determinant polynomial. *Comptes Rendus Mathematique*, 354(9):931 – 935, 2016.

25 Neeraj Kayal, Vineet Nair, Chandan Saha, and Sébastien Tavenas. Reconstruction of full rank algebraic branching programs. *ACM Trans. Comput. Theory*, 11(1), November 2018. URL: `https://doi.org/10.1145/3282427`, `doi:10.1145/3282427`.

26 Mrinal Kumar. On top fan-in vs formal degree for depth-3 arithmetic circuits. `https://eccc.weizmann.ac.il/report/2018/068/`, 2018.

27 Mrinal Kumar. A quadratic lower bound for homogeneous algebraic branching programs. *computational complexity*, 28(3):409–435, 2019. URL: `https://doi.org/10.1007/s00037-019-00186-3`, `doi:10.1007/s00037-019-00186-3`.

28 J. M. Landsberg. Geometric complexity theory: an introduction for geometers. *ANNALI DELL'UNIVERSITA' DI FERRARA*, 61(1):65–117, May 2015.

29 J. M. Landsberg. *Geometry and Complexity Theory*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2017. `doi:10.1017/9781108183192`.

30 Meena Mahajan and V. Vinay. Determinant: combinatorics, algorithms, and complexity. *Chicago J. Theoret. Comput. Sci.*, pages Article 5, 26 pp. (electronic), 1997.

31 Guillaume Malod and Natacha Portier. Characterizing Valiant's algebraic complexity classes. *Journal of Complexity*, 24(1):16 – 38, 2008.

32 K.D. Mulmuley and M. Sohoni. Geometric Complexity Theory. I. An approach to the P vs. NP and related problems. *SIAM J. Comput.*, 31(2):496–526 (electronic), 2001.

33 K.D. Mulmuley and M. Sohoni. Geometric Complexity Theory. II. Towards explicit obstructions for embeddings among class varieties. *SIAM J. Comput.*, 38(3):1175–1206, 2008.

34 D. Mumford. *Algebraic geometry. I: Complex projective varieties*. Classics in mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1976 edition in Grundlehren der mathematischen Wissenschaften, vol. 221.

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

36 Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *J. ACM*, 60(6):Art. 40, 15, 2013. URL: `https://doi.org/10.1145/2535928`, `doi:10.1145/2535928`.

37 Claus-Peter Schnorr. Improved lower bounds on the number of multiplications/divisions which are necessary to evaluate polynomials. In *International Symposium on Mathematical Foundations of Computer Science*, pages 135–147. Springer, 1977.

38 Srikanth Srinivasan. Strongly exponential separation between monotone VP and monotone VNP. arXiv:1903.01630, 2019.

39 Volker Strassen. Polynomials with rational coefficients which are hard to compute. *SIAM Journal on Computing*, 3(2):128–149, 1974.

40 Seinosuke Toda. Classes of Arithmetic Circuits Capturing the Complexity of the Determinant. *IEICE TRANS. INF. & SYST.*, E75-D(1):116–124, 1992.

41 L. G. Valiant. Completeness classes in algebra. In *Conference Record of the Eleventh Annual ACM Symposium on Theory of Computing (Atlanta, Ga., 1979)*, pages 249–261. ACM, New York, 1979.

42 Amir Yehudayoff. Separating monotone VP and VNP. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, page 425–429, New York, NY, USA, 2019. Association for Computing Machinery. URL: `https://doi.org/10.1145/3313276.3316311`, `doi:10.1145/3313276.3316311`.