On the gradient of the coefficient of the characteristic polynomial

Christian Ikenmeyer

November 7, 2025

Abstract

We prove the bivariate Cayley–Hamilton theorem, a powerful generalization of the classical Cayley–Hamilton theorem. The bivariate Cayley–Hamilton theorem has three direct corollaries that are usually proved independently: The classical Cayley–Hamilton theorem, the Girard-Newton identities, and the fact that the determinant and every coefficient of the characteristic polynomial has polynomially sized algebraic branching programs (ABPs) over arbitrary commutative rings. This last fact could so far only be obtained from separate constructions, and now we get it as a direct consequence of this much more general statement.

The statement of the bivariate Cayley–Hamilton theorem involves the gradient of the coefficient of the characteristic polynomial, which is a generalization of the adjugate matrix. Analyzing this gradient, we obtain another new ABP for the determinant and every coefficient of the characteristic polynomial. This ABP has one third the size and half the width compared to the current record-holder ABP constructed by Mahajan–Vinay in 1997. This is the first improvement on this problem for 28 years.

Our ABP is built around algebraic identities involving the first order partial derivatives of the coefficients of the characteristic polynomial, and does not use the ad-hoc combinatorial concept of clow sequences. This answers the 26-year-old open question by Mahajan–Vinay from 1999 about the necessity of clow sequences.

We prove all results in a combinatorial way that on a first sight looks similar to Mahajan–Vinay, but it is closer to Straubing's and Zeilberger's constructions.

Keywords: Cayley–Hamilton theorem, Girard–Newton identities, determinant, characteristic polynomial, algebraic complexity theory, GapL, parameterized complexity

Contents

1 Introduction

The determinant polynomial

$$\det_d \;=\; \sum_{\pi\in\mathfrak{S}_d} \operatorname{sgn}(\pi)\, x_{i,\pi(i)}$$

is the most prominent of the coefficients of the characteristic polynomial. It is ubiquitous throughout mathematics and has been studied for over 200 years. In counting complexity theory, it is complete for the class GapL, and in algebraic complexity theory, it is complete for the class VBP. The Cayley–Hamilton theorem is one of the most fundamental results in linear algebra. It states that every square matrix is a root of its own characteristic polynomial. If X_n is the $n \times n$ matrix of variables $X_n = \left(x_{i,j}\right)_{\substack{1\le i\le n\\ 1\le j\le n}}$, the characteristic polynomial is defined as

$$\det(X_n + t\cdot I_d) \;=\; \sum_{d\ge 0} \chi_{n,n-d}\cdot t^d \tag{1.1}$$

where I_n is the $n \times n$ identity matrix and t is an indeterminate. The Cayley–Hamilton theorem states that $\sum_{i=0}^{n}(-1)^i \chi_{n,n-i} X_n^i$ is the zero matrix. Equation (1.1) defines the homogeneous degree d coefficient polynomials $\chi_{n,d}$, which will play a major role in this paper. From (1.1) it follows that

$$\chi_{n,d} \;=\; \sum_{S\in\binom{\{1,\dots,n\}}{d}} \det([X_n]_{S,S}), \tag{1.2}$$

where $[X]_{S,S}$ is the square submatrix of X with row indices S and column indices S, and $\binom{A}{d}$ is the set of cardinality d subsets of the set A. Note that $\chi_{d,d} = \det_d$ and $\chi_{n,1} = \operatorname{tr}_n := \operatorname{tr}(X_n)$ is the trace.

Commutative rings The above discussion about the determinant, the characteristic polynomial, and the Cayley–Hamilton theorem works over arbitrary commutative rings R, and this is the generality we work in. This is the common setup in commutative algebra, see for example [Eis13, §0.1]: "Nearly every ring treated in this book is commutative, and we shall generally omit the adjective". The specific search for algorithms that work over all commutative rings goes at least back to [Val92] and the concept of Valiant's "ARP" (all rings polynomially-sized circuits), and is again prominently featured in [MV97b]. Let $R[\mathbf{x}]$ denote the polynomial ring over R, and let $R[\mathbf{x}]_d$ denote its homogeneous degree d component. Note that $\chi_{n,d} \in R[\mathbf{x}]_d$.

In algebraic complexity theory, constructions do not always work over arbitrary commutative rings, for example when the Girard–Newton identities are involved. This was an issue in the recent breakthrough lower bound [LST25], which was made field-independent later in [For24] (best paper award at CCC). We shed some light on the Girard–Newton identities in §2, and in fact our bivariate Cayley–Hamilton theorem is a generalization of those identities. Field-independent replacements for field-dependent constructions are a goal in several areas of mathematics, for example algebraic geometry, representation theory, and invariant theory, see for example [Eis13, AFP$^+$19, DV22].

Motivation: Width as the bivariate perspective on ABPs Every homogeneous degree d polynomial f can be written as the bottom-right entry of a product of exactly d many $n \times n$ matrices whose entries are homogeneous linear polynomials, for some large n. The *width* $w(f)$ is defined as the smallest n such that this is possible. We call a d-tuple of such $n \times n$ matrices a pure algebraic branching program (pABP). A trivial upper bound for $w(f)$ is the number of monomials of f: In this straightforward upper bound construction, the first matrix has nonzero entries only in the last row, the last matrix has nonzero entries only in the last column, and all other matrices are diagonal. The width is robust in the sense that it is the same notion when we study so-called homogeneous algebraic branching programs (ABPs) instead of pABPs, see Proposition 7.1 below.

An ABP is a directed acyclic graph in which each vertex has a layer from $0, \ldots, d$, and each edge is either going from a vertex in layer i to another vertex in layer i and is labeled by an element of R (these are called constant edges), or the edge is going from a vertex in layer i to a vertex in layer $i+1$ and is labeled by an element of $R[\mathbf{x}]_1$, see for example Figure 6a. One vertex in layer 0 is called the source s, and one vertex in layer d is called the sink t. To an ABP G we assign its computation result f, which is the weighted sum over all s-t-paths, where each summand is weighted by the product of its edge labels. We say that G computes f. The width of an ABP is defined as the maximum number of vertices in any layer $1, \ldots, d-1$. The smallest width of an ABP computing f equals $w(f)$, see Proposition 7.1. The ABP is a graph theoretic variant of the algebraically pure notion of the pABP, and we will see in Proposition 2.6 that it is sometimes convenient to work with ABPs instead of pABPs. The width is also closely related to the so-called affine algebraic branching program size, see §7.

We can think of the pair $\big(w(f), \deg(f)\big)$ as a highly algebraically idealized way of measuring the running time requirements to evaluate f, which leans heavily towards the algebraic side, but captures arithmetic circuit size up to a quasipolynomial blowup [vzG87]. The degree measures a part of the computational complexity f, but since for any $f \in R[\mathbf{x}]_d$ the degree is exactly d, the width is the more interesting quantity. However, unlike all other ways of measuring complexity, width and degree must always be viewed together as a tuple, because there are sequences of polynomials of constant width with unbounded degree (and hence the evaluation time diverges when d increases) and there are sequences of polynomials of constant degree with unbounded width (for example, the degree 3 $n \times n$ matrix multiplication polynomials). There is the following duality between the degree and the width with respect to subadditivity and submultiplicativity. For $f, h \in R[\mathbf{x}]_d$ we have

$$\begin{aligned} \deg(f+h) &\leq \max(\deg(f), \deg(h)) & \deg(fh) &\leq \deg(f) + \deg(h) \\ w(f+h) &\leq w(f) + w(h) & w(fh) &\leq \max(w(f), w(h)) \end{aligned} \tag{1.3}$$

where we used the conventions $\deg(0) = 0$ and $w(0) = 0$, see §7.II. The determinantal complexity $\mathsf{dc}(f)$ is defined as the smallest n such that f can be written as the determinant of an $n \times n$ matrix with entries from $R[\mathbf{x}]_0 \oplus R[\mathbf{x}]_1$, i.e., affine linear entries. The fact that the entries must be affine cannot avoided here, because the determinant polynomial $\det_d$ is just single-indexed. Moreover, while $\mathsf{dc}(fh) \leq \mathsf{dc}(f) + \mathsf{dc}(h)$, no nice relation is known for $\mathsf{dc}(f+h)$ (only with a blowup, and only via a reduction to ABPs). Fortunately, determinantal complexity is closely related to width as follows:

$$\tfrac{1}{d}\,\mathsf{dc}(f) \;\leq\; w(f) \;\overset{(*)}{\leq}\; d^4\,\mathsf{dc}(f), \tag{1.4}$$

where $(*)$ is currently only known to hold over fields, see Proposition 7.5. Clearly, $\mathsf{dc}(\det_d) = d$, but $(*)$ cannot be used to prove $w(\det_d) \in \mathsf{poly}(d)$, because the proof of $(*)$ uses this fact as an ingredient. Compared to other ways of measuring computational complexity, the width can be used to study n and d independently, and hence it is well suited for studying bivariate complexity, which is commonly known as parameterized complexity, see [DF13, Preface].

Motivation: The bivariate viewpoint in complexity theory The study of bivariate polynomial families where one index is the degree has recently been pioneered in algebraic complexity theory by [BE19], translating results from Boolean parameterized complexity theory [DF99, Nie06, FG06, DF13] into algebraic complexity theory. From the Boolean perspective, [DF13] write "the multivariate perspective has proved useful, even arguably essential". We will see in this paper that the multivariate perspective is also valuable in linear algebra. The general theory in [DF13] allows different so-called "parameters", and finding the "correct" parameter can be an art. But when studying homogeneous polynomials, the canonical choice for the parameter is the degree, see [BE19, Def. 4.7(2), Def. 4.15(2)]. Treating the degree as a separate parameter very recently made

its appearance in several other papers in algebraic complexity theory [DGI+24, CKV24, vdBDG+25]. For double-indexed polynomials $f_{n,d}$ we define the exponents

$$\gamma(f_{\cdot,d}) := \limsup_{n\to\infty} \big(\log_n(w(f_{n,d}))\big).$$

As seen in (1.4), these exponents are invariant under exchanging the width with determinantal complexity. They are also invariant under exchanging the width with the affine algebraic branching program size, see (7.2) and Proposition 7.3. The complexity class VBFPT consists of those $f_{n,d}$ for which the set $\{\gamma(f_{\cdot,d}) \mid d \in \mathbb{N}\}$ is bounded, see also Proposition 7.7. The univariate class VBP consists of [1] those single-indexed f_d for which $w(f_d) \in \mathsf{poly}(d)$. The variant of VBFPT that uses arithmetic circuit complexity instead of width is defined in [BE19], which gives rise to a potentially larger class VFPT. One can relate different conjectures in algebraic complexity theory via the clique polynomial, for example. Let $\mathrm{Clique}_{n,d} = 0$ if $d \neq \binom{k+1}{2}$ for some k, and let

$$\mathrm{Clique}_{n,\binom{k+1}{2}} \;=\; \sum_{1\le v_1<v_2<\cdots<v_k\le n} \prod_{1\le i\le j\le k} x_{v_i,v_j}.$$

Then by [BE19] and [Bür00a, Proof of Thm 3.10] we have

$$\begin{array}{ccccc} \mathrm{Clique}_{n,d} \notin \mathrm{VFPT} & \Rightarrow & \mathrm{Clique}_{n,d} \notin \mathrm{VBFPT} & \Rightarrow & \mathrm{Clique}_{2d,d} \notin \mathrm{VBP} \\ \Updownarrow & & \Updownarrow & & \Updownarrow \\ \mathrm{VFPT} \neq \mathrm{VW}[1] & \Rightarrow & \mathrm{VBFPT} \neq \mathrm{VW}[1] & \Rightarrow & \mathrm{VBP} \neq \mathrm{VNP} \end{array}$$

The definitions of VNP [Bür00b] and VW[1] [BE19] will not play a role in this paper. The rightmost vertical equivalence holds in characteristic $\neq 2$. In characteristic $\neq 2$, the conjecture VBP $\neq$ VNP is also known as Valiant's determinant vs permanent conjecture. It can be conveniently expressed bivariately as $w(\mathrm{Clique}_{n,d}) \notin \mathsf{poly}(n,d)$.

The bivariate viewpoint, i.e., the separation of n and the degree d, is also standard for every notion of a *rank* of a tensor or polynomial, see for example [BL13].

Main results: The bivariate viewpoint in linear algebra We show that the bivariate viewpoint appears naturally in linear algebra, a priori outside of algebraic complexity theory, but with direct consequences such as $w(\chi_{n,d}) \in \mathsf{poly}(n,d)$. For this purpose, we prove our first main result, the bivariate Cayley–Hamilton theorem 5.1. For a polynomial f in n^2 variables $x_{i,j}$, $1 \le i,j \le n$, let ∇f denote the $n \times n$ matrix whose entry at position (i,j) is the partial derivative $\frac{\partial f}{\partial x_{i,j}}$, which we also denote by $\partial_{i,j}(f)$. Note that $\nabla \det_n$ is the cofactor matrix of X_n, and its transpose $(\nabla \det_n)^{\mathsf{T}}$ is by definition the so-called adjugate matrix of X_n.

Our bivariate Cayley–Hamilton theorem 5.1 states that for all n, d we have

$$(\nabla \chi_{n,d+1})^{\mathsf{T}} \;=\; \sum_{i=0}^{d} (-1)^i \, \chi_{n,d-i} \, X_n^i. \qquad \text{(Thm. 5.1)}$$

Our theorem has several immediate corollaries, in particular the classical Cayley–Hamilton theorem, the Girard–Newton identities, and $w(\chi(n,d)) \in \mathsf{poly}(n,d)$, as depicted in Figure 1. The proofs of all of the corollaries are very short, and we prove them all in §2.

We study $(\nabla \chi_{n,d})^{\mathsf{T}}$ more closely in §6. As our second main result, we construct an ABP for computing all $\chi_{n,d}$. All intermediate computations results of our ABP are entries of $(\nabla \chi_{n,d})^{\mathsf{T}}$. Our construction is one third the size and width of the current smallest ABP [MV97b], see §6. This is a modest improvement, but it marks the first improvement on this problem for 28 years.

[1] In our definition of VBP, the sequences are indexed by the degree and not by the number of variables, which gives a cleaner theory, see [vdBDG+25]. This makes no difference for the fundamental complexity theoretic considerations.

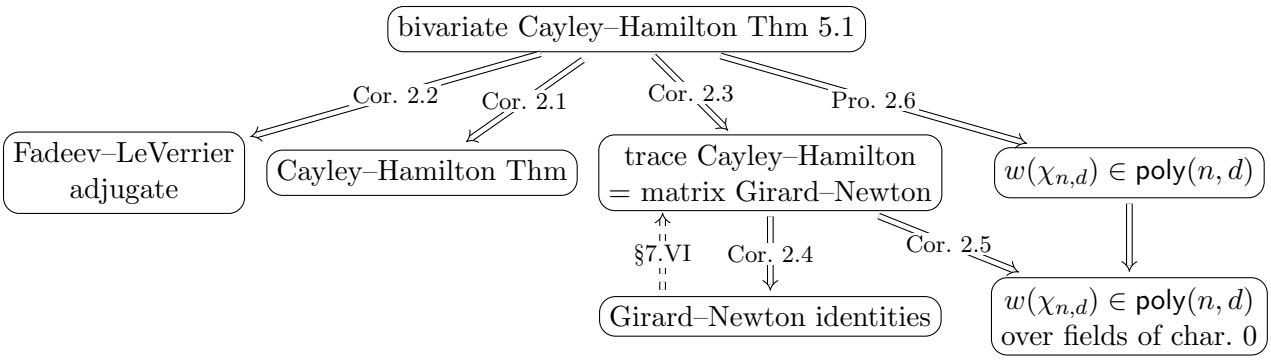


Figure 1: Direct implications between the theorems. The dashed implication only works over algebraically closed fields: one gets the matrix Girard–Newton identities from the classical Girard–Newton identities, and from there one derives $w(\chi_{n,d}) \in \mathsf{poly}(n,d)$, but only over fields of characteristic zero. To obtain $w(\chi_{n,d}) \in \mathsf{poly}(n,d)$ over arbitrary commutative rings, one either uses our bivariate Cayley–Hamilton theorem, or one has to use an entirely separate construction such as [Tod92, Val92, MV97b], see §3.

Our ABP only ever computes the entries of $(\nabla\chi_{n,d})^{\mathsf{T}}$ and nothing else, and does not use any ad-hoc combinatorial constructions. In particular, it does not involve the concept of clow sequences, which are combinatorial gadgets that have no known nice algebraic interpretation, see §4.I. Therefore, our construction answers the question raised in §4 of [MV99] about the necessity of clow sequences for these types of algorithms in the negative. This answer to this 26-year-old open question is our third main result.

Some of our results are algebraic in nature, and our algorithmic results follow directly from algebraic identities. We prove all results via the combinatorial approach to matrix algebra pioneered by [Foa65, CF69, Str83, Zei85], see many more references in [Zei85, MV99]. The construction in [MV97b] is combinatorial, but not in exactly the same way as [Str83, Zei85], who give combinatorial proofs for algebraic identities between inherently algebraic objects, while [MV97b] work with the combinatorial gadget of clow sequences. Our proofs are in the spirit of [Str83, Zei85], since we give combinatorial proofs of algebraic identities between objects from algebra. In our case these are the first order partial derivatives of $\chi_{n,d}$, which appear in the bivariate Cayley–Hamilton theorem.

2 Corollaries

In this section we prove the corollaries of Theorem 5.1 illustrated in Figure 1. For a matrix X we write $[X]_{i,j}$ for the entry in row i and column j.

2.1 Corollary (The Cayley–Hamilton theorem)**.**

$$0 \;=\; \sum_{i=0}^{n} (-1)^i \, \chi_{n,n-i} \, X_n^i. \qquad \diamond$$

Proof. Set $n = d$ in Theorem 5.1. Since $\chi_{n,d} = 0$ for $d > n$, it follows that for $n = d$ the left-hand side of Theorem 5.1 vanishes, which finishes the proof. □

The Fadeev–LeVerrier algorithm computes the adjugate matrix along the way. We will not discuss this algorithm here, because it uses divisions and hence does not work over arbitrary commutative rings, but the following formula is used to prove that it computes the adjugate, see [Gan59, IV(46)].

2.2 Corollary (The adjugate matrix in the Fadeev–LeVerrier algorithm)**.**

$$(\nabla \det_n)^{\mathsf{T}} \;=\; \sum_{i=0}^{n-1} (-1)^i \, \chi_{n,n-1-i} \, X_n^i. \qquad \diamond$$

Proof. We set $d = n-1$ in Theorem 5.1. □

2.3 Corollary (The trace Cayley–Hamilton theorem, also known as the matrix Girard–Newton identities)**.**

$$-d \cdot \chi_{n,d} \;=\; \sum_{i=1}^{d} (-1)^i \, \chi_{n,d-i} \, \operatorname{tr}(X_n^i). \qquad \diamond$$

Proof. Note that by (1.2) we have

$$\partial_{a,a}(\chi_{n,d+1}) \;=\; \sum_{a \notin S \in \binom{\{1,\dots,n\}}{d}} \det[X_n]_{S,S}.$$

Since for every $S \in \binom{\{1,\dots,n\}}{d}$ there are $n-d$ many $a \in \{1,\dots,n\} \setminus S$, it follows that $\operatorname{tr}(\nabla \chi_{n,d+1}) = (n-d)\chi_{n,d}$. Hence, taking the trace in Theorem 5.1 on both sides gives $(n-d)\chi_{n,d} = \sum_{i=0}^{d} (-1)^i \chi_{n,d-i} \operatorname{tr}(X_n^i)$. We conclude the proof by subtracting $n\,\chi_{n,d}$ on both sides of the equation. □

2.4 Corollary (The Girard–Newton identities)**.** *Let $e_{n,d} = \chi_{n,d}(\operatorname{diag}(x_1,\dots,x_n))$ be the elementary symmetric polynomial ($e_{n,0} = 1$). Let $p_{n,d} := x_1^d + \cdots + x_n^d$ be the power sum polynomial ($p_{n,0} = n$). We have*

$$-d \cdot e_{n,d} \;=\; \sum_{i=1}^{d} (-1)^i \, e_{n,d-i} \, p_{n,i}. \qquad \diamond$$

Proof. We restrict Corollary 2.3 to diagonal matrices $D = \operatorname{diag}(x_1,\dots,x_n)$ and immediately obtain the desired statement, because $\chi_{n,d}(D) = e_{n,d}$ and $\operatorname{tr}(D^d) = p_{n,d}$. □

Recall from §1 that an ABP G computes $\sum_{s\text{-}t\text{-path } p \text{ in } G} \beta(p)$, where $\beta(p) = \prod_{e \in p} \beta(e)$, and $\beta(e)$ is the edge label of the edge e. A sub-ABP G_v of G at a vertex v is defined as the ABP that consists of all edges on all paths from s to v, where v is the sink of G_v, and s is the source of G_v. A polynomial f is defined to be computed by an ABP G *along the way* if there exists a vertex v in G such that G_v computes f. We also say that G computes f at v. Sub-ABPs can also have sources other than s, and we see those in the constructions of Corollary 2.5 and Proposition 2.6.

2.5 Corollary. *Over fields of characteristic zero, $w(\chi_{n,d}) \in \mathsf{poly}(n,d)$ follows directly from the trace Cayley–Hamilton theorem (Corollary 2.3).* $\diamond$

Proof. We construct an ABP without constant edges. The ABP will compute $\det_n = \chi_{n,n}$, and it will also compute each $\chi_{n,d}$, $0 \le d \le n$ along the way, at a vertex that we call v_d. See Figure 2 for an illustration. Starting with the vertices $v_0,\dots,v_n$, we connect each v_d with all v_{d-i}, $1 \le i \le d \le n$, by ABPs $G_{d-i,d}$ that compute $\frac{(-1)^i}{-d} \operatorname{tr}(X_n^i)$ via identifying the source of $G_{d-i,d}$ with v_{d-i} and the sink of $G_{d-i,d}$ with v_d. Note that each connection maintains the property that at each vertex layer j the ABP computes homogeneous degree j polynomials. The resulting program has $\binom{n+1}{2}$ such connections, each of width $w(\operatorname{tr}(X_n^i)) \le n^2$. By Corollary 2.3, the ABP computes $\chi_{n,d}$ at each vertex v_d. Hence, $w(\chi_{n,d}) \le \binom{n+1}{2} \cdot n^2$ over fields of characteristic 0. □

In order to get the same result over arbitrary commutative rings, we use Theorem 5.1 instead of Corollary 2.3 and make a bivariate argument, see the following proposition.

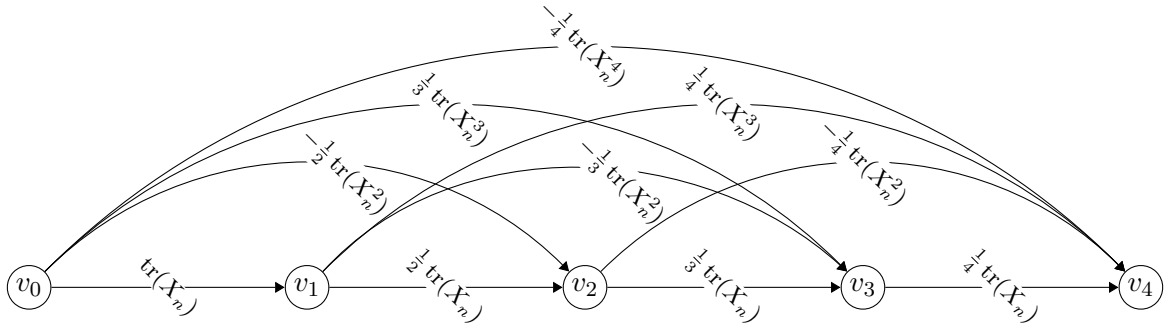


Figure 2: Over fields of characteristic 0. An ABP computing along the way all $\chi_{n,d}$ for $d \leq 4$, where any $n \in \mathbb{N}$ is fixed. Each edge represents a sub-ABP whose source is some v_i and whose sink is some v_j, $i < j$.

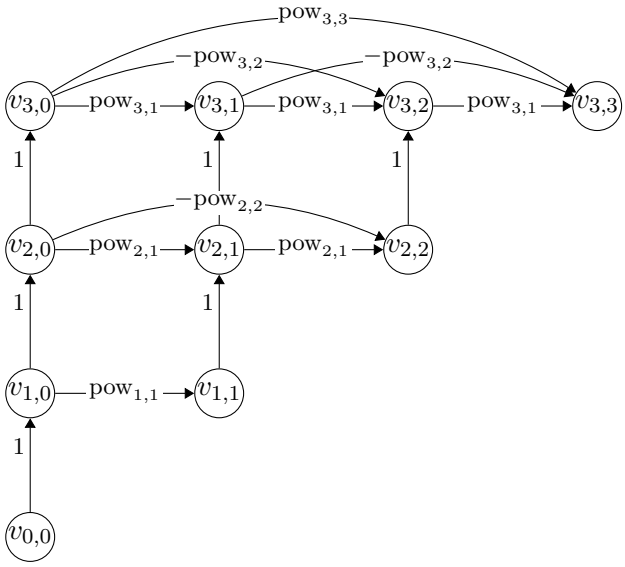


Figure 3: An ABP computing along the way all $\chi_{n,d}$ for $d \leq n \leq 3$. Each non-constant edge represents a sub-ABP.

2.6 Proposition. *Over arbitrary commutative rings, we have* $w(\chi_{n,d}) \in \mathsf{poly}(n, d)$. ◇

Proof. Let $\mathrm{pow}_{n,i} := [X_n^i]_{n,n}$ be the bottom-right entry of the i-th matrix power of X_n, and note $w(\mathrm{pow}_{n,i}) \leq n$. From (1.2) it follows that $[\nabla\chi_{n,d+1}]_{n,n} = \chi_{n-1,d}$. We take the bottom-right matrix entry (i.e., position (n, n)) on both sides of Theorem 5.1:

$$\chi_{n-1,d} \;=\; \sum_{i=0}^{d} (-1)^i \, \chi_{n,d-i} \, \mathrm{pow}_{n,i}. \tag{2.7}$$

We use that $\mathrm{pow}_{n,0} = 1$ to rewrite equation (2.7) in the form

$$\chi_{n,d} \;=\; \chi_{n-1,d} + \sum_{i=1}^{d} (-1)^{i+1} \, \chi_{n,d-i} \, \mathrm{pow}_{n,i}. \tag{2.8}$$

Equation (2.8) gives a recipe to construct an ABP, illustrated in Figure 3. Our ABP will compute $\chi_{n,d}$ at vertex $v_{n,d}$. We start with a sequence of vertices $v_{i,0}$, $0 \leq i \leq n$, with edges labeled with

1 from each $v_{i,0}$ to $v_{i+1,0}$. The source is $v_{0,0}$. Equation (2.8) allows us to take a program that computes $\chi_{n-1,d}$ and each $\chi_{n,d-i}$, $i \in \{1,\dots,d\}$, and enlarge it to a program that also computes $\chi_{n,d}$ at a vertex that we call $v_{n,d}$, as follows. We add an edge with label 1 from $v_{n-1,d}$ to $v_{n,d}$, provided $n > d$ (otherwise, $\chi_{n-1,d} = 0$). Then, for each $i \in \{1,\dots,d\}$ we connect $v_{n,d-i}$ and $v_{n,d}$ via an ABP that computes $(-1)^{i+1}\operatorname{pow}_{n,i}$, identifying sources and sinks as in the proof of Corollary 2.5. We enlarge the program via this construction several times in the following order of pairs (n,d) that guarantees that all required vertices are present at each step of the construction: $(1,1),(2,1),\dots,(n,1),(1,2),(2,2),\dots,(n,2),(3,1),\dots,\dots,(n,d)$. The final ABP consists of $O(nd)$ many named vertices $v_{i,j}$. Every $v_{i,j}$ has exactly j many $\operatorname{pow}_{i,j'}$, $1 \le j' \le j$, ending at it, which gives a total of $O(nd^2)$ many $\operatorname{pow}_{i,j'}$. We have $w(\operatorname{pow}_{i,j'}) \le n$ in every case, which results in $w(\chi_{n,d}) \in O(n^2d^2)$. □

Figure 3 is the first time over arbitrary commutative rings that a polynomially sized ABP for $\det_d$ can be easily depicted. Its width is $O(n^2d^2)$, which is surprisingly small given the fact that it is a direct application of Theorem 5.1, which is first and foremost a theorem in linear algebra, not in algorithms. The algorithmic and algebraic questions appear to be deeply interlinked here.

3 Related work

Girard–Newton, Cayley–Hamilton, and combinatorial proofs in algebra The Girard–Newton identities have numerous applications in constructions in algebraic complexity theory. For example, recently the Girard–Newton identities have been applied in two breakthrough results in algebraic complexity theory [LST25, AW24], and they are the center of investigation in recent papers, both in algebraic complexity theory [FLST24] and algebra [dV25].

The Cayley–Hamilton theorem is a fundamental theorem in linear algebra with countless applications and generalizations. [SC04] studies the Cayley–Hamilton theorem from a proof complexity perspective. In algebraic complexity theory the Cayley–Hamilton theorem can be used to explain the symmetries of Strassen's fast matrix multiplication algorithm [IL19]. [Gri25] first discusses the long history of the trace Cayley–Hamilton theorem, the different proof difficulties for the cases $d \ge n$ and $d < n$, and the available proofs over fields. [Gri25] then states a lemma (Lemma 3.11(c) therein, already implicit in [Buc84, §3]) over arbitrary commutative rings that he then applies in the proofs of both the trace Cayley–Hamilton theorem and the Cayley–Hamilton theorem. The Cayley–Hamilton theorem follows directly from this lemma, but more work is required to reach the trace Cayley–Hamilton theorem. Before our paper, [Gri25] was the most unified view on the subject. The formula on the right-hand side of Theorem 5.1 appears in earlier works, see for example [Gan59, IV(46)] or [Gri25, Lemma 3.11(c)], but the central importance of the matrix on the left-hand side was overlooked and it was left underanalysed with only some of its properties proved, for example its trace. The matrix on the left-hand side was only known in the special case of $d = n$ (Corollary 2.1) and $d = n - 1$ (Corollary 2.2).

Subtleties in proving the Girard–Newton identities are explained in [Min03]. A short combinatorial proof of the Girard–Newton identities is presented in [Zei84], and a combinatorial proof of the Cayley–Hamilton theorem is presented in [Rut64, Str83]. Numerous other algebraic results have been proved combinatorially, for example the MacMahon master theorem [Foa65], the matrix-tree theorem [Orl78, Tem81, Cha82, Min97], the Jacobi identity [Foa80], Vandermonde's determinant identity [Ges79], and others. These results are listed in [MV99] and surveyed in [Zei85], [SW12, Ch. 4].

ABPs, GapL, and clow sequences There are many ways for computing $\det_d$ efficiently, but not all of them directly give a proof that $w(\det_d) \in \mathsf{poly}(d)$ over arbitrary commutative rings.

[Coo85] noted that the Samuelson–Berkowitz algorithm [Sam42, Ber84] can be used to obtain $\det_d$ efficiently from matrix powers, but this does not directly prove $w(\det_d) \in \mathsf{poly}(d)$. It is stated in [Tod91] that this observation in [Coo85] can be converted into a proof that $w(\det_d) \in \mathsf{poly}(d)$. A proof is provided in [Tod92]. This is also proved in [Val92, §3]. These proofs also directly give $w(\chi_{n,d}) \in \mathsf{poly}(n,d)$. Another proof is provided in [MV97b, MV97a], which uses a combinatorial construction and gives the smallest ABP known so far. All these constructions have natural Boolean counterparts, and they can be used to show that the integer determinant computation is complete for the counting complexity class GapL, see [AO96]. This is a surprising fact, because GapL is not defined in an algebraic way. The class GapL consists of all functions $f : \{0,1\}^* \to \mathbb{Z}$ for which there exist two nondeterministic logspace Turing machines M_+ and M_- such that $\forall w \in \{0,1\}^*$ we have $f(w) = \#\mathrm{acc}_{M_+}(w) - \#\mathrm{acc}_{M_-}(w)$, where $\#\mathrm{acc}_M(w)$ is the number of accepting computation paths of M on input w.

See [Rot01, §2.8] for a discussion of the history and attribution of the Samuelson–Berkowitz algorithm.

[MV97b] prove $w(\chi_{n,d}) \in \mathsf{poly}(n,d)$ via the cancellation properties of so-called *clow sequences*, see §4.I below. Clow sequences and their cancellations have also been successfully used in [ARZ99]. [Sol02] finds that the Samuelson–Berkowitz algorithm can be described naturally in terms of clow sequences. [MV99] analyze several algorithms [Sam42, Csa75, Chi85, MV97b] and their internal use of clow sequences and come to the conclusion that all published efficient algorithms for $\chi_{n,d}$ that are based on signed weighted sums of partial cycle covers use the concept of clow sequences or a generalization of clow sequences. The ABP that we construct in Theorem 6.1 computes only first order partial derivatives of partial cycle covers, and does not involve the more general concept of clow sequences (partial cycle covers are clow sequences, but not vice versa). Hence, our construction answers the question raised in [MV99, §4] about the necessity of the use of clow sequences of such algorithms in the negative, see also §4.I.

Structure, homogenization, and geometric complexity theory The width $w(f)$ can be defined as the smallest n such that f is restriction of the iterated matrix multiplication polynomial

$$\mathrm{IMM}_{n,d} \;=\; \sum_{1 \le i_1,\ldots,i_{d-1} \le n} y^{(1)}_{n,i_1}\, y^{(2)}_{i_1,i_2}\, y^{(3)}_{i_2,i_3} \cdots y^{(d-1)}_{i_{d-2},i_{d-1}}\, y^{(d)}_{i_{d-1},n},$$

where a restriction is defined as a map that sends variables to linear combinations of variables. The polynomial $\mathrm{IMM}_{n,d}$ has been recently the object of study in the breakthrough result [LST25]. The high amount of structure of $\mathrm{IMM}_{n,d}$ makes it not only possible to efficiently sample random polynomials of fixed width and degree, but also to analyze their probabilistic properties and obtain algorithms for solving systems of random polynomial equations of given width [BCL23], which is a solution to Smale's 17th problem.

Early papers in geometric complexity theory (GCT) use the determinantal complexity dc as their main complexity measure, and (1.4) indicates that width and dc should work equally well from a computational complexity perspective, but since the determinant is a single-index polynomial, it is necessary to either discard the idea that (border) complexity $\le n$ polynomials form a GL orbit closure, or one must introduce *padding*, see [MS01, MS08, BLMW11]. However, padding has serious effects on the representation theoretic multiplicities [KL14], which led to main GCT conjectures being disproved in [IP17, BIP19]. These proofs rely heavily on the padding, and if the width is used instead of dc, then no such negative results are known. [DGI$^+$24] introduce the noncommutative elementary symmetric polynomial evaluated at 3×3 variable matrices as a simpler replacement for $\mathrm{IMM}_{n,d}$, which makes the study of small finite cases easier. This polynomial is also double-indexed, but it is not characterized by its stabilizer. That polynomial was recently used in [FLST24] to study the homogenization of arithmetic formulas.

4 Preliminaries

For a function $\zeta : \mathbb{N}\times\mathbb{N} \to \mathbb{N}$ we write $\zeta \in \mathsf{poly}(n, d)$ if there exists a bivariate polynomial $p \in \mathbb{Z}[n, d]$ such that for all $n \geq 1$ and $d \geq 1$ we have $\zeta(n, d) \leq p(n, d)$.

All graphs in this paper are directed graphs, or *digraphs* for short, without multi-edges, but we allow loops (i.e., edges from a vertex to itself). Formally, a *digraph* G consists of a *vertex set* $V(G)$ and an *edge set* $E(G) \subseteq V(G) \times V(G)$. A *walk* q of *length* $l \geq 0$ in G is a sequence of vertices $(v_0, v_1, \ldots, v_l)$ for which $\forall i \in \{1, \ldots, l\} : (v_{i-1}, v_i) \in E(G)$. The edge set of a walk is defined as $\{(v_{i-1}, v_i) \mid 1 \leq i \leq l\}$. The sign $\mathrm{sgn}(q)$ is defined as $(-1)^l$. For $s, t \in V(G)$ let $W_l(G, s, t)$ denote the set of length l walks in G from vertex $s = v_0$ to vertex $t = v_l$. A *path* p is a walk in which all v_i are pairwise distinct. In particular, a path never uses a loop. A path can consist of a single vertex. The vertex v_0 is called the tail of p, and v_l is called the head of p. Let $P(G, s, t)$ denote the set of paths in G from vertex s to vertex t. Let $P_l(G, s, t) \subseteq P(G, s, t)$ denote the subset of length l paths. A *cycle* is a set of edges that is obtained from the edge set of a path p by adding the edge from the head of p to its tail. In particular, a loop is a cycle. The length of a cycle is defined as its number of edges. The sign of a cycle c is defined as the sign of a path that is obtained by removing an edge from c. We write $((v_0, v_1, \ldots, v_l))$ for the cycle that has the edge set $\{(v_0, v_1), (v_1, v_2), \ldots, (v_{l-1}, v_l), (v_l, v_0)\}$, and we write $((v_0))$ for the cycle formed by the loop at vertex v_0. For a walk q in G and $e \in E(G)$ we write $e \in q$ if e is an edge in q. Similarly, for $v \in V(G)$, we write $v \in q$ if v is a vertex of q.

We will mostly consider *labeled digraphs*, which are pairs (G, β_G) of a digraph G and a label function $\beta_G : E(G) \to R[\mathbf{x}]$. For an edge $e \in E(G)$ of a labeled digraph (G, β_G), we define the weight $\beta(e) := \beta_G(e)$. We usually do not give an explicit name to the label function, and we write "G is a labeled digraph", and we access the weight via $\beta(e)$. For a walk or cycle q on a labeled digraph G, define the weight $\beta(q)$ of q as

$$\beta(q) \;:=\; \textstyle\prod_{e\in q} \beta(e).$$

Labeled digraphs are drawn with their labels on the edges, where edges with label 0 are usually not drawn.

In a digraph G, a set of pairwise vertex-disjoint cycles is called a *partial cycle cover*. The sign of a partial cycle cover is defined as the product of the signs of its cycles (1 if it is the empty partial cycle cover). The length $\ell(q)$ of a partial cycle cover q is the sum of the lengths of its cycles (0 if it is the empty partial cycle cover). Let $\mathrm{CC}(G)$ denote the set of partial cycle covers of G, and let $\mathrm{CC}_d(G)$ denote the set of partial cycle covers of G of length d. Let K_n denote the complete digraph on n vertices, i.e., the digraph with vertex set $\{v_1, \ldots, v_n\}$ and edge set $\{(v_i, v_j) \mid 1 \leq i, j \leq n\}$. We attach edge labels to K_n via $\beta(v_i, v_j) := x_{i,j}$, making K_n a labeled digraph. We use the short notation $\mathrm{CC} = \mathrm{CC}(K_n)$, $\mathrm{CC}_d = \mathrm{CC}_d(K_n)$, $P(v_a, v_b) = P(K_n, v_a, v_b)$, $W(v_a, v_b) = W(K_n, v_a, v_b)$, $V = V(K_n)$, and so on. We have

$$\det_d = \textstyle\sum_{q\in \mathrm{CC}_d(K_d)} \mathrm{sgn}(q)\, \beta(q). \tag{4.1}$$

By (1.2) it follows

$$\chi_{n,d} = \textstyle\sum_{q\in \mathrm{CC}_d} \mathrm{sgn}(q)\, \beta(q). \tag{4.2}$$

The set $\mathrm{CCP}_d(v_a, v_b)$ consists of all pairs (q, z) for which $z \in P(v_a, v_b)$, $q \in \mathrm{CC}$, $\ell(z) + \ell(q) = d$, and z and q are vertex-disjoint. CCP stands for (partial) cycle cover and path.

4.I Clow sequences

The combinatorial concept of a clow sequence is defined in [MV97b] as follows, and it is a bit more technically involved than what we have introduced so far. We will not use this concept, but we list its definition here to make it apparent that our construction answers the question in [MV99, §4].

A clow (CLOsed Walk) is a walk $q = (v_0, \ldots, v_l)$ on G with $v_0 = v_l$ such that the smallest entry in $\{v_1, \ldots, v_l\}$ appears exactly once. This entry is called the head $h(q)$ of the clow. A clow sequence $(q_1, q_2, \ldots, q_k)$ is a sequence of clows such that $h(q_1) < h(q_2) < \cdots < h(q_k)$. As noted in [MV97b], a (partial) cycle cover is a clow sequence, but not vice versa. In our construction Theorem 6.1, all polynomials computed along the way in our ABP correspond to elements in some CCP. Since clow sequences can self-intersect, our construction only uses a small subset of the clow sequences. It is reasonable to assume that some existing proofs in the literature that rely on the combinatorics of clow sequences simplify when phrasing them via CCP instead, which are just partial cycle covers with one edge removed.

As elaborated in §3 in the context of GCT, having constructions without ad-hoc combinatorial gadgets can be preferable, especially when working with advanced tools from algebra and representation theory. Our construction in Theorem 6.1 uses only the first order partial derivatives of the $\chi_{n,d}$, which is an inherently algebraic notion. No inherently algebraic notion is known to correspond to clow sequences.

5 The Bivariate Cayley–Hamilton Theorem

In this section we prove the bivariate Cayley–Hamilton theorem. The proof is a combination of the combinatorial proof of the Cayley–Hamilton theorem in [Str83] and the combinatorial proof of the Girard–Newton identities in [Zei84].

5.1 Theorem (Bivariate Cayley–Hamilton Theorem)**.** *For all $n, d \in \mathbb{N}$ we have*

$$(\nabla \chi_{n,d+1})^{\mathsf{T}} \;=\; \sum_{i=0}^{d} (-1)^i \, \chi_{n,d-i} \, X_n^i. \qquad \diamond$$

Proof. We prove this matrix equation for each position (a, b) independently, so fix $1 \le a, b \le n$ and consider the entry at row a and column b. We write $[X]_{a,b}$ for the entry of the matrix X at position (a, b).

For the left-hand side of the claimed identity, by (4.2) we have

$$[(\nabla \chi_{n,d+1})^{\mathsf{T}}]_{a,b} \;=\; [\nabla \chi_{n,d+1}]_{b,a} \;=\; \sum_{\substack{r \in \mathrm{CC}_{d+1} \\ (v_b, v_a) \in r}} \mathrm{sgn}(r) \, \frac{\beta(r)}{x_{b,a}} \tag{5.2}$$

(the division here and all divisions in this paper are just for notational convenience, not actual algorithmic division operations).

For the right-hand side of the claimed identity, for every $i \in \{0, \ldots, d\}$ we have

$$[(-1)^i X_n^i]_{a,b} = \sum_{z \in W_i(v_a, v_b)} \mathrm{sgn}(z) \, \beta(z).$$

Therefore,

$$\begin{aligned} \sum_{i=0}^{d} \chi_{n,d-i} \cdot [(-1)^i X_n^i]_{a,b} \;&=\; \sum_{i=0}^{d} \sum_{\substack{q \in \mathrm{CC}_{d-i} \\ z \in W_i(v_a, v_b)}} \mathrm{sgn}(q) \, \beta(q) \; \mathrm{sgn}(z) \, \beta(z) \\ &=\; \sum_{\substack{q \in \mathrm{CC} \\ z \in W(v_a, v_b) \\ \ell(q) + \ell(z) = d}} \mathrm{sgn}(q) \, \beta(q) \; \mathrm{sgn}(z) \, \beta(z). \end{aligned} \tag{5.3}$$

We now prove that the right-hand side of (5.3) equals the right-hand side of (5.2), which finishes the proof. We first handle all summands in the right-hand side of (5.2) and show how they appear

in the right-hand side of (5.3). We then show that the remaining summands in the right-hand side of (5.3) cancel out with each other.

For each $r \in \mathrm{CC}_{d+1}$ with $(v_b, v_a) \in r$, we construct a pair $(q, z) \in \mathrm{CCP}_d(v_a, v_b)$ by removing the edge (v_b, v_a) from r (note that if $a = b$, then the loop (v_a, v_a) is removed, and hence z is the length 0 path from v_a to v_a). We observe that $\operatorname{sgn}(r)\beta(r)/x_{b,a} = \operatorname{sgn}(q)\beta(q)\operatorname{sgn}(z)\beta(z)$. Every pair $(q, z) \in \mathrm{CCP}_d(v_a, v_b)$ is obtained from some r in this way as follows. From (q, z) we construct r by adding back the edge (v_b, v_a).

It remains to show that all other summands in the right-hand side of (5.3) cancel out with each other. These remaining pairs (z, q), $q \in \mathrm{CC}$, $z \in W(v_a, v_b)$, $\ell(q) + \ell(z) = d$, are exactly those for which

- either z uses a vertex at least twice (i.e., $z \in W(v_a, v_b) \setminus P(v_a, v_b)$),
- or z and q are not vertex-disjoint,

or both. Let X denote the set of these pairs (z, q). We will decompose X into a disjoint union $X = \dot{\bigcup}_{v \in V}(S_v \,\dot\cup\, T_v)$ of $2n$ many sets S_v and T_v. For each v we will give a weight-preserving sign-inverting bijection between S_v and T_v. This then finishes the proof.

Every $(z, q) \in X$ belongs to exactly one set S_v or T_v according to the following rule. We start at the vertex v_a and traverse the walk $z \in W(v_a, v_b)$ until we either use a vertex v for the second time (this means that $(z, q) \in S_v$) or z uses a vertex v of q (this means that $(z, q) \in T_v$), at which time we stop the traversal. Note that each (z, q) is contained in a unique set: if a vertex v that is a vertex of q is used a second time by z, then this vertex of q would have been detected in the traversal as a vertex of q before visiting it a second time.

For $(z, q) \in S_v$, we see that z starts with the vertices of a path p from v_a to v, then continues with the vertices of a cycle c from v to v, and then z continues with the vertices of a walk t from v to v_b. Note that q and c are vertex-disjoint, because otherwise the traversal of z would have found their common vertex before reaching v a second time. We define z' as a modification of z as follows. The walk z' starts with p and then directly continues with t without taking the cycle c. We define q' to be the union of q and c. The situation is illustrated in Figure 4. We observe that $(z', q') \in T_v$ by construction. Note $\ell(q') + \ell(z') = d$. We have $\operatorname{sgn}(q) = (-1)^{\ell(c)-1}\operatorname{sgn}(q')$ and $\operatorname{sgn}(z) = (-1)^{\ell(c)}\operatorname{sgn}(z')$, hence $\operatorname{sgn}(q)\operatorname{sgn}(z) = -\operatorname{sgn}(q')\operatorname{sgn}(z')$ as desired.

(a) $(z, q) \in S_{v_2}$ with $z = (v_1, v_2, v_3, v_2, v_4, v_4)$ and $q = \{((v_4, v_6))\}$. We have $p = (v_1, v_2)$, $c = ((v_2, v_3))$, and $t = (v_2, v_4, v_4)$.

(b) $(z', q') \in T_{v_2}$ with $z' = (v_1, v_2, v_4, v_4)$ and $q' = \{((v_2, v_3)), ((v_4, v_6))\}$. We have $p = (v_1, v_2)$, $c = ((v_2, v_3))$, and $t = (v_2, v_4, v_4)$.

Figure 4: The situation in the proof of Theorem 5.1. The left and right subfigure are partner summands in the sign-inverting involution. The figures are adaptions of figures in [Str83].

For $(z', q') \in T_v$, the walk z' starts with the vertices of a path p from v_a to v, where v is part of a cycle c of q', and then z' continues with the vertices of a walk t from v to v_b. We define z to start with the vertices of the path p, then continue with the vertices of the cycle c, then continue with the walk t. We define q to be q' with c removed. The situation is illustrated in Figure 4. We observe that $(z, q) \in S_v$ by construction. Note $\ell(q) + \ell(z) = d$. We have $\operatorname{sgn}(q) = (-1)^{\ell(c)-1}\operatorname{sgn}(q')$ and $\operatorname{sgn}(z) = (-1)^{\ell(c)}\operatorname{sgn}(z')$, hence $\operatorname{sgn}(q)\operatorname{sgn}(z) = -\operatorname{sgn}(q')\operatorname{sgn}(z')$ as desired. □

6 The entries of the gradient

Theorem 5.1 highlights the importance of the gradient transpose, $(\nabla\chi_{n,d})^{\mathsf{T}}$. We analyze the entries of $(\nabla\chi_{n,d})^{\mathsf{T}}$ more closely in this section.

For a finite set of polynomials S we define $w(S)$ to be the width of an ABP that computes every element of S along the way. Such an ABP is allowed to have several sinks in vertex layer d. The number of vertices for the ABP computing $\{\chi_{n,d} \mid d \le n\}$ in [MV97b, MV97a] (excluding source and sinks) is n^3, and the width is n^2 (see the discussion about pruning in [MV97b, §5]). With a similar technique as in Proposition 2.6 and a bit more care we prove the following Theorem 6.1, which improves the leading coefficients of the size by a factor of 3 and width by a factor of 2 compared to [MV97b]. Instead of clow sequences, we use cycle-cover-and-paths (CCP).

6.1 Theorem. $w(\{\chi_{n',d'} \mid n' \le n,\, d' \le d\}) \le \frac{n^2}{2} + \text{l.o.t.}$ *The total number of vertices in this construction is* $\frac{dn^2}{2} - \frac{d^3}{6} + \text{l.o.t.}$*, which for* $d = n$ *gives* $\frac{n^3}{3} + \text{l.o.t.}$ ◇

Proof. Let $r_{n,d}$ be the last row of $(\nabla\chi_{n,d+1})^{\mathsf{T}}$. We decompose the $n\times n$ matrix X_n into an $n\times(n-1)$ matrix L_n and an $n\times 1$ matrix C_n, i.e., $X_n = (L_n|C_n)$. The following block matrix identity is the main ingredient to build an ABP.

$$r_{n,d} = \Big(- r_{n,d-1}L_n \,\Big|\, \sum_{i=d}^{n-1} r_{i,d-1}C_i \Big). \tag{6.2}$$

We now prove (6.2), similarly to Theorem 5.1. For a vector r, let $[r]_a$ denote the a-th entry of r.

$$[r_{n,d}]_a \;=\; \sum_{\substack{q\in \mathrm{CC}_{d+1}\\ (v_a,v_n)\in q}} \operatorname{sgn}(q)\frac{\beta(q)}{x_{a,n}} \;=\; \sum_{(q,z)\in \mathrm{CCP}_d(v_n,v_a)} \operatorname{sgn}(q)\,\beta(q)\,\operatorname{sgn}(z)\,\beta(z)$$

In particular,

$$[r_{n,d}]_n = \chi_{n-1,d} \tag{6.3}$$

and

$$r_{i,d-1}C_i \;=\; \sum_{a=1}^{i} \sum_{\substack{q\in \mathrm{CC}_d(K_i)\\ (v_a,v_i)\in q}} \operatorname{sgn}(q)\frac{\beta(q)}{x_{a,i}}\,x_{a,i} \;=\; \sum_{\substack{q\in \mathrm{CC}_d(K_i)\\ v_i\in q}} \operatorname{sgn}(q)\,\beta(q). \tag{6.4}$$

For $1 \le a < n$ we have

$$\begin{aligned}
[r_{n,d-1}L_n]_a \;&=\; \sum_{b=1}^{n} \sum_{(q,z)\in \mathrm{CCP}_{d-1}(v_n,v_b)} \operatorname{sgn}(q)\,\beta(q)\,\operatorname{sgn}(z)\,\beta(z)\,x_{b,a}\\
&\overset{(*)}{=}\; \sum_{b=1}^{n} \sum_{\substack{(q,z)\in \mathrm{CCP}_{d-1}(v_n,v_b)\\ v_a\notin q,\ v_a\notin z}} \operatorname{sgn}(q)\,\beta(q)\,\operatorname{sgn}(z)\,\beta(z)\,x_{b,a}\\
&=\; - \sum_{(q,z')\in \mathrm{CCP}_d(v_n,v_a)} \operatorname{sgn}(q)\,\beta(q)\,\operatorname{sgn}(z')\,\beta(z') \;=\; -[r_{n,d}]_a.
\end{aligned} \tag{6.5}$$

Since q and z in the sums are vertex-disjoint, to see $(*)$, it suffices to prove

$$\sum_{b=1}^{n} \sum_{\substack{(q,z)\in \mathrm{CCP}_{d-1}(v_n,v_b)\\ v_a\in q}} \mathrm{sgn}(q)\,\beta(q)\;\mathrm{sgn}(z)\,\beta(z)\,x_{b,a}$$

$$= \; -\sum_{b'=1}^{n} \sum_{\substack{(q',z')\in \mathrm{CCP}_{d-1}(v_n,v_{b'})\\ v_a\in z'}} \mathrm{sgn}(q')\,\beta(q')\;\mathrm{sgn}(z')\,\beta(z')\,x_{b',a}.$$

This is proved analogously to the proof of Theorem 5.1 via a weight-preserving sign-inverting bijection between the summands on the left-hand side and on the right-hand side. Let $T_a = \{(q,z) \in \mathrm{CCP}_{d-1}(v_n,.) \mid v_a \in q\}$, and let $S_a = \{(q',z') \in \mathrm{CCP}_{d-1}(v_n,.) \mid v_a \in z'\}$ where $\mathrm{CCP}_{d-1}(v_n,.) = \bigcup_{j\in\{1,\dots,n\}} \mathrm{CCP}_{d-1}(v_n,v_j)$. It will be important that $a \neq n$ by assumption.

For $(q,z) \in T_a$, let v_b be the head of z. We traverse the edges of z from v_n to v_b (these can be zero edges), then the edge (v_b,v_a), and then the cycle c in q that contains v_a (this can be a loop). This is a so-called directed tadpole graph, see Figure 5. We construct q' by removing the cycle c

(a) $n = 6$, $a = 4$, $b = 5$, $b' = 3$, $z = (v_6, v_5)$, $q = \{((v_1,v_2)),((v_3,v_4))\}$. We have $c = ((v_3,v_4))$.

(b) $n = 6$, $a = 4$, $b = 5$, $b' = 3$, $z' = (v_6, v_5, v_4, v_3)$, $q' = \{((v_1,v_2))\}$. We have $c = ((v_3,v_4))$.

Figure 5: The situation in the proof of Theorem 6.1. The left and right subfigure are partner summands in the sign-inverting involution.

from q. We construct z' by starting with all edges in z, adding the edge (v_b,v_a) to z', and then traversing c from v_a until we reach the vertex $v_{b'}$ that satisfies $(v_{b'},v_a) \in c$, i.e., we add all but one of the edges of c to z'. By construction, the set of edges of z and q together with (v_b,v_a) equals the set of edges of z' and q' together with $(v_{b'},v_a)$. Since $v_a \in z'$, we have $(q',z') \in S_a$. The total number of edges stays the same, but the parity of the number of cycles changes in the process, which gives the desired sign change.

For $(q',z') \in S_a$, let $v_{b'}$ be the head of z'. We have that z' starts at v_n, z' contains v_a (note that $a \neq n$, hence v_a has a well-defined predecessor in z' that we call v_b), and ends at $v_{b'}$. The edges of z' from v_a to $v_{b'}$ together with the edge $(v_{b'},v_a)$ form a cycle c. We remove all these edges from z', and we also remove the edge (v_b,v_a), to obtain the path z with head v_b. We add the cycle c to q' to obtain q. By construction, the set of edges of z' and q' together with $(v_{b'},v_a)$ equals the set of edges of z and q together with (v_b,v_a). The total number of edges stays the same, but the parity of the number of cycles changes in the process, which gives the desired sign change. This proves $(*)$, and hence (6.5), which is the left side of the block matrix in (6.2).

We now prove the right side of the block matrix in (6.2). In the following, we group the partial cycle covers of length d in K_{n-1} into different sets according to the index i of their highest numbered vertex:

$$[r_{n,d}]_n \overset{(6.3)}{=} \chi_{n-1,d} \;=\; \sum_{i=d}^{n-1} \sum_{\substack{q\in \mathrm{CC}_d(K_i)\\ v_i\in q}} \mathrm{sgn}(q)\,\beta(q) \overset{(6.4)}{=} \sum_{i=d}^{n-1} r_{i,d-1} C_i. \tag{6.6}$$

This proves the right side of the block matrix in (6.2), so the proof of (6.2) is now complete.

Rewriting (6.6), we have

$$\chi_{n,d} = \sum_{i=d}^{n} r_{i,d-1} C_i. \tag{6.7}$$

We start by constructing an ABP that uses this identity and (6.2) to compute $\chi_{n,d}$. To construct an ABP that computes all $r_{i,d-1}$, $d \le i \le n$, we construct an ABP that computes all vectors $r_{i,j}$, $j < i \le n$, $1 \le j < d$ via (6.2). By (6.3) this computes all $\chi_{i',j}$, $j \le i' < n$, $1 \le j < d$ along the way. At degree j we compute each vector $r_{i,j}$, $i \in \{j+1, \dots, n\}$. Each such vector has i many elements, hence the total number of vertices in vertex layer j is $(j+1)+(j+2)+\cdots+n = \frac{(n-j)(n+j+1)}{2}$. This is maximal in vertex layer 1, where we have $\binom{n+1}{2}-1$ many vertices. Hence, $w(\chi_{n,d}) \le \binom{n+1}{2}-1$. The total number of vertices (excluding source and sink) is $\sum_{j=1}^{d-1} \frac{(n-j)(n+j+1)}{2} = (d-1)\binom{n+1}{2} - \binom{d+1}{3}$.

We modify this ABP to compute the $\chi_{n,j}$, $1 < j < d$, with one additional vertex each via (6.7). Moreover, we use (6.7) to also compute all $\chi_{i,d}$, $d \le i < n$, with one additional vertex each. All traces $\chi_{i,1}$, $1 \le i < n$, are computed along the way via (6.3). The trace $\chi_{n,1}$ can be computed with one additional vertex. □

For the sake of completeness, we now give the transition matrices for the construction in Theorem 6.1. We write (6.2) via block matrices:

$$\big(r_{d,d-1}, r_{d+1,d-1}, r_{d+2,d-1}, \dots, r_{n,d-1}\big) \overbrace{\begin{pmatrix} (0|C_d) & (0|C_d) & \cdots & (0|C_d) & (0|C_d) & (0|C_d) \\ (-L_{d+1}|0) & (0|C_{d+1}) & \cdots & (0|C_{d+1}) & (0|C_{d+1}) & (0|C_{d+1}) \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\ 0 & \ddots & \ddots & (0|C_{n-3}) & \vdots & \vdots \\ \vdots & 0 & 0 & (-L_{n-2}|0) & (0|C_{n-2}) & (0|C_{n-2}) \\ \vdots & \vdots & \vdots & 0 & (-L_{n-1}|0) & (0|C_{n-1}) \\ 0 & 0 & 0 & 0 & 0 & (-L_n|0) \end{pmatrix}}^{=:M_{n,d}} = \big(r_{d+1,d}, r_{d+2,d}, r_{d+3,d}, \dots, r_{n,d}\big).$$

It is sometimes interesting to have an ABP whose edges are only labeled by constants or scalar multiples of variables, see Proposition 7.5 below. Since these matrices have only variables or their negations (or zero) in each cell, and C_d (the adjacency matrix of the last edge layer) also has only variables in each cell, every edge in the ABP (besides the first edge layer) is either labeled with a variable $x_{i,j}$ or with $-x_{i,j}$. In the first edge layer, the edges are labeled with entries from the vectors $r_{n,1}$ for $2 \le n \le d$. We have

$$r_{n,1} \;=\; \big(-x_{n,1} \quad -x_{n,2} \quad \cdots \quad -x_{n,n-1} \quad \mathrm{tr}_{n-1}\big),$$

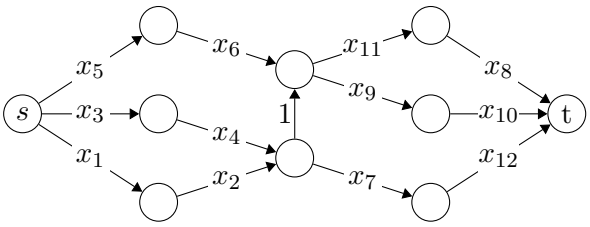
where $\mathrm{tr}_n := x_{1,1} + \cdots + x_{n,n}$. The final matrix product is

$$(r_{2,1}, r_{3,1}, \dots, r_{d,1}) M_{d,2} M_{d,3} \cdots M_{d,d-1} C_d = \det_d. \tag{6.8}$$

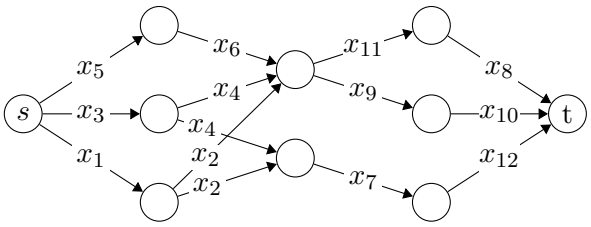
There is no need in the ABP to have edges labeled tr_1, tr_2, …, but we can instead label these edges with just $x_{1,1}$, $x_{2,2}$, …, and add edges with label 1 from each vertex that computes $x_{i,i}$ to the vertex that computes $x_{i+1,i+1}$, exactly as in Figure 2.

The elementary symmetric polynomial $e_{n,d}$ is the restriction of $\chi_{n,d}$ to diagonal matrices, and we have $w(e_{n,d}) \le n$. Theorem 6.1 gives $w(\chi_{n,d}) \in O(n^2)$ via rather sparse matrices. Moreover, the width of the construction is concentrated on the first layers. Therefore, the following question seems reasonable.

6.9 Open Problem. Is $w(\chi_{n,d}) \in O(n)$? ◇



(a) An ABP. Note that there is a constant edge labeled with "1". A corresponding pABP is on the right.



(b) An example pABP. There are no constant edges.

Figure 6: An ABP and a pABP, both computing the same $f \in R[\mathbf{x}]_d$. On the left: The ABP. On the right: The ABP after the constant edges are removed. The result is a pABP.

7 Appendix

The arguments in this appendix are either well-known or variants of well-known results. We included the appendix to make the paper more self-contained.

In §1 we defined a pABP as a d-tuple of $n \times n$ matrices of homogeneous linear polynomials whose matrix product's bottom right entry is the computation result. The width $w(f)$ is the smallest n so that f is such a computation result. Obviously, in the first matrix only the last row has any effect on the computation result, and in the last matrix only the last column has any effect on the computation result, so from now on we can assume that all other entries are zero. These pABPs are in bijective correspondence to ABPs without constant edges in the obvious way: The matrices specify the edge labels between the vertex layers. In this section we interpret pABPs in this way, and not as a tuple of matrices. This definition is for example used in [KS14, Def 59], under the name ABP.

7.I ABPs and pABPs give the same width

Recall from §1 the definition of an ABP G, which is a layered digraph labeled with homogeneous linear polynomials with single source and sink that computes

$$\sum_{p \in P(G,s,t)} \beta(p).$$

Recall that the width of an ABP is the minimum of the number of vertices in all layers but the first and last. The homogeneous width $w_{\mathrm{h}}(f)$ is defined as the smallest n such that f is computed by a width n ABP.

7.1 Proposition. *For every $f \in R[\mathbf{x}]_d$ we have $w(f) = w_{\mathrm{h}}(f)$.* ◇

Proof. Obviously, $w_{\mathrm{h}}(f) \leq w(f)$. The other direction requires a classical construction. Figure 6 gives an illustration. If we have edges in an ABP *within a vertex layer*, then these edges are labeled by elements from R, and the set of these edges is not allowed to contain a directed cycle. These edges are called *constant edges*. Note that even though an ABP can have several vertices in vertex layer 0 and vertex layer d, there is only a single source and a single sink, and $w_{\mathrm{h}}(f)$ is defined as the maximum number of vertices in a vertex layer $1 \leq k < d$, i.e., we ignore vertex layer 0 and d. To an ABP we assign a pABP via the following process. We choose a vertex v that has no incoming constant edges (which must exist due to the absence of directed cycles), but at least one outgoing constant edge (v, w) with label $\alpha \in R$. We remove (v, w), and for each edge (u, v) with label $\beta(u, v)$ we add a new edge (u, w) with label $\alpha\beta(u, v)$. If an edge (u, w) already existed, then instead of

having two parallel edges, we keep one edge whose label is the sum of the two labels. If v is the source vertex, then v and u are merged into one new source vertex, keeping all their outgoing edges, and as before parallel edges are merged into single edges by adding their edge labels. The whole process removes at least 1 constant edge, and does not change the set of vertices in any layer besides layer 0. We continue this process until all constant edges are removed. At the end, we delete all non-source vertices in vertex layer 0, and all non-sink vertices in vertex layer d, so that we end up with a pABP. Let $G_1, G_2, \ldots, G_j$ denote the list of ABPs arising between the steps of this process. The invariant that is preserved between the steps is

$$\sum_{p \in P(G_1,s,t)} \beta(p) \;=\; \sum_{p \in P(G_2,s,t)} \beta(p) \;=\; \cdots \;=\; \sum_{p \in P(G_j,s,t)} \beta(p),$$

which proves the claim. □

7.II Degree and width

Proof of (1.3). The statements for the degree are obvious, but note that the second inequality can indeed be strict, because R can have zero divisors. The statements for the width are classical and can be proved by explicit constructions that appear for example in [Val79]. The width-of-sum statement is obtained by taking a pABP G_f for f and a pABP G_h for h and get a pABP for the sum $f + h$ by identifying the two sources with each other and identifying the two sinks with each other. The width-of-product statement is obtained by taking a pABP G_f for f and a pABP G_h for h and get a pABP for the product fh by identifying the sink in G_f with the source in G_h. Both constructions are illustrated in Figure 7. □

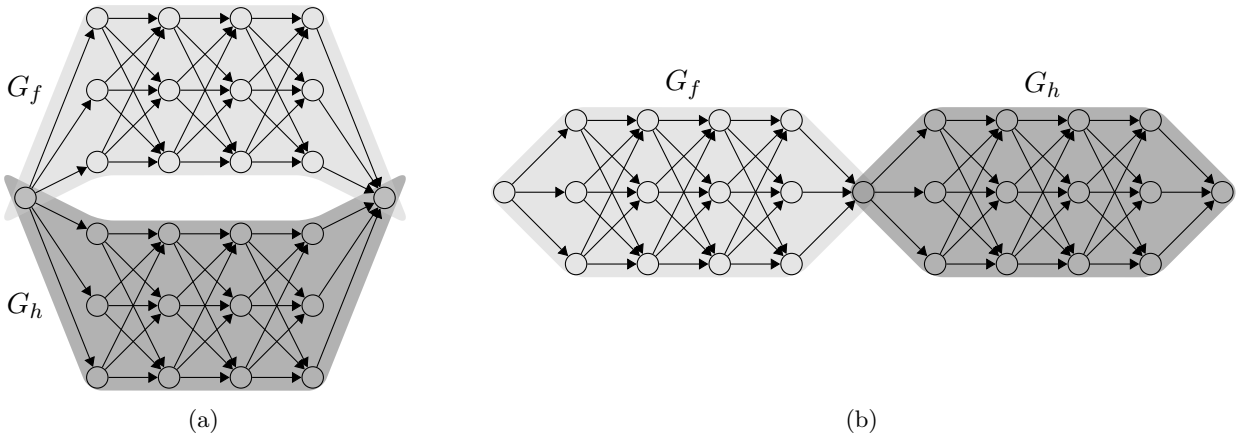


Figure 7: The constructions in the proof of (1.3). Addition is on the left, multiplication is on the right.

7.III Width compared to affine ABPs

An affine algebraic branching program (aABP) is a directed acyclic graph with a single source s and single sink t. The edges are labeled with (not necessarily homogeneous) polynomials of degree ≤ 1. The size of an aABP is the number of its vertices (not counting source or sink). An aABP computes a polynomial via the sum of the products of the edge labels, where the sum is over all paths from the source to the sink. Note that it is not required that all s-t-paths have the same

length. We write $\mathsf{ac}(f)$ for the smallest size of an aABP computing f. For $f \in R[\mathbf{x}]_d$, we clearly have

$$\mathsf{ac}(f) \leq d \cdot w(f). \tag{7.2}$$

The following proposition is classical and shows the other direction.

7.3 Proposition (Homogenization)**.** *Let h be a polynomial, and let f be a homogeneous component of h. Then $w(f) \leq \mathsf{ac}(h)$. In particular, $w(f) \leq \mathsf{ac}(f)$.* ◇

Proof. Let $d := \deg(f)$. We take a minimal size aABP G that computes h and we replace every vertex v with $d+1$ many vertices $v_0, \ldots, v_d$. We will set the edges so that if a vertex v computes some polynomial F, then v_i will compute the i-th homogeneous component of F, for $0 \leq i \leq d$, and we will not compute higher degree components. If $\ell + \alpha$ is an edge label in G from v to w with ℓ homogeneous linear and α constant, then we add edges with label ℓ from each v_i to w_{i+1}, and we add edges with label α from each v_i to w_i. For the source s we delete all s_i with $i > 0$, and for the sink t we delete all t_i with $i < d$, so that we have a unique source and sink. This finishes the construction. The result is an ABP that has at most $\mathsf{ac}(h)$ many vertices in each vertex layer but the first and last, in other words, the width of the ABP is at most $\mathsf{ac}(h)$. Hence $w(h) \leq \mathsf{ac}(h)$. □

7.IV Width compared to determinantal complexity

7.4 Proposition ([Val79])**.** *Let $f \in R[\mathbf{x}]_d$. We have $\mathsf{dc}(f) \leq d \cdot w(f)$.* ◇

Proof. Let G be a width n ABP that computes f. Then G has at most $(d-1)n+2$ many vertices. We identify the source and the sink, and to every other vertex we add a loop with label 1. Let G' denote the resulting graph, and let M be its adjacency matrix. Note that G' has $(d-1)n+1 \leq dn$ many vertices. Since $\det(M)$ is the sum over all signed cycle covers of G' (see (4.1)), we observe that the set of cycle covers in G' is in weight-preserving bijection to the set of s-t-paths in G, where the bijection is just adding/removing all loops. Moreover, all s-t-paths have length d, hence $\det(M) = (-1)^{d+1} f$. If d is even, then we swap two rows of M to obtain M' with $\det(M') = f$. □

Almost 45 years lie between Proposition 7.4 and its recent counterpart Proposition 7.5.

7.5 Proposition (Width variant of [CKV24, Thm 4.1])**.** *Let R be a field, and $f \in R[\mathbf{x}]_d$. Then $w(f) \leq d^4 \cdot \mathsf{dc}(f)$.* ◇

Proof. Let M be an $s \times s$ matrix whose entries are in $R[\mathbf{x}]_{\leq d}$ such that $\det(M) = f$. Let $M = M_1 + M_0$, where M_0 is a matrix of constants, and M_1 is a matrix of homogeneous linear polynomials. The matrix M_0 does not have full rank, because $\det(M)$ has constant part zero. Since R is a field, we find $g, h \in \mathrm{GL}_s$ with $\det(gh) = 1$ such that $gM_0h = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$. Define A, B, C, D via

$$gM_1h \;=\; \begin{pmatrix} A & B \\ C & -D \end{pmatrix},$$

where D and I have the same size, so that

$$gMh = \begin{pmatrix} A & B \\ C & I - D \end{pmatrix}.$$

We have $N := (I-D)^{-1} = \sum_{i \geq 0} D^i$ in the ring of formal power series. We have the Schur complement factorization:

$$\begin{pmatrix} A & B \\ C & I - D \end{pmatrix} = \begin{pmatrix} I & BN \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BNC & 0 \\ 0 & I - D \end{pmatrix} \begin{pmatrix} I & 0 \\ DC & I \end{pmatrix}.$$

Since the first and third matrix in the factorization have determinant 1, it follows that

$$\det\begin{pmatrix} A & B \\ C & I-D \end{pmatrix} = \det\begin{pmatrix} A-BNC & 0 \\ 0 & I-D \end{pmatrix} = \det(A-BNC)\det(I-D).$$

The lowest nonzero homogeneous part of this determinant is $\det(A-BNC)$. Since f is homogeneous of degree d, it follows that $A-BNC$ is a $d \times d$ matrix. Moreover, since f is homogeneous of degree d, f equals the homogeneous degree d part of $\det(A - B(\sum_{i=0}^{d-2} D^i)C)$, because for $i > d-2$ the monomials of the matrix BD^iC are of degree $> d$.

An aABP for an entry of $W = A - B(\sum_{i=0}^{d-2} D^i)C$ has size at most $s(d-1)$. We take the ABP for $\det_d$ from Theorem 6.1, which only has variables or their negations or constants as edge labels. We replace each variable $x_{i,j}$ with the aABP for $W_{i,j}$, and each negated variable $-x_{i,j}$ with the aABP for $-W_{i,j}$: Remove the edge (a,b) that has label $x_{i,j}$, then add a copy of the aABP for $W_{i,j}$ and identify the source with a, and the sink with b. The result is an aABP of size at most $O(sd^4)$. Since $w(f) \leq \mathsf{ac}(f)$ due to Proposition 7.3, it follows $w(f) \leq O(sd^4)$. □

No purely combinatorial proof of Proposition 7.5 is known. Such an algorithm could be helpful in resolving the following open problem.

7.6 Open Problem. Does there exist a function $\xi : \mathbb{N} \to \mathbb{N}$ such that for every commutative ring R and every $f \in R[\mathbf{x}]_d$ we have $w(f) \leq \xi(d) \cdot \mathsf{dc}(f)$? ⋄

7.V Exponents and fixed-parameter tractability

Recall that VBFPT consists of those $f_{n,d}$ for which the set $\{\gamma(f_{\cdot,d}) \mid d \in \mathbb{N}\}$ is bounded. This nice phrasing is possible, because the model of computation is nonuniform, see [BE19, footnote 5]. The following proposition shows the connection to the classical phrasing.

7.7 Proposition. *$\{\gamma(f_{\cdot,d}) \mid d \in \mathbb{N}\}$ is bounded iff $\exists \xi : \mathbb{N} \to \mathbb{N}, c \in \mathbb{N}\ \forall n, d : w(f_{n,d}) \leq \xi(d) \cdot n^c$.* ⋄

Proof. Given ξ, c with $\forall n, d : w(f_{n,d}) \leq \xi(d) \cdot n^c$. Then $\forall n, d : \log_n(w(f_{n,d})) \leq c + \log_n(\xi(d))$. For all d, we have $\lim_{n\to\infty} \big(c + \log_n(\xi(d))\big) = c$, and hence $\forall d : \gamma(f_{\cdot,d}) \leq c$.

Let $c' \geq 0$ such that $\forall d : \gamma(f_{\cdot,d}) \leq c'$. Hence, $\forall d, \varepsilon > 0\, \exists n_{d,\varepsilon}\, \forall n > n_{d,\varepsilon} : \log_n(w(f_{n,d})) \leq c' + \varepsilon$. In particular, $\forall d\, \exists n_d\, \forall n > n_d : \log_n(w(f_{n,d})) \leq c' + 1$. In other words, $\forall d\, \exists n_d\, \forall n > n_d : w(f_{n,d})) \leq n^{c'+1}$. Now, let $\xi(d) := \max\{w(f_{n,d})) \mid n \leq n_d\}$. Then $\forall d : w(f_{n,d})) \leq \xi(d) \cdot n^{c'+1}$. The claim follows by setting $c = c' + 1$. Note that the argument works for any $c > c' \geq 1$, which gives different ξ. □

7.VI Algebraically closed fields

In this paper, R is an arbitrary commutative ring. Not every such ring can be embedded into a field. For the sake of completeness, in this section we assume that R is an algebraically closed field, and we show how some arguments can be made via Zariski density arguments.

The Cayley–Hamilton Theorem The Cayley–Hamilton theorem over algebraically closed fields can be proved as follows. Note that for an $n \times n$ matrix A, every eigenvalue λ to an eigenvector v satisfies $Av = \lambda v$, in other words $(A - \lambda I_n)v = 0$, hence λ is a zero of $\det(A - tI_n) = 0$. Conversely, if λ is a zero of $\det(A - tI_n)$, then $\det(A - \lambda I_n) = 0$ and hence $A - \lambda I_n$ is not invertible, which implies that there exists v with $(A - \lambda I_n)v = 0$, hence λ is an eigenvalue of A.

The set of matrices that satisfy the Cayley–Hamilton theorem is Zariski closed, because the coefficients of the characteristic polynomial of A depend polynomially on the entries of A. The set of diagonalizable matrices with distinct eigenvalues is Zariski dense in the space of all $n \times n$ matrices. Hence, it suffices to prove the Cayley–Hamilton theorem for such matrices. For such

matrices, the characteristic polynomial of A factors: $\chi_A := \det(A - tI_n) = (t - \lambda_1) \cdots (t - \lambda_n)$, where λ_i are the eigenvalues of A. Let v_i be eigenvectors of λ_i. Now, for $P = \big(v_1 | \cdots | v_n\big)$, we have $A = P \operatorname{diag}(\lambda_1, \ldots, \lambda_n) P^{-1}$. Hence, for $p(t) = t^k$ we have $p(A) = P \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k) P^{-1}$, and by linearity it holds for all polynomials $p(t)$ that $p(A) = P \operatorname{diag}(p(\lambda_1), \ldots, p(\lambda_n)) P^{-1}$. Since the λ_i are the roots of χ_A, it follows that $\chi_A(A) = P \operatorname{diag}(0, \ldots, 0) P^{-1} = 0$, which finishes the proof of the Cayley–Hamilton theorem. Since R is infinite, by multivariate interpolation the Cayley–Hamilton theorem is also true as an identity of matrices of polynomials.

The matrix Girard–Newton Identities We now deduce the matrix Girard–Newton identities via a similar density argument from the classical Girard–Newton identities over algebraically closed fields. Let $\text{tr-pow}_{n,j} := \operatorname{tr}(X_n^j)$, where $X_n = \big(x_{i,j}\big)$ is the $n \times n$ matrix of variables. Our task is to prove

$$\sum_{j=0}^{d-1} (-1)^j \cdot \chi_{n,j} \cdot \text{tr-pow}_{n,d-j} \quad + \quad (-1)^d \cdot d \cdot \chi_{n,d} \ = 0 \tag{7.8}$$

from the fact that

$$\sum_{j=0}^{d-1} (-1)^j \cdot e_j(\mathbf{x}) \cdot p_{d-j}(\mathbf{x}) \quad + \quad (-1)^d \cdot d \cdot e_d(\mathbf{x}) \ = 0, \tag{7.9}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$, and $e_j(\mathbf{x}) = \sum_{S \subseteq \{1,\ldots,n\}, |S|=j} e_{S_1} \cdots e_{S_j}$ is the elementary symmetric polynomial, and $p_j(\mathbf{x}) = x_1^j + \cdots + x_n^j$ is the power sum polynomial.

Note that (7.8) for diagonal matrices follows directly from (7.9) by observing that for the diagonal matrix $D = \operatorname{diag}(x_1, \ldots, x_n)$ we have $\chi_{n,j}(D) = e_j(\mathbf{x})$ and $\text{tr-pow}_{n,j}(D) = p_j(\mathbf{x})$. Since both $\chi_{n,j}$ and $\text{tr-pow}_{n,j}$ are invariant under conjugation of their input variable matrices (note that $\chi_{n,j}$ is a coefficient of $\det(X_n - t \cdot I_n) = \det(PX_nP^{-1} - t \cdot I_n)$), and since (7.8) is true when evaluated at diagonal matrices, it follows (7.8) holds when evaluated at any diagonalizable matrix. Since $n \times n$ diagonalizable matrices are dense in the set of all $n \times n$ matrices, (7.8) holds for all $n \times n$ matrices. Since R is infinite, by multivariate interpolation (7.8) is also true as an identity of polynomials.

Acknowledgments CI thanks Robert Andrews for discussions about $\chi_{n,d}$, Markus Bläser for discussions about VBFPT, and Darij Grinberg for feedback on a draft version, pointers to the literature, and making [Gri25] available. CI was partially supported by EPSRC grant EP/W014882/2.

Generative AI Generative AI erroneously claimed that Theorem 5.1 was already known and appeared in the Fadeev–LeVerrier algorithm. During our investigation of this claim, we proved Corollary 2.2.

References

[AFP+19] Marian Aprodu, Gavril Farkas, Ştefan Papadima, Claudiu Raicu, and Jerzy Weyman. Koszul modules and Green's conjecture. *Inventiones mathematicae*, 218(3):657–720, 2019.

[AO96] Eric Allender and Mitsunori Ogihara. Relationships among PL, #L, and the determinant. *RAIRO-Theoretical Informatics and Applications*, 30(1):1–21, 1996.

[ARZ99] Eric Allender, Klaus Reinhardt, and Shiyu Zhou. Isolation, matching, and counting uniform and nonuniform upper bounds. *Journal of Computer and System Sciences*, 59(2):164–181, 1999.

[AW24] Robert Andrews and Avi Wigderson. Constant-depth arithmetic circuits for linear algebra problems. In *65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2367–2386, 2024.

[BCL23] Peter Bürgisser, Felipe Cucker, and Pierre Lairez. Rigid continuation paths II. Structured polynomial systems. In *Forum of Mathematics, Pi*, volume 11, page e12. Cambridge University Press, 2023.

[BE19] Markus Bläser and Christian Engels. Parameterized Valiant's Classes. In *14th International Symposium on Parameterized and Exact Computation (IPEC 2019)*, volume 148 of *LIPIcs*, pages 3:1–3:14, 2019. The arXiv version contains the full proofs: 1907.12287v2.

[Ber84] Stuart J. Berkowitz. On computing the determinant in small parallel time using a small number of processors. *Information Processing Letters*, 18(3):147–150, 1984.

[BIP19] Peter Bürgisser, Christian Ikenmeyer, and Greta Panova. No occurrence obstructions in geometric complexity theory. *Journal of the American Mathematical Society*, 32(1):163–193, 2019.

[BL13] Jarosław Buczyński and Joseph M Landsberg. Ranks of tensors and a generalization of secant varieties. *Linear Algebra and its Applications*, 438(2):668–689, 2013.

[BLMW11] Peter Bürgisser, Joseph M Landsberg, Laurent Manivel, and Jerzy Weyman. An overview of mathematical issues arising in the geometric complexity theory approach to VP≠VNP. *SIAM Journal on Computing*, 40(4):1179–1209, 2011.

[Buc84] Arthur Buchheim. Mathematical notes. *The Messenger of Mathematics*, 13:65–66, 1883–1884.

[Bür00a] Peter Bürgisser. *Completeness and reduction in algebraic complexity theory*, volume 7. Springer Science & Business Media, 2000.

[Bür00b] Peter Bürgisser. Cook's versus Valiant's hypothesis. *Theoretical Computer Science*, 235(1):71–88, 2000.

[CF69] Pierre Cartier and Dominique Foata. *Problemès combinatoires de commutation et réarrangements*, volume 85. Springer, 1969.

[Cha82] Seth Chaiken. A combinatorial proof of the all minors matrix tree theorem. *SIAM Journal on Algebraic Discrete Methods*, 3(3):319–329, 1982.

[Chi85] Alexander L Chistov. Fast parallel calculation of the rank of matrices over a field of arbitrary characteristic. In *International Conference on Fundamentals of Computation Theory*, volume 199 of *Lecture Notes in Computer Science*, pages 63–69, New York, Berlin, 1985. Springer.

[CKV24] Abhranil Chatterjee, Mrinal Kumar, and Ben Lee Volk. Determinants vs. algebraic branching programs. *computational complexity*, 33(2):11, 2024.

[Coo85] Stephen A Cook. A taxonomy of problems with fast parallel algorithms. *Information and control*, 64(1-3):2–22, 1985.

[Csa75] Laszlo Csanky. Fast parallel matrix inversion algorithms. In *16th Annual Symposium on Foundations of Computer Science (sfcs 1975)*, pages 11–12. IEEE, 1975.

[DF99] Rodney G Downey and Michael Ralph Fellows. *Parameterized complexity*. Monographs in computer science. Springer, 1999.

[DF13] R. G. Downey and M. R. Fellows. *Fundamentals of Parameterized Complexity*. Texts in Computer Science. Springer, London, UK, 2013.

[DGI+24] Pranjal Dutta, Fulvio Gesmundo, Christian Ikenmeyer, Gorav Jindal, and Vladimir Lysikov. Homogeneous Algebraic Complexity Theory and Algebraic Formulas. In *15th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 287 of *LIPIcs*, pages 43:1–43:23, 2024.

[DV22] Harm Derksen and Emanuele Viola. Fooling polynomials using invariant theory. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 399–406, 2022.

[dV25] Sjoerd de Vries. On Newton's identities in positive characteristic. *Journal of Algebra*, 668:348–364, 2025.

[Eis13] David Eisenbud. *Commutative algebra: with a view toward algebraic geometry*, volume 150. Springer Science & Business Media, 2013.

[FG06] Jörg Flum and Martin Grohe. *Parameterized complexity theory*. Springer, 2006.

[FLST24] Hervé Fournier, Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. On the power of homogeneous algebraic formulas. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, pages 141–151, 2024.

[Foa65] Dominique Foata. Etude algébrique de certains problèmes d'analyse combinatoire et du calcul des probabilités. In *Annales de l'ISUP*, volume 14, pages 81–241, 1965.

[Foa80] Dominique Foata. A combinatorial proof of Jacobi's identity. *Annals of Discrete Mathematics*, 6:125–135, 1980.

[For24] Michael A Forbes. Low-depth algebraic circuit lower bounds over any field. In *39th Computational Complexity Conference (CCC 2024)*, volume 300, pages 31:1–31:16, 2024.

[Gan59] F. R. Gantmacher. *The Theory of Matrices*, volume 1. Chelsea Publishing Company, New York, NY, USA, 1959.

[Ges79] Ira Gessel. Tournaments and Vandermonde's determinant. *Journal of Graph Theory*, 3:305–307, 1979.

[Gri25] Darij Grinberg. The trace Cayley–Hamilton theorem. arXiv:2510.20689v1, 2025.

[IL19] Christian Ikenmeyer and Vladimir Lysikov. Strassen's 2×2 matrix multiplication algorithm: a conceptual perspective. *Annali dell'Università di Ferrara*, 65(2):241–248, 2019.

[IP17] Christian Ikenmeyer and Greta Panova. Rectangular Kronecker coefficients and plethysms in geometric complexity theory. *Advances in Mathematics*, 319:40–66, 2017.

[KL14] Harlan Kadish and Joseph M Landsberg. Padded polynomials, their cousins, and geometric complexity theory. *Communications in Algebra*, 42(5):2171–2180, 2014.

[KS14] Neeraj Kayal and Ramprasad Saptharishi. A selection of lower bounds for arithmetic circuits. In *Perspectives in Computational Complexity: The Somenath Biswas Anniversary Volume*, pages 77–115. Springer, 2014.

[LST25] Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. Superpolynomial lower bounds against low-depth algebraic circuits. *Journal of the ACM*, 72(4):1–35, 2025.

[Min97] Michel Minoux. Bideterminants, arborescences and extension of the matrix-tree theorem to semirings. *Discrete Mathematics*, 171(1-3):191–200, 1997.

[Min03] Ján Mináč. Newton's identities once again! *The American mathematical monthly*, 110(3):232–234, 2003.

[MS01] Ketan D Mulmuley and Milind Sohoni. Geometric complexity theory I: An approach to the P vs. NP and related problems. *SIAM Journal on Computing*, 31(2):496–526, 2001.

[MS08] Ketan D Mulmuley and Milind Sohoni. Geometric complexity theory II: Towards explicit obstructions for embeddings among class varieties. *SIAM Journal on Computing*, 38(3):1175–1206, 2008.

[MV97a] Meena Mahajan and V Vinay. A combinatorial algorithm for the determinant. In *8th Symposium on Discrete Algorithms (SODA)*, pages 730–738, 1997.

[MV97b] Meena Mahajan and V Vinay. Determinant: Combinatorics, algorithms, and complexity. *Chicago Journal of Theoretical Computer Science*, 1997(5), 1997.

[MV99] Meena Mahajan and V Vinay. Determinant: Old algorithms, new insights. *SIAM journal on Discrete Mathematics*, 12(4):474–490, 1999.

[Nie06] Rolf Niedermeier. *Invitation to fixed-parameter algorithms*, volume 31. OUP Oxford, 2006.

[Orl78] James B Orlin. Line-digraphs, arborescences, and theorems of Tutte and Knuth. *Journal of Combinatorial Theory, Series B*, 25(2):187–198, 1978.

[Rot01] Günter Rote. *Division-Free Algorithms for the Determinant and the Pfaffian: Algebraic and Combinatorial Approaches*, pages 119–135. Springer Berlin Heidelberg, 2001.

[Rut64] DE Rutherford. Xix.–the Cayley-Hamilton theorem for semi-rings. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 66(4):211–215, 1964.

[Sam42] Paul A Samuelson. A method of determining explicitly the coefficients of the characteristic equation. *The Annals of Mathematical Statistics*, 13(4):424–429, 1942.

[SC04] Michael Soltys and Stephen Cook. The proof complexity of linear algebra. *Annals of Pure and Applied Logic*, 130(1-3):277–323, 2004.

[Sol02] Michael Soltys. Berkowitz's algorithm and clow sequences. *The Electronic Journal of Linear Algebra*, 9:42–54, 2002.

[Str83] Howard Straubing. A combinatorial proof of the Cayley–Hamilton theorem. *Discrete Mathematics*, 43(2-3):273–279, 1983.

[SW12] Dennis Stanton and Dennis White. *Constructive combinatorics*. Springer Science & Business Media, 2012.

[Tem81] H. N. V. Temperley. *Graph Theory and Applications*. Ellis Horwood Ltd., Chichester, 1981.

[Tod91] Seinosuke Toda. Counting problems computationally equivalent to computing the determinant. Technical Report CSIM 91-07, Department of Computer Science and Information Mathematics, University of Electro-Communications, Chofu-shi, Tokyo 182, Japan, May 1991.

[Tod92] Seinosuke Toda. Classes of arithmetic circuits capturing the complexity of computing the determinant. *IEICE Transactions on Information and Systems*, 75(1):116–124, 1992.

[Val79] Leslie Valiant. Completeness classes in algebra. In *Proceedings of the ACM Symposium on Theory of Computing*, pages 249–261, 1979.

[Val92] Leslie G Valiant. Why is boolean complexity theory difficult. *Boolean Function Complexity*, 169(84-94):4, 1992.

[vdBDG+25] Maxim van den Berg, Pranjal Dutta, Fulvio Gesmundo, Christian Ikenmeyer, and Vladimir Lysikov. Algebraic Metacomplexity and Representation Theory. In *40th Computational Complexity Conference (CCC)*, volume 339 of *LIPIcs*, pages 26:1–26:35, 2025.

[vzG87] Joachim von zur Gathen. Feasible arithmetic computations: Valiant's hypothesis. *J. Symb. Comput.*, 4(2):137–172, 1987.

[Zei84] Doron Zeilberger. A combinatorial proof of Newton's identities. *Discrete mathematics*, 49(3):319, 1984.

[Zei85] Doron Zeilberger. A combinatorial approach to matrix algebra. *Discrete Mathematics*, 56:61–72, 1985.