THEORY OF COMPUTING LIBRARY
GRADUATE SURVEYS 10 (2025), pp. 1–166
www.theoryofcomputing.org

Introduction to geometric complexity theory

Markus Bläser   Christian Ikenmeyer

Received August 4, 2018; Revised April 5, 2021; Published May 31, 2025

Abstract. Geometric complexity theory is an approach towards proving lower bounds in algebraic complexity theory via methods from algebraic geometry and representation theory. It was introduced by Mulmuley and Sohoni and has gained significant momentum over the last few years. Since deep methods from several different areas of mathematics are involved, geometric complexity theory has a steep learning curve. There are great survey articles on geometric complexity theory, but those require a significant level of mathematical background or often only sketch many of the proofs, see, e. g., Regan (*Bull. EATCS* 2002), Mulmuley (*J. ACM* 2011), Bürgisser, Landsberg, Manivel, Weyman (*SIAM J. Comput.* 2011), Grochow (PhD thesis, U. of Chicago 2012), Ikenmeyer (PhD thesis, Paderborn U., 2012), Landsberg (*Ann. U. di Ferrara*, 2015). This survey tries to be a gentle introduction for graduate students and even advanced undergraduate students in computer science that requires almost no background knowledge except for the usual knowledge in linear algebra and some basic knowledge in analysis. All the necessary concepts from algebraic geometry and representation theory are introduced and almost all proofs are given. We focus on two questions, the permanent versus determinant problem and the border rank problem for matrix multiplication. There have been many more

ACM Classification: F.1.3, F.2.3, I.1.2

AMS Classification: 68Q05, 68Q15, 68Q17, 20C30, 20G05

Key words and phrases: geometric complexity theory, algebraic complexity theory, representation theory, determinant, permanent, tensor rank, border rank, occurrence obstructions

 DOI: 10.4086/toc.gs.2025.010

results in the past few years, which we cannot cover, however, this survey should give the reader the neccessary background to understand them. The survey culminates in two recent results, a negative one for the permanent versus determinant question and a positive one for the matrix multiplication problem. We present the proof that occurrence obstructions essentially cannot resolve the permanent versus determinant question. However, occurence obstructions are only the most basic tool of geometric complexity theory and it might be well possible that the more general concept of multiplicity obstructions will resolve the problem. On the other hand, as a proof of concept, we show that occurrence obstructions indeed can give lower bounds for the border rank of matrix multiplication.

Contents

Acknowledgement. This survey is based on lecture notes of courses on geometric complexity theory given by the authors at Saarland University during the summer term 2017, winter term 2017/18, and summer term 2018. We thank all the participants, in particular, Julian Dörfler, Nick Fischer, Lennart Haas, Thomas Haslbauer, Umangathan Kandasamy, Felix Rech, Igor Schlegel, Julian Rosemann, and Philip Wellnitz.

1 Boolean circuits and arithmetic circuits

1.1 Introduction

Computational complexity theory is concerned with the study of the inherent complexity of computational problems. Its flagship conjecture is the famous $\mathrm{P} \neq \mathrm{NP}$ conjecture, which is one of the seven Millenium Problems of the Clay Mathematics Institute [27], ranking this conjecture as the most important one at the intersection of mathematics and theoretical computer science. To this day several thousand computational problems are classified as NP-complete, i. e., they have a polynomial time algorithm iff $\mathrm{P} = \mathrm{NP}$[1]. The practical importance of the $\mathrm{P} \neq \mathrm{NP}$ conjecture is at least twofold: First of all, many NP-complete problems are of high practical relevance and have to be solved every day in commercial and scientific applications, for example the traveling salesperson problem, integer programming, facility location, subset sum, knapsack, longest path, multiprocessor scheduling, tensor rank. Secondly, all current security notions in cryptography heavily rely on $\mathrm{P} \neq \mathrm{NP}$. Indeed, $\mathrm{P} = \mathrm{NP}$ would break all existing cryptographic ciphers. A lot of effort by many researchers has been put into resolving the $\mathrm{P} \neq \mathrm{NP}$ conjecture, but progress has been slow; see for example [34] for a survey, and [1] for a more recent survey that includes geometric complexity theory, in particular its recent revision [2]. Complexity questions in algebraic models date back at least to [67]. Valiant [80] argued that algebraic complexity is at the heart of our issues with Boolean complexity, introduced algebraic complexity classes, and highlighted the importance of the determinant for understanding efficient computation. Mulmuley and Sohoni [63, 64] realized that it is natural to use algebraic geometry and representation theory to study Valiant's questions. Some of these geometric ideas date back to Strassen [76] who discovered them in his study of the rank of the matrix multiplication tensor. Mulmuley and Sohoni's approach is now called *geometric complexity theory*.

Geometric complexity theory is an approach towards computational complexity lower bound questions via methods from algebraic geometry and representation theory. It has gained significant momentum over the last few years, but it has a steep learning curve which is a result of the many different areas of mathematics involved. This course tries to be a gentle introduction that requires almost no background knowledge. There are great survey articles on geometric

[1] We use the standard shorthand notation "iff" for "if and only if"

complexity theory, but those require a significant level of mathematical maturity and often only sketch many of the proofs, see, e. g., [68, 62, 24, 40, 46, 53].

1.2 The non-uniform P ≠ NP question: NP $\not\subseteq$ P/poly

To make our lives easier we will not directly discuss the P vs NP question, but its so-called *non-uniform analogue*, i. e., its *circuit complexity* version. If the non-uniform analogue is true, then also P ≠ NP.

We start with the basic definition of a circuit.

Definition 1.1 (Circuit)**.** Fix a set $\mathbb{F}$ (in our case $\mathbb{F}$ will be $\mathbb{F}_2 = \{0,1\}$ or the set $\mathbb{C}$ of complex numbers) and a set $S = \{s_i\}$ of functions s_i of arbitrary arity a_i, i. e., s_i maps from $\mathbb{F}^{a_i}$ to $\mathbb{F}$, where each $a_i \in \mathbb{N}_{\geq 1}$. (For example, for *Boolean circuits*, choose $\mathbb{F} = \mathbb{F}_2$ and $S = \{\texttt{and}, \texttt{or}, \texttt{not}\}$). A *circuit* C is a directed graph (abbreviated *digraph*) that contains no directed cycle such that the following properties hold (see Figure 1):

- A subset of the vertices with indegree 0 is labeled by indeterminates. These vertices are called the *input gates*. The other vertices with indegree 0 are labeled with elements of $\mathbb{F}$ and are called *constant gates*. All other vertices are called *computation gates*. A computation gate with outdegree 0 is called an *output gate*.
- Each computation gate g is labeled with a function $s_i \in S$ with arity a_i coinciding with the indegree of g.

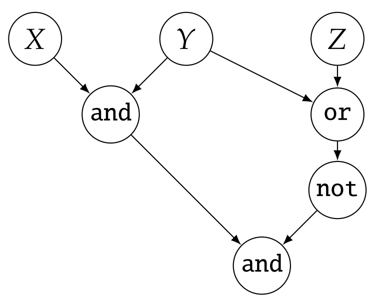

Figure 1: A circuit of size 7 computing the function $\{0,1\}^3 \to \{0,1\}$ given by $(X \texttt{ and } Y) \texttt{ and } \texttt{not}(Y \texttt{ or } Z)$. Here $\mathbb{F} = \mathbb{F}_2$ and $S = \{\texttt{and}, \texttt{or}, \texttt{not}\}$. The circuit has 3 input gates, no constant gate, 4 computation gates, one of which is an output gate.

Let m be the number of input gates of a circuit C. Since by definition circuits contain no directed cycle, for each gate g we can define a function $C_g \colon \mathbb{F}^m \to \mathbb{F}$ in the natural way by induction over the structure of the digraph as follows: For each input gate g labeled with a constant α we define C_g to be the constant function α. For each input gate g labeled with a variable X_j we define $C_g(x_1, \dots, x_m) = x_j$. For a computation gate g with label s and parents $g^1, \dots, g^a$ we define $C_g(x_1, \dots, x_m) = s(C_{g^1}(x_1, \dots, x_m), \dots, C_{g^a}(x_1, \dots, x_m))$.

We say that the functions C_g on the output gates g of C are *computed* by C. We call a circuit a *single-output circuit*, if it has only one output gate and in this case $C\colon \mathbb{F}^m \to \mathbb{F}$ denotes the function of the output gate. The *size* $|C|$ of a circuit C is defined to be the number of its vertices. The depth of a circuit is defined as the largest number of edges from an input gate to an output gate.

Definition 1.2. The *circuit complexity* $c_{\mathbb{F},S}(h)$ of a function $h : \mathbb{F}^m \to \mathbb{F}$ is the minimal size of a circuit C computing h.

Remark 1.3. In the literature, sometimes the number of edges or the number of computation gates is used as the definition of circuit complexity. In most contexts this does not make a significant difference.

A *Boolean circuit* is defined to be a circuit with $\mathbb{F} = \mathbb{F}_2$ and $S = \{\texttt{and}, \texttt{or}, \texttt{not}\}$. When we speak of a *univariate polynomial*, we mean a polynomial in one variable with *real* coefficients. The polynomials which we discuss in later chapters will be multivariate and will have *complex* coefficients. Those serve a completely different purpose and need to be distinguished from their univariate namesakes.

Definition 1.4. A sequence $(n_m)_{m\in\mathbb{N}}$ of natural numbers is called *polynomially bounded* if there exists a univariate polynomial q such that for all $m \in \mathbb{N}$ we have $n_m \le q(m)$.

For a sequence of functions (h_m) we obtain a sequence of natural numbers $c_{\mathbb{F},S}(h_m)$. Formally a *family* of objects is the same as a sequence of objects. We use the word family when we are interested in the sequence of complexity values.

Definition 1.5. Fix $\mathbb{F} := \mathbb{F}_2$ and $S = \{\texttt{and}, \texttt{or}, \texttt{not}\}$. The class P/poly consists of all function families (h_m) with $h_m\colon \mathbb{F}^m \to \mathbb{F}$ whose complexity sequence $c_{\mathbb{F},S}(h_m)$ is polynomially bounded.

Example 1.6. Let $h_m : \{0,1\}^m \to \{0,1\}$ denote the palindrome function: $h_m(w) = 1$ iff $w_i = w_{m+1-i}$ for all $1 \le i \le m$. It is easy to construct a Boolean circuit that computes h_m whose size is polynomially bounded in m.

More generally, for computer scientists, take any language $L \subseteq \{0,1\}^*$ in P. Then define the function $h_m : \{0,1\}^m \to \{0,1\}$ to be the indicator function of L restricted to input words of length exactly m. Then $(h_m) \in \mathrm{P/poly}$. In other words $\mathrm{P} \subseteq \mathrm{P/poly}$. This result is a bit technical and we will not discuss it any further.

1.2.1 SAT and NP

We do not define the class NP here, but we define the $\mathrm{NP} \not\subseteq \mathrm{P/poly}$ conjecture via the satisfiability function. The technical details in this subsection are only used locally.

A *Boolean formula* is a finite character string consisting of variables $x(1), x(2), x(3), \ldots$ and parantheses symbols (,), as well as the classical logical junctors and, or, not. For example:

$$(x(1)\ \texttt{and not}\ x(3))\ \texttt{or not}(x(1)\ \texttt{or}\ x(4))\ \texttt{or}\ x(2)$$

is a Boolean formula. A Boolean formula is called *satisfiable* if we can replace every variable $x(i)$ by either *true* or *false* such that the resulting statement is true. In our example, one of these assignments would be $x(1) = \text{true}$, $x(2) = \text{true}$, $x(3) = \text{true}$, $x(4) = \text{false}$, and hence the Boolean formula is satisfiable. We fix any reasonable way of encoding Boolean formulas as finite bit strings, so, for example, we could choose $\texttt{0000} = 0$, $\texttt{0001} = 1$, $\texttt{0010} = 2, \ldots, \texttt{1001} = 9$, $\texttt{1010} = x$, $\texttt{1011} = ($, $\texttt{1100} =)$, $\texttt{1101} = \texttt{and}$, $\texttt{1110} = \texttt{or}$, $\texttt{1111} = \texttt{not}$. For example the Boolean formula $x(2)$ and $x(3)$ is represented by $\texttt{101010110010110011011010101100111100}$. Using this encoding we can interpret the set of satisfiable Boolean formulas as a subset of the set of finite length bit strings. Now we are ready to define the satisfiability problem. Let (h_m) be the sequence of functions $h_m : \{0,1\}^m \to \{0,1\}$ defined by the property

$$h_m(w) = 1 \text{ iff } w \in \{0,1\}^m \text{ encodes a satisfiable Boolean formula.}$$

The NP $\not\subseteq$ P/poly conjecture can be stated as $(h_m) \notin \mathrm{P/poly}$, or equivalently as

the sequence of Boolean circuit complexities $c_{\mathbb{F}_2,S}(h_m)$ is not polynomially bounded.

Remark 1.7. One would usually choose the name SAT_m for the function h_m. h_m stands for a *hard* function, i. e., a function for which we want to prove complexity lower bounds. We will encouter this usage of h_m more often in the following chapters

1.3 From Boolean circuits to arithmetic circuits

If we interpret the set $\mathbb{F}_2 = \{0,1\}$ as the set of cosets modulo 2, then we see that besides the Boolean operations the set $\mathbb{F}_2$ is also a *ring* (even a *field*) and hence has an addition and a multiplication operation. The operation tables look as follows:

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

The following rephrased version of Definition 1.5 has a more algebraic flavor.

Proposition 1.8 (Arithmetic characterization of P/poly). *Let $\mathbb{F} := \mathbb{F}_2$ and let $S := \{+, \times\}$, where "+" and "×" have arity 2 and represent addition and multiplication. The class* P/poly *consists of all families (h_m) of functions $h_m : \mathbb{F}^m \to \mathbb{F}$ whose circuit complexity sequence $c_{\mathbb{F},S}(h_m)$ is polynomially bounded.*

Proof. For $\mathbb{F} = \mathbb{F}_2$, a circuit using $S = \{\texttt{and}, \texttt{or}, \texttt{not}\}$ can be converted into a circuit using $S = \{+, \times\}$ and vice versa by replacing gates with subcircuits of constant size, as follows:

- X and $Y = X \times Y$
- not $X = X + 1$
- X or $Y = X \times Y + X + Y$
- $X + Y = X$ xor $Y = (X \text{ or } Y)$ and not$(X \text{ and } Y)$ □

Circuits where $\mathbb{F}$ is a field and S is the set $\{+, \times\}$ of arithmetic operations are called *arithmetic circuits over* $\mathbb{F}$. Computation gates labeled with $+$ are called *addition gates* and computation gates labeled with $\times$ are called *multiplication gates*.

Given the characterization from Proposition 1.8, it is straightforward to work over other rings than $\mathbb{F}_2$. *Infinite fields* for example have a big advantage, as we will see in Lemma 1.9. For a fixed ring $\mathbb{F}$, single-output arithmetic circuits with m input gates not only naturally compute a function $\mathbb{F}^m \to \mathbb{F}$, but they also *compute a polynomial* in the polynomial ring $\mathbb{F}[X_1, \ldots, X_m]$ in m variables by induction on the circuit structure, see Figure 2. Two polynomials are consirered to

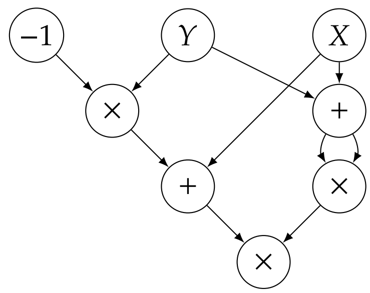

Figure 2: A circuit computing the polynomial $X^3 + X^2Y - XY^2 - Y^3$. Here $\mathbb{F} = \mathbb{C}$ and $S = \{\times, +\}$. The circuit has 2 input gates, one constant gate, 5 computation gates, and 1 output gate.

be equal if for each monomial their corresponding coefficients coincide. Single-output arithmetic circuits over $\mathbb{F}_2$ that compute different polynomials can compute the same function, as the following small example shows: Let $h_1(X, Y) = X^2Y$ and $h_2(X, Y) = XY^2$. Clearly h_1 and h_2 are different polynomials, but as functions they coincide:

$$\forall x \in (\mathbb{F}_2)^2 : h_1(x) = h_2(x).$$

This is a cumbersome subtlety which does not arise over infinite fields.

Lemma 1.9. *Let $\mathbb{F}$ be an infinite field. Then for two polynomials $h_1, h_2 \in \mathbb{F}[X_1, X_2, \ldots, X_m]$ we have*

$$h_1 = h_2 \text{ as polynomials} \quad \text{iff} \quad \text{for all } x \in \mathbb{F}^m \text{ we have } h_1(x) = h_2(x).$$

Proof. We show by induction that a polynomial that vanishes on the whole $\mathbb{F}^m$ is the zero polynomial. For $m = 1$ the result follows easily from successive polynomial division by linear factors: A nonzero degree d polynomial cannot have more than d zeros. For $m > 1$ we can decompose every h that vanishes on $\mathbb{F}^m$ as $h = \sum_{i=0}^{\deg h} g_i X_m^i \in \mathbb{F}[X_1, \ldots, X_{m-1}][X_m]$. Fix a point $(x_1, \ldots, x_{m-1}) \in \mathbb{F}^{m-1}$. Define $p(y) := h(x_1, \ldots, x_{m-1}, y) = \sum_{i=0}^{\deg h} g_i(x_1, \ldots, x_{m-1}) y^i \in \mathbb{F}[y]$. Note that $p(y)$ vanishes on $\mathbb{F}$ and hence p is the zero polynomial. But the coefficients of p are the $g_i(x_1, \ldots, x_{m-1})$. Thus equating coefficients of p yields that for all i we have $g_i(x_1, \ldots, x_{m-1}) = 0$. Since the point $(x_1, \ldots, x_{m-1})$ was chosen arbitrarily, all g_i vanish on the whole $\mathbb{F}^{m-1}$. By the induction hypothesis each g_i is the zero polynomial. Therefore h is the zero polynomial. ◻

In the light of Lemma 1.9 we see that if we are working over an infinite field we can focus on the *polynomials* computed by arithmetic circuits instead of the *functions* computed by them.

Our field of choice will be the complex numbers $\mathbb{C}$ from now on.

Definition 1.10. The *arithmetic complexity* $L(h)$ of a polynomial h is the size of the smallest single-output arithmetic circuit computing h.

The notation $L(h)$ stands for the *length of a straight-line program computing* h, which is the same as the arithmetic circuit complexity.

Example 1.11. $L(\prod_{i=1}^{m} X_i) = O(m)$.

Example 1.12. $L(\sum_{i=1}^{m}(X_i)^m) = O(m \log_2(m))$ using the repeated squaring algorithm.

Let $\mathfrak{S}_m$ denote the symmetric group on m letters, i. e., the group of bijective maps $\{1, \dots, m\} \to \{1, \dots, m\}$. We define the *permanent polynomial* as follows:

$$\mathrm{per}_m := \sum_{\pi \in \mathfrak{S}_m} \prod_{i=1}^{m} X_{i,\pi(i)} \quad \in \mathbb{C}[X_{1,1}, X_{1,2}, \dots, X_{m,m}]$$

Notice the striking similarity to the determinant:

$$\det_m := \sum_{\pi \in \mathfrak{S}_m} \mathrm{sgn}(\pi) \prod_{i=1}^{m} X_{i,\pi(i)} \quad \in \mathbb{C}[X_{1,1}, X_{1,2}, \dots, X_{m,m}],$$

where $\mathrm{sgn}(\pi) \in \{-1, 1\}$ denotes the *sign* of the permutation π.

Computing the permanent of a matrix is NP-hard [81], while determinants of matrices can be efficiently computed using Gaussian elimination. We postpone the definition of the Valiant's complexity classes VP and VNP, but Valiant's famous VP $\neq$ VNP conjecture can be stated as

the sequence $L(\mathrm{per}_m)$ is not polynomially bounded.

Remark 1.13. It is known [5] that $L(\sum_{i=1}^{n} X_i^n Y_i^n) = \Omega(n \log n)$, which is the best known lower bound for L.

A simplification via algebra

Proving circuit complexity lower bounds for Boolean functions is difficult. Replacing the base field $\mathbb{F}_2$ with $\mathbb{C}$ lets us study polynomials instead. Valiant's famous conjecture, a conjecture similar to P $\neq$ NP, says that $L(\mathrm{per}_m)$ is not polynomially bounded.
This is a first step towards a rich set of algebraic tools that will become available in later chapters.

2 Waring rank and border Waring rank

2.1 Waring rank

We start our study of arithmetic circuits with a very special case of circuits. For this we use a new gate in our arithmetic circuits: raising a polynomial to some fixed power d. We call these gates *degree d powering gates*. A circuit is *layered* if we can assign to each gate a natural number (its *layer*) so that edges from gates in layer i only go to gates in layer $i+1$. Also for our addition gates we allow arbitrarily high arities.

Definition 2.1. A layered arithmetic circuit C is called a $\Sigma\Lambda^d\Sigma^{\mathrm{hom}}$-circuit if C is a tree of depth 3 (not counting leaf vertices) whose leaves are variables and constants, the next layer consists of multiplication gates whose parents are exactly one variable and one constant, the next layer consists of addition gates with arbitrary arity, the next layer consists of degree d powering gates, the last layer consists of a single summation gate of arbitrary arity. The *size* of C is defined as the number of powering gates.

A polynomial in many variables is called *multivariate*. To each monomial we assign a *degree*, which is the sum of its exponents. For example $\deg(XY^2) = 3$ and $\deg(X^2Y^3Z) = 6$. If all monomials of a polynomial h have the same degree d, then we say that h is *homogeneous of degree d*. For example, $XY + 3X^2$ is homogeneous (of degree 2), but $XY + X + 1$ is not homogeneous. Constants are homogeneous of degree 0. The zero polynomial is homogeneous of all degrees. The permanent per_m is homogeneous of degree m.

Homogeneous polynomials of degree d are sometimes called *forms*. In particular homogeneous degree-1 polynomials are called *linear forms*, but it is less ambiguous to say *homogeneous linear form*.

For any fixed set of variables, the set of homogeneous degree d polynomials forms a vector space that we denote by $\mathbb{C}[X_1,\dots,X_N]_d$. Moreover, the degree function makes the polynomial ring $\mathbb{C}[X_1,\dots,X_N]$ a graded algebra (see Appendix A).[2] We observe that each $\Sigma\Lambda^d\Sigma^{\mathrm{hom}}$-circuit computes a homogeneous degree d polynomial.

Claim 2.2. *Equivalently, we can allow the leaves of the circuits in Definition 2.1 to be labeled with homogeneous linear forms, followed by degree d powering gates and then a single addition gate. Explicitly writing these linear forms as sums of scalar multiples of variables does not change the size of the circuit.*

Proof. The complexity is defined as the number of powering gates, which does not change when replacing the computation of the homogeneous linear forms with just a leaf whose label is the homogeneous linear form or vice versa. □

Definition 2.3. For a homogeneous degree d polynomial h, the *Waring rank* is defined as the smallest size of a $\Sigma\Lambda^d\Sigma^{\mathrm{hom}}$-circuit computing h. Alternatively, the Waring rank of h is

[2] A *graded algebra* basically means that we have a well-defined notion of a degree of an element. The polynomial ring is our main example of a graded algebra. Another important example is the quotient ring of a polynomial ring divided by an ideal generated by *homogeneous* polynomials. In the non-homogeneous case however, the situation is different. For example, in the ring $\mathbb{C}[X_1]/(X_1^2 - X_1)$ we have $X_1^3 = X_1^2$, therefore $\mathbb{C}[X_1]/(X_1^2 - X_1)$ does not inherit the degree function from $\mathbb{C}[X_1]$.

the smallest number of summands such that h can be expressed as a sum of d-th powers of homogeneous linear forms.

The Waring rank is an important quantity in classical algebraic geometry, also called *symmetric rank*.

Example 2.4. $XY = (\frac{X}{2} + \frac{Y}{2})^2 + (i\frac{X}{2} - i\frac{Y}{2})^2$, therefore the Waring rank of XY is at most 2.

Example 2.5 (taken from a presentation by Luke Oeding in 2017)**.** Let $h := X^2Y \in \mathbb{C}[X,Y]_3$. Then

$$6h = (X+Y)^3 + (-X+Y)^3 + (\sqrt[3]{-2}\,Y)^3.$$

A direct calculation can show that there is no better way: The Waring rank of $6h$ is 3. The following lemma shows that indeed the Waring rank of h is 3.

Remark 2.6. Over fields that are not algebraically closed the Waring rank is usually defined as the smallest number of nonzero coefficients in a linear combination of powers.

Lemma 2.7. *Waring rank is invariant under nonzero rescaling: h and αh have the same Waring rank for $\alpha \neq 0$.*

Proof. Let h be of degree d with Waring rank n and let C be the smallest $\Sigma\Lambda\Sigma^{\mathrm{hom}}$-circuit computing h. Let $0 \neq \alpha \in \mathbb{C}$. We rescale all the homogeneous linear forms at the leafs with $\sqrt[d]{\alpha}$ to obtain a circuit of the same size computing αh. For the other direction we apply the same argument, but we rescale by $\sqrt[d]{\alpha^{-1}}$. □

Here we see again how convenient the choice of $\mathbb{C}$ as a base field is: dth roots of numbers are guaranteed to exist.

Proposition 2.8. *Waring rank is always finite.*

Proof. We can apply Fischer's formula [33] to each monomial separately:

$$\gamma_n X_1 X_2 \cdots X_n = \sum_{\substack{b_i \in \{-1,1\} \\ 2 \le i \le n}} \left(\prod_{\ell=2}^n b_\ell\right)(X_1 + b_2X_2 + b_3X_3 + \cdots + b_nX_n)^n, \tag{2.1}$$

where $\gamma_n := n! \cdot 2^{n-1}$. □

2.2 The discriminant

How can we prove a Waring rank complexity lower bound for specific h? We study homogeneous degree 2 polynomials in two variables X and Y. The set of these polynomials is denoted by $\mathbb{C}[X,Y]_2$. Every $h \in \mathbb{C}[X,Y]_2$ can be written as

$$h = aX^2 + bXY + cY^2.$$

Recall the case when $Y = 1$: The polynomial $aX^2 + bX + c$ has a double root iff $b^2 - 4ac = 0$. Note that $aX^2 + bX + c$ has a double root α iff $aX^2 + bX + c = (X - \alpha)^2$. The same holds here:

$$\text{There exist scalars } \alpha, \beta \in \mathbb{C} \text{ with } aX^2 + bXY + cY^2 = (\alpha X + \beta Y)^2 \text{ iff } b^2 - 4ac = 0.$$

Thus to prove that the Waring rank of h is at least 2, we simply can verify that $b^2 - 4ac \neq 0$. This is a first example of a general method to prove complexity lower bounds.

For instance, expressing $XY = aX^2 + bXY + cY^2$ we obtain $a = c = 0$ and $b = 1$, so that $b^2 - 4ac = 1 \neq 0$. Therefore the Waring rank of XY is at least 2. This means that the $\Sigma\Lambda\Sigma^{\mathrm{hom}}$-circuit in Example 2.4 is optimal.

Proving lower bounds means that we have a set of polynomials W (the "easy" functions, here the set of polynomials of Waring rank 1) and we want to prove that some polynomial h is not in W (a "hard" polynomial). One approach is to find a function f—defined on the space of polynomials that we consider—that vanishes on W but $f(h) \neq 0$. Such f are called *separating functions* or *obstructions* against $h \in W$. Obstructions are a very general concept and we will see specialized types of obstructions based on representation theory in Sections 10, 11, and 12. In Sections 2.3 and 2.4 we discuss the case where the separating functions are continuous.

2.3 Border Waring rank

As in Example 2.5, let $h := X^2Y \in \mathbb{C}[X, Y]_3$. The Waring rank of h is 3, but:

$$\lim_{\varepsilon \to 0} \left(\frac{1}{3\varepsilon} X^3 + \frac{1}{3\varepsilon} (\varepsilon Y - X)^3 \right) = \lim_{\varepsilon \to 0} \left(X^2Y - \varepsilon XY^2 + \frac{1}{3} \varepsilon^2 Y^3 \right) = X^2Y. \tag{2.2}$$

So there is a curve of polynomials of Waring rank ≤ 2 that converges to h.

Let us formally define what this means. Let $\mathbb{A} := \mathbb{C}[X_1, \ldots, X_N]_n$. This is a finite dimensional vector space with $\dim \mathbb{A} = \binom{N+n-1}{n}$. Every element $h \in \mathbb{A}$ can be written as

$$h = \sum_{\lambda \in \mathbb{N}^N, |\lambda| = n} \alpha_\lambda X_1^{\lambda_1} \cdots X_N^{\lambda_N} \tag{2.3}$$

for some constants α_λ. We define the *norm* or *length* of a polynomial h

$$|h| := \sum_{\lambda \in \mathbb{N}^N, |\lambda| = n} |\alpha_\lambda|. \tag{2.4}$$

It is easy to check that this satisfies the axioms of a norm ($|h| \geq 0$, $|h| = 0$ iff $h = 0$, $|\alpha h| = |\alpha| \cdot |h|$, $|h_1 + h_2| \leq |h_1| + |h_2|$).

The distance between $h_1 \in \mathbb{A}$ and $h_2 \in \mathbb{A}$ is defined as $\mathrm{dist}(h_1, h_2) := |h_1 - h_2|$. This satisfies the axioms of a metric.

In particular, the triangle inequality holds:

$$\mathrm{dist}(h_1, h_3) \leq \mathrm{dist}(h_1, h_2) + \mathrm{dist}(h_2, h_3). \tag{2.5}$$

For example,

$$\mathrm{dist}(2X_1^2, 2X_1^2 + \frac{1}{100} X_1 X_2) \quad = \quad |2 - 2| + |0 - \frac{1}{100}| = \frac{1}{100}.$$

Definition 2.9. Let $\mathbb{A}$ be a finite dimensional complex vector space and let $h \in \mathbb{A}$. We say that a sequence $(h_i)_i$ with all $h_i \in \mathbb{A}$ *converges to* $h \in \mathbb{A}$, if

$$\forall \varepsilon \in \mathbb{R}_{>0}\ \exists i_0 \in \mathbb{N}\ \forall i > i_0 : \mathrm{dist}(h - h_i) < \varepsilon.$$

In this case we write $\lim_{i\to\infty} h_i = h$ and say that h is the *limit* of the sequence $(h_i)_i$. A sequence for which a limit exists is called *convergent*.

(2.2) is an example of a convergent sequence if we set $\varepsilon := \frac{1}{i}$. It is easy to see that every convergent sequence has a unique limit.

Remark 2.10. Choosing different norms in (2.4) has no effect on the convergence behaviour of sequences. The limits stay the same.

Definition 2.11. For a homogeneous degree d polynomial h the *border Waring rank* is defined as the smallest n such that h is the limit of a sequence of polynomials of Waring rank $\le n$.

In Example 2.3 we see that the border Waring rank of X^2Y is at most 2. Clearly, the border Waring rank of h cannot exceed the Waring rank of h, because for all h the constant sequence $(h, h, \ldots)$ converges to h.

2.4 Closures and continuous separating functions

We want to show complexity lower bounds by using the nonvanishing of separating functions such as the discriminant. But in this section we see that *continuous* functions cannot distinguish between Waring rank and border Waring rank. On the other hand we discover that if we search for functions that prove lower bounds on border Waring rank, then we can restrict our search to continuous functions only. We will see later that we can restrict our search space significantly further using algebraic geometry and representation theory.

We will use a very simple definition of continuity:

Definition 2.12. A function $f : \mathbb{A} \to \mathbb{A}'$ between two finite dimensional metric spaces spaces is called *continuous* if for every convergent sequence $(h_i)_i$ in $\mathbb{A}$, the sequence $f(h_i)$ converges in $\mathbb{A}'$ to $f(\lim_{i\to\infty} h_i)$.

Claim 2.13. *Functions defined by multivariate polynomials are continuous. ("Multivariate polynomials are continuous").*

Proof sketch. It is easy to see that for all $1 \le k \le N$ the coordinate function $f_k : \mathbb{C}^N \to \mathbb{C}$, $(T_1, \ldots, T_N) \mapsto T_k$ is continuous. Moreover, it is not hard to derive the facts that finite products and finite sums of continuous functions are continuous (here the proof for products is only slightly more involved). It follows by induction that all multivariate polynomials $f \in \mathbb{C}[T_1, \ldots, T_N]$ are continuous. ◻

Example 2.14. Claim 2.13 implies that the discriminant $b^2 - 4ac$ is continuous.

We will now see that if we use continuous functions to prove Waring rank lower bounds, then we actually prove *border* Waring rank lower bounds. Moreover, if we want to prove border Waring rank lower bounds, then we can restrict our search for separating functions to continuous functions. As a first step we reformulate border Waring rank in the language of $\mathbb{C}$-closures.

Definition 2.15. Given a (not necessarily linear) subset $W \subseteq \mathbb{A}$, the $\mathbb{C}$-*closure* $\overline{W}$ in $\mathbb{A}$ is defined as the set of the limits of all convergent sequences whose elements are taken from W.

Example 2.16. Consider $\mathbb{C} \setminus \{0\} \subseteq \mathbb{C}$. Then $\overline{\mathbb{C} \setminus \{0\}} = \mathbb{C}$.

Clearly, $W \subseteq \overline{W}$, because for all $h \in W$, the constant sequence $(h, h, \dots)$ converges to h. Moreover, if $V \subseteq W$, then $\overline{V} \subseteq \overline{W}$, because every sequence with elements from V is also a sequence with elements from W. Using the definition above, we see that the set of border Waring rank $\leq n$ polynomials is the $\mathbb{C}$-closure of the set of Waring rank $\leq n$ polynomials, that is,

$$\{h \in \mathbb{C}[X_1, \dots, X_M]_m \mid \text{border Waring rank}(h) \leq n\}$$
$$= \overline{\{h \in \mathbb{C}[X_1, \dots, X_M]_m \mid \text{Waring rank}(h) \leq n\}}\,.$$

Definition 2.17. A subset $W \subseteq \mathbb{A}$ is called $\mathbb{C}$-*closed in* $\mathbb{A}$, if $\overline{W} = W$, i. e., the limit of every convergent sequence $(h_i)_i$ with $h_i \in W$ is contained in W.

Lemma 2.18. *Let $W \subseteq \mathbb{A}$ be any subset. After taking the $\mathbb{C}$-closure in $\mathbb{A}$ once, taking the $\mathbb{C}$-closure in $\mathbb{A}$ again has no additional effect: $\overline{\overline{W}} = \overline{W}$. In particular, $\mathbb{C}$-closures in $\mathbb{A}$ are $\mathbb{C}$-closed in $\mathbb{A}$.*

Proof. Clearly $\overline{W} \subseteq \overline{\overline{W}}$. Let $h \in \overline{\overline{W}}$ be arbitrary and let $(h_i)_i$ denote a sequence converging to h with $h_i \in \overline{W}$, i. e., for each i there exists a sequence $(h_{i,j})_j$ such that $\lim_{j\to\infty} h_{i,j} = h_i$ and the $h_{i,j}$ are elements of W. Let $h_i(\varepsilon)$ denote the first entry in $(h_{i,j})_j$ such that the distance between $h_{i,j}$ and h_i is less than ε. Taking $\varepsilon = \frac{1}{i}$, it follows that the sequence $(h_i(\frac{1}{i}))_i$ converges to h and all elements $h_i(\frac{1}{i})$ are taken from W. Therefore $h \in \overline{W}$. □

The next Proposition 2.19 shows that continuous functions cannot distinguish between a set and its $\mathbb{C}$-closure. We will see in Lemma 2.20 that continuous functions are exactly those functions that can be used to distinguish points from $\mathbb{C}$-closed sets. Note that this is exactly what we need for proving border Waring rank lower bounds.

Proposition 2.19. *A continuous function $f : \mathbb{A} \to \mathbb{C}$ vanishes on a set $W \subseteq \mathbb{A}$ iff f vanishes on the $\mathbb{C}$-closure $\overline{W}$.*

Proof. Since $W \subseteq \overline{W}$, one direction is clear. Let $h \in \overline{W} \setminus W$. Let h_i be a sequence in W with $\lim_{i\to\infty} h_i = h$. Then $f(h_i) = 0$ and since f is continuous we have $f(\lim_{i\to\infty} h_i) = \lim_{i\to\infty} f(h_i) = \lim_{i\to\infty} 0 = 0$. □

Lemma 2.20. *Let $\overline{W} \subseteq \mathbb{A}$ be a $\mathbb{C}$-closed set. Then there exists a* continuous *function $f : \mathbb{A} \to \mathbb{C}$ that vanishes on $\overline{W}$* and nowhere else.

Remark 2.21. Note that the statement of Lemma 2.20 is trivial if we drop the requirement of f being *continuous*. However, it is also absolutely useless then.

Proof. For the easy direction, we see that the existence of an f that vanishes on $\overline{W}$ with $f(h) \neq 0$ clearly implies $h \notin \overline{W}$. For the other direction we need to construct a separating function f vanishing on $\overline{W}$ but not on h. Intuitively f is the distance function to $\overline{W}$. We define the distance f of h to $\overline{W}$ to be the infimum

$$f(h) := \inf\{\mathrm{dist}(h, h_1) \mid h_1 \in \overline{W}\},$$

which is the largest $\alpha \in \mathbb{R}$ such that for all $h_1 \in \overline{W}$ we have $\mathrm{dist}(h, h_1) \geq \alpha$. Clearly if $h \in \overline{W}$, then $f(h) = 0$. Therefore if $f(h) \neq 0$ we have $h \notin \overline{W}$. For the other direction we have to show that every h with $f(h) = 0$ lies in $\overline{W}$. Let h satisfy $f(h) = 0$. This means that there exists a sequence (h_i) of elements of $\overline{W}$ such that the distance sequence $\mathrm{dist}(h, h_i)$ converges to 0. Therefore $\lim_{i\to\infty} h_i = h$, but since $\overline{W}$ is $\mathbb{C}$-closed in $\mathbb{A}$ it follows that $h \in \overline{W}$.

It remains to show that f is continuous. It is sufficient to show that for all h and h_1, we have

$$|f(h) - f(h_1)| \leq \mathrm{dist}(h, h_1),$$

because for a sequence h_i converging to h this implies $\lim_{i\to\infty} f(h_i) = f(h)$.

Let $h_2 \in \overline{W}$ be arbitrary. Then by (2.5) we have $\mathrm{dist}(h, h_2) \leq \mathrm{dist}(h, h_1) + \mathrm{dist}(h_1, h_2)$ and therefore $f(h) \leq \mathrm{dist}(h, h_1) + \mathrm{dist}(h_1, h_2)$ or in other words $\mathrm{dist}(h, h_1) \geq f(h) - \mathrm{dist}(h_1, h_2)$. Since h_2 was arbitrary we obtain $\mathrm{dist}(h, h_1) \geq f(h) - f(h_1)$. Note that we here used the fact that if a non-strict inequality holds for a subset of the real numbers, then it also holds for its infimum. Reversing the roles of h and h_1 we obtain $\mathrm{dist}(h, h_1) \geq f(h_1) - f(h)$ and therefore $\mathrm{dist}(h, h_1) \geq |f(h) - f(h_1)|$. □

If we want to prove border Waring rank lower bounds, then we can restrict our search for separating functions to continuous functions. The proofs did not involve anything specific about Waring rank, so this holds in far higher generality. Using some algebraic geometry we will see later that we can restrict our search further to homogeneous polynomials (for example the discriminant is a homogeneous polynomial). Using some representation theory we will restrict the search space even further to homogeneous polynomials in irreducible representations.

A general approach to lower bounds

Proving lower bounds means that we have a set of functions W (the "easy" functions) and we want to prove that some function h is not in W (a "hard" function). One approach is to find a function f—defined on the space of functions that we consider—that vanishes on W but $f(h) \neq 0$. While finding any such f is as hard as showing that $h \notin W$, we can restrict our search to "nice" functions f. Here "nice" means continuous. You should be aware that with this approach, we can only prove that $h \notin \overline{W}$. If we are unlucky, $h \in \overline{W} \setminus W$ and we will never be able to prove this with this approach.

Example: Waring rank

The Waring rank of a homogeneous polynomial h of degree d is the smallest number of summands such that h can be expressed as a sum of d-th powers of homogeneous linear forms.
The smallest r such that h is the limit of a sequence of polynomials of Waring rank $\leq r$ is the border Waring rank.
The discriminant $b^2 - 4ac$ is a polynomial that vanishes on all polynomials $aX^2 + bXY + cY^2 \in \mathbb{C}[X,Y]_2$ of border Waring rank 1.

3 Actions, orbits, and orbit closures

Complexity lower bounds are about separating points h from $\mathbb{C}$-closures $\overline{W}$ by using functions f that vanish on $\overline{W}$ but not on h. In the previous chapter we saw that since $\overline{W}$ is $\mathbb{C}$-closed, we can restrict our search to continuous functions f only. In this chapter we find more properties of $\overline{W}$ that will help to reduce the search space for f even further: Our sets $\overline{W}$ are *group orbit closures*. As our main example we consider the Waring rank problem: We express the Waring rank problem as a monoid orbit problem and the border Waring rank problem as an orbit closure problem.

3.1 Monoid actions

For a set V, let $V \to V$ denote the set of maps from V to V. If V is a vector space, then let $\mathsf{End}(V)$ denote its monoid of endomorphisms, i. e., the monoid of *linear* maps $V \to V$ (recall Appendix A). For $V = \mathbb{C}^N$ we can identify $\mathsf{End}(V) = \mathbb{C}^{N\times N}$, i. e., the space of $N \times N$ complex matrices. We use the shorthand $\mathsf{End}_N := \mathsf{End}(\mathbb{C}^N)$. Here, we are interested in $V = \mathbb{C}[X_1, \dots, X_n]_d$, the space of homogeneous polynomials of degree d, as we will see in Section 3.2.

Definition 3.1. Let G be a monoid and V be a set.

1. An *action* of G on V is a monoid homomorphism $\varrho : G \to (V \to V)$.

2. If V is a vector space, then a *linear action* of G on V defined as a monoid homomorphism $\varrho : G \to \mathsf{End}(V)$.

We say that G *acts on* V and we write gv as a shorthand for $(\varrho(g))(v)$, $g \in G, v \in V$.

Recall that the axiom for a monoid homomorphism in this case is $\varrho(g \cdot \tilde{g}) = \varrho(g) \circ \varrho(\tilde{g})$, where "$\cdot$" is the monoid operation and "$\circ$" is the composition of maps. You can think of the monoid elements moving the points of V around. The identity element fixes all points.

Caveat: The monoids that we are mainly interested in are also endomorphism spaces, so $G = \mathsf{End}_n$ for some n. This might be a source of confusion.

Example 3.2. Let $G = \mathsf{End}_n$, let $V = \mathbb{C}^n$, and let $\varrho(g) = g$. In this case gv can be interpreted as the usual matrix-vector product.

A more interesting example is obtained by lifting the action to the function space, as we explain in the following section.

3.2 Lifting the action to the function space

Using this action we define the action of End_n on $V = \mathbb{C}[X_1, \dots, X_n]_d$. For a polynomial $h \in \mathbb{C}[X_1, \dots, X_n]_d$, $g \in \mathsf{End}_n$, $x \in \mathbb{C}^n$, define

$$(gh)(x) := h(g^T x), \tag{3.1}$$

where g^T is the transpose and $g^T x$ is the monoid action of End_n on $x \in \mathbb{C}^n$, i. e., the matrix-vector multiplication (see Example 3.2). You can also think of End_n acting on V by replacing the variables $X_1, \dots, X_n$ by homogeneous linear forms in $X_1, \dots, X_n$.

We verify that the definition in (3.1) satisfies $(g\tilde{g})h = g(\tilde{g}h)$ as follows:

$$\begin{aligned} (g(\tilde{g} \cdot h))(x) &\overset{(3.1)}{=} (\tilde{g} \cdot h)(g^T x) \overset{(3.1)}{=} h(\tilde{g}^T(g^T x)) \\ &\overset{3.2}{=} h((\tilde{g}^T \cdot g^T) \cdot x) = h((g \cdot \tilde{g})^T \cdot x) \overset{(3.1)}{=} ((g \cdot \tilde{g})h)(x). \end{aligned}$$

Note that we take the transpose since G acts "from the left" on V and transposing reverses the order of the two monoid elements. Another way to define this action is to use g^{-1} instead of g^T, but that only works if G is a group. On the other hand, if G is an arbitrary group and $g \in G$, then it is unclear what g^T means. In all the cases we encounter it is just a matter of taste which definition to use.

Example 3.3. We have

$$\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}}_{=:g} (X_1 X_2 \cdots X_n) = X_1^n.$$

The following calculation may seem unnecessary, as g sends each X_i to X_1, so it sends $X_1 \cdots X_n$ to X_1^n. However, we go through it carefully to illustrate the importance of the transpose in the definition $(g \cdot f)(x) = f(g^T x)$.

Calculation: Let $h := X_1 \cdots X_n$. Then $(g \cdot h)(x) = h(g^T x)$. For $x = (x_1, \dots, x_n)$, we have $g^T x = (x_1, x_1, \dots, x_1)$, so then $h(g^T x) = h(x_1, x_1, ..., x_1) = x_1^n$. Since $(g \cdot h)(x)$ agrees with the polynomial X_1^n at all points x, the two must be equal. Note that if we had written $h(gx)$ instead, we would have that $gx = (x_1 + x_2 + \cdots + x_n, 0, 0, \dots, 0)$, and then $h(gx) = 0$, clearly not what we wanted. ◻

Example 3.4. Let $g \in \mathbb{C}^{n \times n}$ and let $\ell = (\ell_1, \dots, \ell_n) \in \mathbb{C}^n$ be the first column of g.

$$g(X_1^d) = (\ell_1 X_1 + \ell_2 X_2 + \cdots + \ell_n X_n)^d.$$

Calculation: Let $x \coloneqq \alpha_1 e_1 + \cdots + \alpha_n e_n$. We have $(g(X_1^d))(x) \coloneqq (X_1^d)(g^T x)$. But

$$g^T x = (\ell_1\alpha_1 + \ell_2\alpha_2 + \cdots + \ell_n\alpha_n)e_1 + \beta_2 e_2 + \cdots + \beta_n e_n$$

for some $\beta_2, \ldots, \beta_n \in \mathbb{C}$. Thus,

$$(X_1^d)(g^T x) = (\ell_1\alpha_1 + \ell_2\alpha_2 + \cdots + \ell_n\alpha_n)^d = ((\ell_1 X_1 + \ell_2 X_2 + \cdots + \ell_n X_n)^d)(x). \qquad \square$$

More generally, Example 3.4 shows that

$$g(X_i^d) = (\ell_1' X_1 + \ell_2' X_2 + \cdots + \ell_n' X_n)^d, \tag{3.2}$$

where $(\ell_1', \ldots, \ell_n')$ is the ith column of g.

We will combine these insights with the following two structural properties. This lifted action is linear and an algebra homomorphism, as the following two lemmas show. Let $\mathbb{A} = \mathbb{C}^N$ and define $\mathbb{C}[\mathbb{A}] \coloneqq \mathbb{C}[X_1, \ldots, X_N]$.

Lemma 3.5. *Let $h, h' \in \mathbb{C}[\mathbb{A}]$ and let $g \in G$. For all complex numbers α, α' we have*

$$g(\alpha h + \alpha' h') = \alpha(gh) + \alpha'(gh').$$

Proof. Let $x \in \mathbb{A}$ be arbitrary. We calculate

$$\begin{aligned}
(g(\alpha h + \alpha' h'))(x) &\overset{(3.1)}{=} (\alpha h + \alpha' h')(g^T x) \overset{(*)}{=} \alpha h(g^T x) + \alpha' h'(g^T x) \\
&\overset{(3.1)}{=} \alpha((gh)(x)) + \alpha'((gh')(x)) \\
&\overset{(*)}{=} (\alpha(gh) + \alpha'(gh'))(x),
\end{aligned}$$

where $(*)$ uses the fact that $\mathbb{C}[\mathbb{A}]$ is a vector space. $\square$

More generally, Lemma 3.5 holds by induction for arbitrary finite linear combinations of functions in $\mathbb{C}[\mathbb{A}]$.

Lemma 3.6. *Let $h, h' \in \mathbb{C}[\mathbb{A}]$ and let $g \in G$. Then*

$$g(h \cdot h') = (gh) \cdot (gh').$$

Proof. Let $x \in \mathbb{A}$ be arbitrary. We calculate

$$\begin{aligned}
(g(h \cdot h'))(x) &\overset{(3.1)}{=} (h \cdot h')(g^T x) \overset{(*)}{=} h(g^T x) \cdot h'(g^T x) \\
&\overset{(3.1)}{=} (gh)(x) \cdot (gh')(x) \\
&\overset{(*)}{=} ((gh) \cdot (gh'))(x),
\end{aligned}$$

where $(*)$ follows from the definition of the product of two functions. $\square$

Note that Lemma 3.5 and Lemma 3.6 imply that gh can be calculated by taking products and linear combinations of gX_i, but gX_i is just the homogeneous linear form given by the ith column of g. For example

$$\begin{aligned}\begin{pmatrix}1 & 2\\ 1 & 0\end{pmatrix}(X_1X_2+X_1^2) &= (X_1+X_2)(2X_1)+(X_1+X_2)^2\\ &= 2X_1^2+2X_1X_2+X_1^2+2X_1X_2+X_2^2=3X_1^2+4X_1X_2+X_2^2,\end{aligned}$$

where the action replaces X_1 by the linear form X_1+X_2 given by the first matrix column, and the action replaces X_2 by $2X_1$ given by the second matrix column.

Corollary 3.7. *Let $g\in\mathbb{C}^{n\times n}$ and let $\ell^i\in\mathbb{C}^n$ be the ith column of g. Let $L_i := \ell_1^iX_1+\cdots+\ell_n^iX_n\in\mathbb{C}[X_1,\dots,X_n]_1$.*

$$g(X_1^d+\cdots+X_n^d)=L_1^d+\cdots+L_n^d.$$

Proof. Combine (3.2) with Lemma 3.5. □

Example 3.8. We can rewrite Example (2.2) as

$$\lim_{\varepsilon\to 0}\left(\frac{1}{3\varepsilon}\begin{pmatrix}1 & -1\\ 0 & \varepsilon\end{pmatrix}(X^3+Y^3)\right)=X^2Y.$$

3.3 Orbits

In this section we express the Waring rank problem as a problem on orbits of monoids. We'll see later that *border* Waring rank is a corresponding problem on orbit *closures*.

Definition 3.9. For a monoid G acting on a set V define $Gh := \{gh \mid g\in G\}$ for $h\in V$. We call Gh the *orbit of h*.

Example 3.10. For $G=\mathbb{C}^{N\times N}$ the orbit $GX_1^d\subseteq\mathbb{C}[X_1,\dots,X_N]_d$ is the set of Waring rank 1 homogeneous degree d polynomials in N variables.

Proof. For $g\in G$ let $\ell := g_{1,1}X_1+g_{2,1}X_2+\cdots+g_{N,1}X_N$. Then $gX_1^d=\ell^d$ by Example 3.4.

For the other direction, let ℓ^d have Waring rank 1, $\ell\in\mathbb{C}^N$. Let $g\in\mathbb{C}^{N\times N}$ with first column ℓ. Then by Example 3.4 we have $gX_1^d=\ell^d$. □

We can now phrase the Waring rank problem in the language of monoid orbits as follows. For $n\le N$ we embed $\mathbb{C}[X_1,\dots,X_n]_d\subseteq\mathbb{C}[X_1,\dots,X_N]_d$ in the natural way.

Proposition 3.11. *Let $N\ge n$ and $N\ge m$. Let $G := \mathsf{End}_N$.*

$$\{h\in\mathbb{C}[X_1,\dots,X_m]_d \mid \textit{Waring rank of } h \textit{ is at most } n\}=G(X_1^d+\cdots+X_n^d)\cap\mathbb{C}[X_1,\dots,X_m]_d.$$

Proof. If $h = g(X_1^d + \cdots + X_n^d)$ for some $g \in \mathbb{C}^{N\times N}$, then $h = (g_{1,1}X_1 + g_{2,1}X_2 + \cdots + g_{N,1}X_N)^d + \cdots + (g_{1,n}X_1 + g_{2,n}X_2 + \cdots + g_{N,n}X_N)^d$ by Cor. 3.7 and thus the Waring rank of h is at most n. (Note that we could also have set all $g_{i,j}$ to zero for which $j > n$ or $i > m$.)

Conversely, if the Waring rank of $h \in \mathbb{C}[X_1, \ldots, X_m]$ is at most n, then

$$h = (g_{1,1}X_1 + g_{2,1}X_2 + \cdots + g_{m,1}X_m)^d + \cdots + (g_{1,n}X_1 + g_{2,n}X_2 + \cdots + g_{m,n}X_n)^d.$$

Since $N \ge n$ and $N \ge m$, we can construct a matrix $g \in \mathbb{C}^{N\times N}$ by filling the remaining cells with zeros. Then $h = g(X_1^d + \cdots + X_n^d)$ by Cor. 3.7. □

In the proposition above, we have to intersect the orbit on the right-hand side by $\mathbb{C}[X_1, \ldots, X_m]_d$, since the G-action can introduce variables with index $> m$, which cannot occur on the left-hand side.

We now want to go one step further and look at the inclusion of monoid orbits than of just point membership. In this way we could use properties of the point h to show $h \notin W$.

Lemma 3.12. *The orbit Gh is the smallest set that contains h and is closed under the monoid action.*

Proof. If a set contains h and is closed under the monoid action, then it contains all gh with $g \in G$, so by definition it contains Gh.

Moreover, Gh is closed under the monoid action: Let $g \in G$ be arbitrary and let $h' \in Gh$ be arbitrary. By definition there exist $g' \in G$ such that $h' = g'h$. Thus $gh' = g(g'h) \overset{(*)}{=} (gg')h \in Gh$, where $(*)$ follows from the axioms of a monoid action. □

Corollary 3.13. *Let $N \ge n$ and $N \ge m$. Let $G := \mathsf{End}_N$. Let $h \in \mathbb{C}[X_1, \ldots, X_m]_d$. The Waring rank of h is at most n iff $Gh \subseteq G(X_1^d + \cdots + X_n^d)$.*

Proof. $h \in G(X_1^d + \cdots + X_n^d)$ iff $Gh \subseteq G(X_1^d + \cdots + X_n^d)$ by Lemma 3.12. Prop. 3.11 says that the Waring rank of h is at most n iff $h \in G(X_1^d + \cdots + X_n^d)$, which finishes the proof. □

3.4 Orbit closures

In this section we express the border Waring rank problem as a problem on monoid orbit closures.

The vector space $\mathsf{End}_N = \mathbb{C}^{N\times N}$ is endowed with the standard metric

$$\mathrm{dist}(g, g') = \sum_{i,j=1}^{N} |g_{i,j} - g'_{i,j}|,$$

$g, g' \in \mathsf{End}_N$.

Consider the metric space $\mathsf{End}_N \times \mathbb{C}[X_1, \ldots, X_m]_d$ via $\mathrm{dist}((g,h),(g',h')) := \mathrm{dist}(g,g') + \mathrm{dist}(h,h')$. We postpone the proof of the following simple technical lemma to Section 3.6.

Lemma 3.14. *The map $\mathsf{End}_N \times \mathbb{C}[X_1, \ldots, X_m]_d \to \mathbb{C}[X_1, \ldots, X_N]_d$, $(g,h) \mapsto gh$ given by the action in Section 3.2 is continuous.*

Corollary 3.15.

1. *For a fixed $h \in \mathbb{C}[X_1, \dots, X_m]_d$ the map $\mathsf{End}_N \to \mathbb{C}[X_1, \dots, X_N]_d$, $g \mapsto gh$ is continuous.*
2. *For a fixed $g \in \mathsf{End}_N$ the map $\mathbb{C}[X_1, \dots, X_m]_d \to \mathbb{C}[X_1, \dots, X_N]_d$, $h \mapsto gh$ is continuous.*

Proof. Both maps are restrictions of the continuous map in Lemma 3.14. □

Lemma 3.16. *The monoid orbit closure $\overline{Gh}$ is the smallest set that contains h, is closed under the monoid action, and $\mathbb{C}$-closed.*

Proof. Let X be a set that contains h, is closed under the monoid action, and is $\mathbb{C}$-closed. We have

$$h \in X \Leftrightarrow Gh \subseteq X \Leftrightarrow \overline{Gh} \subseteq X.$$

For the last equivalence we used that if $A \subseteq B$, then $\overline{A} \subseteq \overline{B}$ and that $\overline{B} = B$ for $\mathbb{C}$-closed sets B.

One subtlety remains: A priori it is unclear that $\overline{Gh}$ is closed under the monoid action. We prove this as follows. Let $h' \in \overline{Gh}$, $h' = \lim_{i\to\infty} g_i h$ with $g_i \in G$. Let $g \in G$ be arbitrary. Then $gh' = g \lim_{i\to\infty} g_i h = \lim_{i\to\infty} g(g_i h) = \lim_{i\to\infty}(gg_i)h \in \overline{Gh}$, because the map $h \mapsto gh$ is continuous for every $g \in G$ (Cor. 3.15). □

Corollary 3.17. *Let $N \geq n$ and $N \geq m$. Let $G := \mathsf{End}_N$. Let $h \in \mathbb{C}[X_1, \dots, X_m]_d$. The border Waring rank of h is at most n iff $\overline{Gh} \subseteq \overline{G(X_1^d + \cdots + X_n^d)}$.*

Proof. Using Prop. 3.11 we see that the border Waring rank of h is at most n iff $h \in \overline{G(X_1^d + \cdots + X_n^d)}$. But since $\overline{G(X_1^d + \cdots + X_n^d)}$ is $\mathbb{C}$-closed and closed under the action of G, from Lemma 3.16 it follows that $h \in \overline{G(X_1^d + \cdots + X_n^d)}$ iff $\overline{Gh} \subseteq \overline{G(X_1^d + \cdots + X_n^d)}$. □

3.5 Group orbit closures

It is more common to talk about group orbit closures instead of monoid orbit closures. The reason is that we can replace End_N in Cor. 3.17 by the general linear group $\mathsf{GL}_N := \{g \in \mathsf{End}_N \mid \det(g) \neq 0\}$.

Lemma 3.18 (Density of $\mathsf{GL}_N \subseteq \mathsf{End}_N$)**.** *For every $g \in \mathsf{End}_N$ there exists a sequence (g_i) with each $g_i \in \mathsf{GL}_N$ such that $\lim_{i\to\infty} g_i = g$.*

Proof. Consider $\det(g + \varepsilon \mathrm{Id}_N)$, which is a nonzero univariate polynomial in ε of degree $\leq N$. Thus it has at most N zeros. From the sequence (g_i), $g_i := g + \frac{1}{i}\mathrm{Id}_N$ we remove those g_i with zero determinant (these are at most N many). Then (g_i) converges to g with all $g_i \in \mathsf{GL}_N$. □

Proposition 3.19. *Let $N \geq n$ and $N \geq m$. Let $G := \mathsf{GL}_N$. Let $h \in \mathbb{C}[X_1, \dots, X_m]_d$. The border Waring rank of h is at most n iff $\overline{Gh} \subseteq \overline{G(X_1^d + \cdots + X_n^d)}$.*

Proof. We prove that in general, $\overline{\mathsf{GL}_N h} = \overline{\mathsf{End}_N h}$. According to Cor. 3.15 the map $\varphi : g \mapsto gh$ is continuous. In general, for any continuous map φ and any set G we have $\varphi(\overline{G}) \subseteq \overline{\varphi(G)}$. The proof of this fact is short: let $g \in \overline{G}$ with $g = \lim_{i\to\infty} g_i$. Then $\varphi(g) = \varphi(\lim_{i\to\infty} g_i) = \lim_{i\to\infty} \varphi(g_i) \in \overline{\varphi(G)}$.

Using $\varphi(\overline{G}) \subseteq \overline{\varphi(G)}$, we take closures on both sides: $\overline{\varphi(\overline{G})} \subseteq \overline{\varphi(G)}$ (Lemma 2.18). Since clearly $\overline{\varphi(G)} \subseteq \overline{\varphi(\overline{G})}$, we have $\overline{\varphi(\overline{G})} = \overline{\varphi(G)}$. Setting $G = \mathsf{GL}_N$ and using that $\overline{G} = \mathsf{End}_N$ by Lemma 3.18, the statement follows. □

Remark 3.20. Although $\mathsf{End}_N = \overline{\mathsf{GL}_N}$, we often have the strict inclusion $\mathsf{End}_N h \subsetneq \overline{\mathsf{GL}_N h}$. For example, the set $\mathsf{End}_2 X^3 + Y^3$ is the set of all Waring rank ≤ 2 polynomials in $\mathbb{C}[X,Y]_3$ (see Corollary 3.13), whereas $\overline{\mathsf{GL}_2 X^3 + Y^3}$ is the set of all border Waring rank ≤ 2 polynomials in $\mathbb{C}[X,Y]_3$. The example in (2.2) shows that the former is strictly contained in the latter. A more complicated example is discussed in [45]: $\mathsf{End}_9 \mathrm{det}_3$ is strictly contained in $\overline{\mathsf{GL}_9 \mathrm{det}_3}$.

3.6 The orbit map

In this section we prove Lemma 3.14. Indeed, we prove the following more general statement.

Proposition 3.21. *Let $\varphi : \mathsf{End}_N \times \mathbb{C}[X_1,\ldots,X_N]_d \to \mathbb{C}[X_1,\ldots,X_N]_d$, $\varphi(g,h) = gh$. Let $\eta := \dim \mathbb{C}[X_1,\ldots,X_N]_d = \binom{N+d-1}{d}$. For $1 \leq i \leq \eta$ define φ_i to be the ith coordinate function of φ. Then each φ_i is given by a polynomial in the $N^2 + \eta$ coordinate variables of $\mathsf{End}_N \times \mathbb{C}[X_1,\ldots,X_N]_d$.*

Since polynomials are continuous and combining continuous coordinate functions gives a continuous function, Proposition 3.21 implies Lemma 3.14.

Proof of Prop. 3.21. Let the entry in row i and column j of g be denoted by g^i_j.

For a list $(i^1, i^2, \ldots, i^d)$ of numbers let $S(i^1, i^2, \ldots, i^d)$ denote the set of all lists that have the same entries as $(i^1, i^2, \ldots, i^d)$, but where the positions are permuted. Let $s^{i^1,i^2,\ldots,i^d}_{i_1,i_2,\ldots,i_d}$ denote the sum

$$s^{i^1,i^2,\ldots,i^d}_{i_1,i_2,\ldots,i_d} := \sum_{(j^1,\ldots,j^d)\in S(i^1,i^2,\ldots,i^d)} g^{j^1}_{i_1} g^{j^2}_{i_2} \cdots g^{j^d}_{i_d}.$$

For $1 \leq i_1 \leq i_2 \leq \cdots \leq i_d \leq N$ we have

$$\begin{aligned} g(X_{i_1}\cdots X_{i_d}) &= (g^1_{i_1}X_1 + \cdots + g^N_{i_1}X_N)\cdots(g^1_{i_d}X_1 + \cdots + g^N_{i_d}X_N) \\ &= \sum_{1\leq i^1\leq i^2\leq\cdots\leq i^d\leq N} s^{i^1,i^2,\ldots,i^d}_{i_1,i_2,\ldots,i_d} X_{i^1}\cdot X_{i^2}\cdots X_{i^d} \end{aligned}$$

and thus

$$g\left(\sum_{1\le i_1\le i_2\le\cdots\le i_d\le N}\alpha_{i_1,\dots,i_d}X_{i_1}\cdots X_{i_d}\right)$$
$$=\sum_{1\le i^1\le i^2\le\cdots\le i^d\le N}\left(\sum_{1\le i_1\le i_2\le\cdots\le i_d\le N}\alpha_{i_1,\dots,i_d}s^{i^1,i^2,\dots,i^d}_{i_1,i_2,\dots,i_d}\right)X_{i^1}\cdot X_{i^2}\cdots X_{i^d}.$$

The term in parantheses is homogeneous of degree $d+1$ in the η variables $\alpha_{i_1,\dots,i_d}$ and the N^2 variables g^i_j. □

Orbits and orbit closures

Recall from the previous section that we want to prove that a particular polynomial h is not contained in a set of polynomials W. The monoid $G = \mathsf{End}_n$ acts on $\mathbb{C}[X_1,\dots,X_n]$ or $\mathbb{C}[X_1,\dots,X_n]_d$ by replacing the variables by homogeneous linear forms. Instead of showing that $h\notin W$, we try to prove that h is not contained in a certain G-orbit.
In the case of Waring rank, this works particularly well, since the set of all polynomials of Waring $\le n$ has a complete polynomial, namely $X_1^d+\cdots+X_n^d$.
If we want to prove that $h\notin \overline{W}$, then we replace the orbit by the corresponding orbit closure. This has the nice effect that we can replace End_n by GL_n, which is a group (and very well understood).

4 First algebraic geometry

We explain the crucial link of our observations so far to algebraic geometry, namely that our orbit closures are actually *Zariski-closed*, so that we can restrict our search for obstructions f to separating *polynomials* f, which we call *polynomial obstructions*. This is formally stated in Definition 4.16. The proof of this insight requires a good amount of algebraic geometry and thus we only provide the necessary definitions and examples to understand the referenced theorem statements.

4.1 Zariski-closure

Recall Lemma 2.20: If a set W is $\mathbb{C}$-closed, then there is a continuous function vanishing precisely on W. Moreover, in the other direction Prop. 2.19 shows that if a continuous function vanishes precisely on W, then W is $\mathbb{C}$-closed. It follows that we could *define* $\mathbb{C}$-closed sets to be exactly those sets which can be separated from arbitrary points by continuous functions vanishing on the sets. In Definition 4.1 we use exactly this approach to define what a Zariski-closed set is: Those are the sets that can be separated from arbitrary points by multivariate polynomials vanishing on the sets.

Let $\mathbb{A} = \mathbb{C}[X_1, \dots, X_m]_d$ with $\dim \mathbb{A} = \eta$ and let $T_1, \dots, T_\eta$ be the variables corresponding to the coefficients of the monomials in $\mathbb{A}$. For example, in the discriminant setting we have $T_1(aX^2 + bXY + cY^2) = a$.

In the following, we will have two polynomials $c, h \in \mathbb{A}$. (c is not the variable appearing in the discriminant above!) h is the polynomial for which we search complexity lower bounds, for example, we want to prove a lower bound on its border Waring rank (the name "h" stands for *hard*, as it is the polynomial we want to show is hard, i. e., not in $\overline{W}$). The variable name "c" stands for *complexity* or *complete polynomial*. In the case of the Waring rank, think of c being the power sum $c = X_1^d + \cdots + X_m^d$.

Definition 4.1. A subset $W \subseteq \mathbb{A}$ (think of $W = \overline{Gc}$) is called *Zariski-closed in* $\mathbb{A}$ iff there exists a natural number $r \in \mathbb{N}$ and polynomials $f_1, \dots, f_r \in \mathbb{C}[T_1, \dots, T_\eta]$ such that

$$h \in W \;\Leftrightarrow\; f_1(h) = f_2(h) = \cdots = f_r(h) = 0.$$

We say that the polynomials $f_1, \dots, f_r$ *cut out* W or *define* W.

Example 4.2. Let $\mathbb{A} = \mathbb{C}[X, Y]_2$, so every element h can be written as $aX^2 + bXY + cY^2$. Then $h \mapsto b^2 - 4ac$ cuts out the (border) Waring rank 1 polynomials.

Example 4.3. Consider $\mathbb{A} = \mathbb{C}^2$ and let the Zariski-closed set $W \subseteq \mathbb{A}$ be cut out by the polynomial $(T_1)^2 + (T_2)^2 - 1$. Those points in W that have real coordinates form a circle with radius 1 in $\mathbb{R}^2 \subseteq \mathbb{C}^2$.

Example 4.4. If $\mathbb{A} = \mathbb{C}^{n\times n}$, then the set GL_n of singular $n \times n$ matrices (i. e., matrices with determinant zero) is Zariski-closed. It is cut out by the determinant polynomial.

Example 4.5. If $\mathbb{A} = \mathbb{C}^{n\times n}$, then the set of $n \times n$ matrices that have rank at most $n - 2$ is Zariski-closed. It is cut out by the determinants of all $(n - 1) \times (n - 1)$ submatrices.

In order to prove complexity lower bounds we would like to know a set of polynomials cutting out the set $\overline{Gc}$, but unfortunately this kind of analysis is only feasible for some very small cases.

From now on we use the short notation

$$\mathbb{C}[\mathbb{A}] := \mathbb{C}[T_1, \dots, T_\eta]$$

and call $\mathbb{C}[\mathbb{A}]$ the *coordinate ring of the ambient space*. Note that this replaces the clumsy $\mathbb{C}[\mathbb{A}] = \mathbb{C}[\,\mathbb{C}[X_1, \dots, X_m]_d\,]$.

We will see in Theorem 4.15 that our orbit closures $\overline{Gc}$ are Zariski-closed. This is perfect for our purposes, because for Zariski-closed sets non-membership of a point is equivalent to nonvanishing of a single polynomial f as the following straightforward lemma highlights. This is much stronger than separation by merely continuous functions.

Lemma 4.6. *Let W be Zariski-closed in $\mathbb{A}$. For $h \in \mathbb{A}$ we have $h \notin W$ iff there exists a polynomial $f \in \mathbb{C}[\mathbb{A}]$ such that f vanishes on W and $f(h) \neq 0$.*

Proof. Let W be cut out by $f_1, \dots, f_r$, i. e., $h \in W \Leftrightarrow f_1(h) = \cdots = f_r(h) = 0$. Then $h \notin W$ iff $\exists 1 \le i \le r : f_i(h) \neq 0$. □

The following lemma shows a first property of Zariski-closed sets, namely that they are $\mathbb{C}$-closed.

Lemma 4.7. *If $W \subseteq \mathbb{A}$ is Zariski-closed in $\mathbb{A}$, then W is $\mathbb{C}$-closed in $\mathbb{A}$.*

Proof. Let W be cut out by the polynomials $f_1, \dots, f_r$. Let $(h_i)_i$ be a sequence in W that converges to some $h \in \mathbb{A}$. Then $f_j(h_i) = 0$ for all i, j. Since the f_j are continuous, it follows that $f_j(h) = 0$. Therefore $h \in W$. We conclude that W is $\mathbb{C}$-closed. □

4.2 Algebraic geometry of orbit closures

In this section we write $G := \mathsf{GL}_N$. Lemma 4.7 says that Zariski-closed sets are $\mathbb{C}$-closed. The crucial point is that in our case the converse of Lemma 4.7 holds: Orbit closures $\overline{Gc}$ are not only $\mathbb{C}$-closed but also Zariski-closed.

Definition 4.8. A map $\mathbb{C}^a \to \mathbb{C}^b$ is called a *polynomial map* if all its b coordinate functions are multivariate polynomials in the a standard basis vectors.

Example 4.9. Using Prop. 3.21 we see that for a fixed $c \in \mathbb{A}$, the map $G \to \mathbb{A}$, $g \mapsto gc$ is a polynomial map. Its image is the orbit Gc.

Definition 4.10. We use the following definitions with respect to the Zariski topology.

1. A subset $W \subseteq \mathbb{A}$ to be *open* if its complement $\mathbb{A} \setminus W$ is closed.
2. A subset $W \subseteq \mathbb{A}$ is called *locally closed* if W is an intersection of an open and a closed set. Equivalently (for those who know a little bit of topology) W is locally closed iff W is open in its closure.
3. A subset $W \subseteq \mathbb{A}$ is called *constructible* if W is a finite union of locally closed sets. (Equivalently, the collection of constructible sets is the smallest collection that contains all closed sets and is closed under taking complements and finite unions.)

Example 4.11. The set $\{A \in \mathsf{End}_N \mid \det(A) = 0\}$ is Zariski-closed in End_N. Thus $G \subseteq \mathsf{End}_N$ is open and hence locally closed (and hence constructible).

We state the following theorem without proof.

Theorem 4.12 (Chevalley's Theorem, see e.g [51, AI.3.3 Folgerung 2] or [79, Prop. 15.4.3])**.** *The image of a constructible set under a polynomial map is again constructible.*

Corollary 4.13. *For any $c \in \mathbb{A}$ the orbit Gc is constructible.*

Proof. Combine Example 4.9 and Example 4.11. □

Remark 4.14. One can even prove that Gc is locally closed, see, e. g. [51, II.2.2 c].

We state the following result without proof.

Theorem 4.15 ([51, AI.7.2 Folgerung] or [65, Thm. 2.33])**.** *For constructible sets, Zariski closure and $\mathbb{C}$-closure coincide.*

We can use Theorem 4.15 and Lemma 4.6 to draw the following immediate crucial conclusion that states that *polynomials* can always be used to separate points in $\mathbb{A}$ from $\overline{Gc} \subseteq \mathbb{A}$. This greatly reduces the search space for obstructions and is one of the key ideas in geometric complexity theory. In fact, this insight dates back all the way to Strassen, see [76, the 2nd paragraph on p. 132].

Definition 4.16 (Polynomial Obstruction)**.** We call the polynomials f that separate h from $\overline{Gc}$ by satisfying $f(\overline{Gc}) = \{0\}$ and $f(h) \neq 0$ *polynomial obstructions*.

From our previous discussions we see that polynomial obstructions are *guaranteed to exist* if $h \notin \overline{Gc}$. The hard task is to find them.

4.3 Cones

In this section we write $G := \mathsf{GL}_N$. We will now use the additional structure of $\overline{Gc}$ being a *cone* to restrict our search for obstructions to *homogeneous* polynomials only.

Proposition 4.17. *For $f \in \mathbb{C}[\mathbb{A}]$, if $f(\overline{Gc}) = \{0\}$ and $f(h) \neq 0$, then there exists a homogeneous polynomial $f_{\text{hom}} \in \mathbb{C}[\mathbb{A}]$ such that $f_{\text{hom}}(\overline{Gc}) = \{0\}$ and $f_{\text{hom}}(h) \neq 0$.*

Note that for example the discriminant $b^2 - 4ac$ is homogeneous.

The rest of this section is devoted to formalizing the necessary background in order to prove this result.

Definition 4.18. Recall that $\mathbb{A} = \mathbb{C}^\eta$ is a complex vector space and hence is endowed with a scalar multiplication. For a vector $c \in \mathbb{A}$ and $\alpha \in \mathbb{C}$ let αc denote the scalar multiple of c. A subset $W \subseteq \mathbb{A}$ is called a *cone* if it is closed under scalar multiplication, i. e.,

$$\forall \alpha \in \mathbb{C}, c \in W : \alpha c \in W.$$

Lemma 4.19. *For any polynomial $c \in \mathbb{C}[X_1, \dots, X_N]_d$ the orbit closure $\overline{Gc}$ is a cone.*

Proof. Let $h \in \overline{Gc}$ and let $\alpha \in \mathbb{C}$ be arbitrary. Let $(h_i)_i$ be a sequence in Gc that converges to h. Let $g_i \in G$ such that $h_i = g_i c$. Choose $\beta \in \mathbb{C}$ such that $\beta^d = \alpha$ and let βg_i denote the product of the scalar β and the matrix g_i, i. e., the matrix g_i in which all entries are scaled with β. We observe that $(\beta g_i)c = \alpha(g_i c)$. Since scaling with α is continuous, the sequence $((\beta g_i)c)_i$ converges to αh and hence $\alpha h \in \overline{Gc}$. ◻

Remark 4.20. Usually in algebraic geometry one makes the transition to projective geometry whenever cones are encountered, but here, it is not necessary to do so.

Proposition 4.21. *Let $W \subseteq \mathbb{A}$ be a cone. If a polynomial $f \in \mathbb{C}[\mathbb{A}]$ vanishes on W, then all its homogeneous parts vanish on W.*

Proof. The statement is clear for $W = \emptyset$. Let $h \in W$ be arbitrary. Let f_i be the ith homogeneous part of f, i. e., $f_i(\alpha h) = \alpha^i f_i(h)$. We interpret $f(\alpha h)$ as a univariate polynomial $\tilde{f}$ in α. We have

$$\tilde{f}(\alpha) = f(\alpha h) = \sum_{i=0}^{d} f_i(\alpha h) = \sum_{i=0}^{i} \alpha^i f_i(h).$$

The coefficient of the monomial α^i in $\tilde{f}$ is $f_i(h) \in \mathbb{C}$. Since f vanishes on W and W is a cone we have that $\tilde{f}$ vanishes everywhere on $\mathbb{C}$, so $\tilde{f} = 0$. Therefore all coefficients $f_i(h)$ of $\tilde{f}$ are zero. Since $h \in W$ was arbitrary, we get that f_i vanishes on W. □

Proof of Prop. 4.17. Let $S \subseteq \mathbb{N}_{\geq 0}$ denote the finite set of degrees i such that the homogeneous degree i part of f is nonzero. Since $f(h) \neq 0$ we have that $f \neq 0$ and therefore $S \neq \emptyset$. Decompose f into its nonzero homogeneous parts $f = \sum_{i \in S} f_i$. Since $\overline{Gc}$ is a cone and $f(\overline{Gc}) = \{0\}$, using Proposition 4.21 we see that all f_i vanish on $\overline{Gc}$. Since $0 \neq f(h) = \sum_{i \in S} f_i(h)$, it follows that there exists $i \in S$ such that $f_i(h) \neq 0$. Choose f_{hom} to be such an f_i. □

For a subset $W \subseteq \mathbb{A}$ let

$$I(W) := \{f \in \mathbb{C}[\mathbb{A}] \mid f(w) = 0 \quad \forall w \in W\}$$

denote the *vanishing ideal* of W. This is clearly a complex vector space and is closed under multiplication with arbitrary polynomials, thus $I(W)$ is an ideal in the ring $\mathbb{C}[\mathbb{A}]$. Proposition 4.21 implies that if W is a cone, then $I(W)$ is a *graded* $\mathbb{C}$-algebra, i. e., every element can be decomposed into a unique sum of homogeneous parts, where each part is in $I(W)$. We denote by $I(W)_i$ the ith homogeneous part of $I(W)$. We have $I(W)_i \cdot I(W)_j \subseteq I(W)_{i+j}$. (This is the crucial property of a graded algebra.)

Search space reduction via algebraic geometry

- For orbit closures Zariski closure equals $\mathbb{C}$-closure.
  Consequence: If $h \notin \overline{Gc}$, then it is separated by polynomial (instead of simply a continuous function).
- Orbit closures are cones
  Consequence: If $h \notin \overline{Gc}$, then it is even separated by a homogenous polynomial.

5 Algebraic complexity classes

While the Waring rank is a very instructive and important example, we also want to define complexity measures that are more powerful and closer to actual computations. These are the so-called *Valiant's classes*.[3] This chapter, Chapter 6, and Chapter 7 introduce classical results

[3] They are usually all called Valiant's classes, although the more recent ones like VP_{ws} were not defined by Valiant.

about algebraic complexity classes that lead to questions about group orbit closures in Chapter 8. Our observations in this and some following chapters work over other fields $\mathbb{F}$ than the complex numbers. We will go back to the complex numbers when rephrasing Valiant's classes in terms of orbit closures in Chapter 8.

We denoted by f a separating function and by g a group element, but traditionally in the field of algebraic complexity theory both f and g denote polynomials. We stick with this classical notation.

5.1 VP

Let $X = (X_1, X_2, \dots)$ be an infinite family of indeterminates over some field $\mathbb{F}$.

Definition 5.1. A sequence of polynomials $(f_n) \in \mathbb{F}[X]$ is called a *p-family*[4] if for all n,

1. $f_n \in \mathbb{F}[X_1, \dots, X_{p(n)}]$ for some polynomially bounded function p and
2. $\deg f_n \le q(n)$ for some polynomially bounded function q.

Recall Definition 1.10.

Definition 5.2. The class VP consists of all p-families (f_n) such that $L(f_n)$ is polynomially bounded.

Example 5.3. Let $f_n = X_1^n + X_2^n + \cdots + X_n^n$. It is easy to see that $L(f_n) \le n^2$ and hence $(f_n) \in \mathrm{VP}$.

Example 5.4. Let $\det_n = \sum_{\pi \in \mathfrak{S}_n} \mathrm{sgn}(\pi) X_{1,\pi(1)} \dots X_{n,\pi(n)}$. We will see soon that $\det_n$ has polynomial-sized arithmetic circuits. Therefore, $(\det_n) \in \mathrm{VP}$.

In the above example, the indeterminates have two indices instead of one. Of course we could write $\det_n$ as a polynomial in $X_1, X_2, \dots$ by using a bijection between $\mathbb{N}^2$ and $\mathbb{N}$. However, we prefer the natural naming of the variables (and will do so with other polynomials).

Let $f \in \mathbb{F}[X]$ be a polynomial and $s : X \to \mathbb{F}[X]$ be a mapping that maps indeterminates to polynomials. s can be extended in a unique way to an algebra endomorphism $\mathbb{F}[X] \to \mathbb{F}[X]$. We call s a *substitution*. (Think of the variables replaced by polynomials.)

Definition 5.5.
1. Let $f, g \in \mathbb{F}[X]$. f is called a *projection* of g if there is a substitution $r : X \to X \cup \mathbb{F}$ such that $f = r(g)$. We write $f \le_p g$ in this case. (Since g is a polynomial, it only depends on a finite number of indeterminates. Therefore, we only need to specify a finite part of r.)
2. Let (f_n) and (g_n) be p-families. (f_n) is a *p-projection* of (g_n) if there is a polynomially bounded function $q : \mathbb{N} \to \mathbb{N}$ such that $f_n \le_p g_{q(n)}$. We write $(f_n) \le_p (g_n)$.

[4] This is a problematic naming, because it contains the singled out letter p, which can easily be mistaken as a variable. Unfortunately, the name is completely standard in the literature. "poly-family" would have been a much better name.

Remark 5.6. Definition 5.5 captures Valiant's original notion of projection. Many authors nowadays allow variables to be replaced by arbitrary linear combinations of variables and constants. This rarely makes a difference.

Projections are very simple reductions. Therefore, we can also use them to define hardness for "small" complexity classes like VP. Projections fulfill the usual requirements of a reductions:

Lemma 5.7. *(1) If* $(f_n) \leq_p (g_n)$ *and* $(g_n) \in \mathrm{VP}$*, then* $(f_n) \in \mathrm{VP}$*.*

(2) $\leq_p$ *is a transitive relation.*

Proof.

1. Let q be a polynomially bounded function and s_n be a projection such that $f_n = s_n(g_{q(n)})$ for all n. Let C_m be a circuit computing g_m. We get a circuit computing f_n by replacing every variable X_i in $C_{q(n)}$ by $s_n(X_i)$. This circuit has the same size as $C_{q(n)}$.

2. The composition of two polynomially bounded functions is polynomially bounded and the composition of two substitutions is a substitution again.

□

Definition 5.8.

1. A p-family (f_n) is called VP-*hard* (under p-projections) if $(g_n) \leq_p (f_n)$ for all $(g_n) \in \mathrm{VP}$.

2. It is called VP-*complete* if in addition $(f_n) \in \mathrm{VP}$.

Lemma 5.9. *If* (f_n) *is* VP*-hard and* $(f_n) \leq_p (g_n)$*, then* (g_n) *is* VP*-hard, too.*

Proof. Let $(h_n) \in \mathrm{VP}$ be arbitrary. Since (f_n) is VP-hard, $(h_n) \leq_p (f_n)$. By transitivity, $(h_n) \leq_p (g_n)$. Since (h_n) was arbitrary, the VP-hardness of (g_n) follows. □

Example 5.10. Let $X_{i,j}^{(\ell)}$, $1 \leq i, j, \ell \leq n$, be indeterminates and let $M_\ell = (X_{i,j}^{(\ell)})_{1 \leq i,j \leq n}$ for $1 \leq \ell \leq n$. The polynomial imm_n is a polynomial in n^3 variables and is the $(1, 1)$ entry of the matrix product $M_1 \cdots M_n$. The p-family $\mathrm{imm} = (\mathrm{imm}_n)$ is the called *iterated matrix multiplication*.

By using the trivial algorithm for matrix multiplication, it is easy to see that imm $\in$ VP. We will see in the next chapter that imm and det are equivalent under p-projections. We do not know whether the determinant (or the iterated matrix multiplication polynomial) is VP-complete. However, there are generic problems that are VP-complete. But also more natural complete problems are known, see [31, 59, 25].

Question 5.11. Is det VP-complete?

Remark 5.12. When we replace polynomial upper bounds by quasipolynomial upper bounds (of the form $O(n^{\log^c(n)})$ for constant c) in the definition of VP and p-projections, then the determinant is complete for this class usually called VQP, see [17] and [9] for more complete families. Here, "QP" stands for "quasi-polynomial".

5.2 VP_e

We call an arithmetic circuit a *formula* if the underlying graph structure is a tree. In this case, every computation gate has fanout one, that is, there is only one edge leaving every node and therefore, every intermediate result can only be used once.

Definition 5.13. A p-family (f_n) is contained in the class VP_e if there is a family of formulas (F_n) such that F_n has polynomial size in n and computes f_n.

The "e" in the subscript stands for *expression*, another word for formula. Since every formula is a circuit, we have $\mathrm{VP}_e \subseteq \mathrm{VP}$. It is not known whether this inclusion is strict, but many researchers believe it is.

Question 5.14. Is VP_e a strict subset of VP?

Corollary 5.15. *If there exists a* VP*-complete function in* VP_e*, then* $\mathrm{VP}_e = \mathrm{VP}$.

Proof. This follows from the transitivity of $\leq_p$ (Lemma 5.7(2)) and the the fact that if $(g_n) \in \mathrm{VP}_e$ and $(f_n) \leq_p (g_n)$, then $(f_n) \in \mathrm{VP}_e$, analogously to Lemma 5.7(1). ◻

Definition 5.16. 1. A p-family (f_n) is in the class VNC_i if there is a family of circuits (C_n) computing f_n such that the size of C_n is polynomially bounded in n and the depth of C_n is bounded by $O(\log^i n)$.[5]

2. $\mathrm{VNC} := \bigcup_{i\in\mathbb{N}} \mathrm{VNC}_i$.

It turns out that $\mathrm{VP}_e = \mathrm{VNC}_1$, that is, every p-family that is computable by formulas of polynomial size has efficient parallel algorithms, and that $\mathrm{VP} = \mathrm{VNC}_2$ (see [82]). We will now prove the first of these two statements.

Theorem 5.17. $\mathrm{VP}_e = \mathrm{VNC}_1$.

This theorem is a direct consequence of a depth-reduction theorem of Brent's (see Theorem 5.19). We will first prove a lemma that helps balancing binary trees.

Lemma 5.18. *Let T be a rooted binary tree with n nodes. Then there is an edge e in T such that removing e separates T into two trees both having between $n/3$ and $2n/3$ nodes.*

Proof. We construct a path $u_1, \ldots, u_m$ starting from the root as follows: We set u_1 to be the root of the tree. Let u_i be the current end node of the path and let u_{i+1} be the child of u_i that is the root of the larger subtree (if the two subtrees rooted at the children of u_i are of equal size, we may let u_{i+1} be either child). We stop when the size of the subtree with root u_i is $< 2/3n$ and set $m = i$. The edge e is the edge (u_{m-1}, u_m). The subtree with root u_m has size $< 2/3n$ by construction. The subtree with root u_{m-1} has size $\geq 2/3n$. Since u_m is the root of the larger subtree, the subtree with root u_m has size at least $n/3$. The size of the remaining tree is between $n - 2/3n = n/3$ and $n - n/3 = 2n/3$. ◻

[5] Here $\log^i n$ denotes the ith power of $\log n$ and *not* the iterated logarithm.

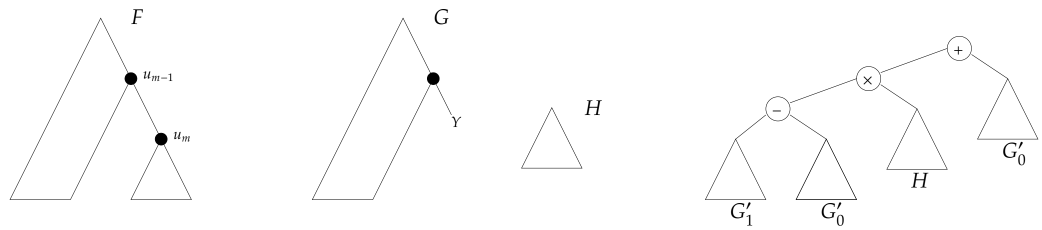

Figure 3: The formula F with the edge (u_{m-1}, u_m), the two formulas G and H, and the new formula computing f

Theorem 5.19 (Brent [14]). *Let F be a formula of size s. Then there is a formula F' of size* $\mathrm{poly}(s)$ *and depth $O(\log s)$ computing the same polynomial as F.*

Proof. By Lemma 5.18, there is an edge in F such that when removing e, we get two parts, each of size between $s/3$ and $2s/3$. The part not containing the output gate of F is again a formula, which we call H. The part containing the output gate is not a formula, since one of the gates has fanin one after removal of e. We add a new child to this gate, which is labeled with a new input variable Y. Call the resulting formula G. G computes a linear form $aY + b$, since Y appears only once in G. (a and b are polynomials in the original input variables.) If we substitute the polynomial h computed by H for Y, then we get the polynomial f computed by F. If we substitute 1 for Y, then we get $a + b$, and if we substitute 0 for Y, then we get b. Therefore, there are formulas of size $\leq 2s/3$ computing $a + b$, b and h. With these formulas, we can proceed recursively. We get formulas G'_1, G'_0, and H' computing $a + b$, b, and h, respectively. We can combine them to a formula computing $ah + b = f$ as depicted in Figure 3: For the size $\sigma(s)$ of this new formula, we get the recursion

$$\sigma(s) = 4 \cdot \sigma(2s/3) + 3$$

and for the depth $d(s)$, we get the recursion

$$d(s) = d(2s/3) + 3.$$

It is a routine check that $\sigma(s) = \mathrm{poly}(s)$ and $d(s) = O(\log s)$. □

5.3 Constant size iterated matrix multiplication

For some $c \in \mathbb{N}$, we define the family $(\mathrm{imm}_n^{(c)})$ like the family (imm_n), except that every polynomial is an iterated matrix product of $c \times c$-matrices (instead of $n \times n$-matrices), so $\mathrm{imm}_n^{(c)}$ is a polynomial in $c^2 n$ variables.

Theorem 5.20 (Ben-Or & Cleve [6]). *Let F be a formula of depth d computing a polynomial f, then f is a projection of* $\mathrm{imm}_{4^d}^{(3)}$.

Proof. We will prove by induction on d that we can find 4^d many 3×3-matrices whose entries are either indeterminates or constants such that the product of these matrices is

$$\begin{pmatrix} 1 & f & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that f is not in position $(1,1)$. But we can move it into the $(1,1)$-entry at the end by applying appropriate permutation matrices. The statement above is obviously true for depth zero formulas, since these formulas compute constants or single variables.

If the depth d is larger than zero, we either have $f = g + h$ or $f = gh$ and g and h are both computed by formulas of depth $\le d - 1$. By the induction hypothesis, there are two sets of 4^{d-1} 3×3-matrices each such that their products are

$$\begin{pmatrix} 1 & g & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

respectively. In the case of an addition gate we have

$$\begin{pmatrix} 1 & g & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & g+h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Therefore we can write f as a projection of a 3×3-iterated matrix multiplication of length $2 \cdot 4^{d-1} \le 4^d$.

In the case of a multiplication gate, we have

$$\begin{pmatrix} 1 & g & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & g & gh \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that h is standing in the "wrong" position. But we can easily fix this by applying permutation matrices from the left and the right. This just corresponds to exchanging the rows or columns of the first and last matrix of the corresponding matrix product, respectively. We proceed with

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -h \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & g & gh \\ 0 & 1 & h \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -g & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & gh \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Note that we now have a $-g$ and $-h$ instead of a g and h. But this can be easily fixed by multiplying the second row and column by -1, which can be achieved by adjusting the first and last matrix of the 4^{d-1} matrices. Altogether, we get that f is a projection of a product of $4 \cdot 4^{d-1} = 4^d$ matrices. □

Corollary 5.21. $\mathrm{imm}^{(3)}$ *is* VP_e*-complete.*

Proof. Let $f = (f_n) \in \mathrm{VP}_e$. Let F_n be a formula of polynomial size computing f_n. By Theorem 5.19, there is an equivalent formula of polynomial size and depth $O(\log n)$. By Theorem 5.20, f_n is a projection of $\mathrm{imm}^{(3)}_{\mathrm{poly}(n)}$. This proves the hardness.

To construct a formula of polynomial size for $\mathrm{imm}^{(3)}_n$, we recursively construct formulas for the nine entries of the product $M_1 \cdots M_n$. W.l.o.g. we can assume that n is a power of 2. To this aim, we divide the product into two products $M_1 \cdots M_{n/2}$ and $M_{n/2+1} \cdots M_n$ of size $n/2$ each. We need two recursive calls of half the size to compute these products. From these two results, we can compute the product $M_1 \cdots M_n$ by a constant size formula, a single matrix multiplication. Since each entry of the two resulting matrices is used three times, we need three distinct copies of the formulas for each entry. Therefore, we get the following recursion for the overall size $s(n)$ of the formulas:

$$s(n) = 6s(n/2) + O(1).$$

Therefore, $s(n) = \mathrm{poly}(n)$. □

Obviously, $\mathrm{imm}^{(c)}$ is VP_e-complete for any $c \geq 3$. On the other hand, $\mathrm{imm}^{(1)}$ is not, since it only computes a single monomial. Allender and Wang [4] prove that $\mathrm{imm}^{(2)}$ is also not VP_e-complete by exhibiting a polynomial that is not the projection of $\mathrm{imm}^{(2)}_n$ for any n! When we consider approximate computations, however, then $\mathrm{imm}^{(2)}_n$ is complete [15] (analogous to border Waring rank, see Chapter 2. Also see Chapter 8).

5.4 Orbit problems

Historically, p-projections have been the reduction of choice in algebraic complexity theory, because they are very simple reductions yet sufficiently powerful to prove hardness results (as we will see in the next two chapters).

Let (f_n) be a p-family. Then there is a polynomially bounded function p such that $f_n \in \mathbb{F}[X_1, \dots, X_{p(n)}]$. We saw how to let End_m act on $\mathbb{F}[X_1, \dots, X_m]$. We let endomorphisms act on the sequence (f_n) by letting $\mathsf{End}_{p(n)}$ act on f_n. An element $g \in \mathsf{End}_{p(n)}$ replaces each variable by a homogeneous linear combination of the variables. In particular, it preserves the degree, that is, $\deg g f_n = \deg f_n$. P-projections, on the other hand, do not preserve the degree. We can use the following "trick" called padding. We only need it for homogeneous polynomials, since all our polynomials under consideration will be homogeneous, but it also works for non-homogeneous polynomials in the obvious way. Assume that $a \leq_p b$ for two homogeneous polynomials a and b and let s be the corresponding substitution. We define a new substitution $\hat{s}$ that whenever $s(X_i)$ is a constant α, then we set $\hat{s}(X_i) = \alpha T$ for some new indeterminate T instead. When $s(X_i)$ is a variable, then $\hat{s}(X_i) = s(X_i)$. We have $\hat{s}(b) = T^{\deg b - \deg a} \cdot a$. Note that $\hat{s}$ is a very special endomorphism, replacing every variable by a scalar multiple of some other variable. Assume $b \in \mathbb{F}[X_1, \dots, X_n]$, then we will now consider it as a polynomial in $\mathbb{F}[T, X_1, \dots, X_n]$ and let End_{n+1} act on it. Since T does not appear in b, we can restrict ourselves to endomorphisms that leave T fixed.

Definition 5.22. Let (f_n) and (h_n) be homogeneous p-families. Let $p(m)$ be minimal such that that $h_m \in \mathbb{F}[X_1, \ldots, X_{p(m)}]$. We write $(f_n) \leq_{end} (h_n)$ if there is a polynomially bounded function q such that

$$T^{\deg h_{q(n)} - \deg f_n} \cdot f_n \in \mathsf{End}_{p(q(n))+1} h_{q(n)}.$$

Note that since p is minimal, it is polynomially bounded by the definition of p-family. We have chosen a fresh variable T for padding. In the literature, X_1 or $X_{p(q(n))}$ has frequently been used for padding (but also arbitrary linear forms have been used). Taking a new variable turns out to be simpler, in particular, $\leq_{end}$ will be transitive. We will see below that in our situation, it actually does not matter which variable we will take.

Lemma 5.23. *Let (f_n) and (h_n) be homogeneous p-families.*

1. *If $(f_n) \leq_p (h_n)$, then $(f_n) \leq_{end} (h_n)$.*

2. *If $(f_n) \leq_{end} (h_n)$ and $(h_n) \in \mathrm{VP}$, then $(f_n) \in \mathrm{VP}$. The same statement is true if VP is replaced by VP_e.*

3. *$\leq_{end}$ is transitive.*

Proof. Let p and q be defined as in Definition 5.22.

1. Let s_n be the substitution mapping f_n to $h_{q(n)}$. As above, we define $\hat{s}_n$ to be the substitution that whenever $s_n(X_i) = \alpha$, then $\hat{s}_n(X_i) = \alpha T$. Then $\hat{s}_n(h_n) = T^{\deg h_{q(n)} - \deg f_n}$. The substitution obviously defines an endomorphism.

2. It is very easy to see that for $g_n \in \mathsf{End}_n$, the sequence $(g_{p(n)+1} h_n)$ is in VP or VP_e, respectively, since $g_{p(n)+1}$ is just a linear transformation of the variables, which can be implemented by formulas of polynomial size. Therefore, the sequence $(T^{\deg h_{q(n)} - \deg f_n} \cdot f_n)$ is in VP or VP_e, respectively. Since T does not appear in f_n, we can set $T = 1$.

3. Let $(f_n) \leq_{end} (h_n)$ and $(h_n) \leq_{end} (a_n)$. There is an endomorphism $g_{p(q(n))+1}$ such that

$$T^{\deg h_{q(n)} - \deg f_n} \cdot f_n = g_{p(q(n))+1} h_{q(n)}. \tag{5.1}$$

Furthermore, there are polynomially bounded functions p' and q' such that

$$T^{\deg a_{q'(m)} - \deg h_m} \cdot h_m \in \mathsf{End}_{p'(q'(m))+1} a_{q'(m)}.$$

Thus, there is an endomorphism $g'_{p'(q'(m))+1}$ such that

$$T^{\deg a_{q'(m)} - \deg h_m} \cdot h_m = g'_{p'(q'(m))+1} a_{q'(m)}. \tag{5.2}$$

We set $m = q(n)$, apply $g_{p(q(n))+1}$ to (5.2), and plug in (5.1), where we interpret $g_{p(q(n))+1}$ as an endomorphism in $\mathsf{End}_{p'(q'(q(n))+1}$ (setting further variables to zero for instance). Therefore,

$$T^{\deg a_{q'(q(n))} - \deg f_n} \cdot f_n = g_{p(q(n))+1} g'_{p'(q'(q(n))+1} a_{q'(q(n))},$$

since we can assume that $g_{p(q(n))+1}(T) = T$.

□

The rest of this section is rather technical. It can be skipped at a first reading. We deal with the question under what conditions it does not matter to take a fresh variable or an existing one for padding. We assume $\mathbb{F}$ to be large enough.

The next lemma shows that it does not matter too much, whether we take a fresh variable T for padding or an existing one. We generalize ideas by Ikenmeyer and Panova [47].

Lemma 5.24. *Let $f, h \in \mathbb{F}[X_1, \dots, X_m]$ be homogeneous polynomials.*

1. *If $T^{\deg h - \deg f} f \in \mathsf{End}_{m+1} h$, then $X_i^{\deg h - \deg f} f \in \mathsf{End}_m h$ (where we interpret h in the first equation as a polynomial in $\mathbb{F}[T, X_1, \dots, X_m]$).*
2. *If $X_i^{\deg h - \deg f} f \in \mathsf{End}_m h$ for some i, then there is a circuit of size polynomial in $L(h)$, $\deg h$, and m computing f.*

Proof. 1. Since T does not appear in h, we can simply replace T by X_i.

2. Write $f = \sum_{j=0}^{\deg f} X_i^j f_j$. As above, there is a circuit of size polynomial in $L(h)$ and m computing $X_i^{\deg h - \deg f} f = \sum_{j=\deg h - \deg f}^{\deg h} X_i^j f_j$. We now take $\deg h + 1$ copies of this circuit, and in each of them, we plug in a different value from $\mathbb{F}$ for X_i. From the results, we can obtain the polynomials $f_0, \dots, f_{\deg f}$ by interpolation (see, e. g.[19]). Once we have these polynomials, we can easily compute f.

□

Let C be a class of p-families. We call C *closed under interpolation* if for every $(f_n) \in C$, $f_n = \sum_{j=0}^{\deg f_n} X_1^j f_{n,j}$ with $f_{n,j} \in \mathbb{F}[X_2, X_3, \dots]$, the family $(\sum_{j=0}^{\deg f_n} Y_j f_{n,j}) \in C$, where $Y_0, Y_1, Y_2, \dots$ are new variables. So essentially, given f_n and considering it as a univariate polynomial in X_1, we can compute the coefficients of f_n. The new variables Y_i are introduced for book-keeping purposes to have again only one polynomial.

We call the class C *closed under substitutions*, if for two p-families $(f_n), (h_n) \in C$, the family obtained by substituting some of the variables X_i in f_n by $h_{j(i)}$ for some p-bounded function j is again in C.

The next lemma is a strengthening of Lemma 5.24, part (2).

Lemma 5.25. *Let C be a class of p-families that is closed under interpolation and substitutions. Let $(f_n), (h_n) \in C$ such that*

1. *$X_i^{\deg h_{q(n)} - \deg f_n} f_n \in \mathsf{End}_{p(q(n))} h_{q(n)}$ for some i (here p and q are as in Definition 5.22),*
2. *every sequence (m_i) of monomials of p-bounded degree is in C, and*
3. *(h_n) is C-hard (under p-projections).*

Then there is a polynomially bounded function q' such that $T^{\deg h_{q'(n)} - \deg f_n} f_n \in \mathsf{End}_{p(q'(n))+1} h_{q'(n)}$.

Proof. Write $X_i^{\deg h_{q(n)} - \deg f_n} f_n = \sum_{j=0}^{\deg f_n} X_i^{\deg h_{q(n)} - \deg f_n + j} f_{n,j}$, where $f_{n,j}$ does not depend on X_i. Since C is closed under interpolation, the family $(\sum_{j=\deg h_{q(n)} - \deg f_n}^{\deg h_{q(n)}} Y_j f_{n,j}) \in C$, too. Renaming the indices we see that also $(\sum_{j=0}^{\deg f_n} Y_j f_{n,j}) \in C$. Now consider the family $(T^{\deg h_{q(n)} - \deg f_n} X_i^j)$ and substitute it for the Y-variables into the former family. We get that $(T^{\deg h_{q(n)} - \deg f_n} f_n) \in C$. Now since (h_n) is C-complete, the statement of the lemma follows. □

Valiant's classes

Objects: families of polynomials of polynomial degree
Computational model: Arithmetic circuits
Reductions: p-projections

- VP is characterized by circuits of polynomial size.
- VP_e is characterized by formulas of polynomial size.
- $\mathrm{imm}^{(3)}$ is complete for VP_e.

6 VP and the determinant

Constant size iterated matrix multiplication $\mathrm{imm}^{(3)}$ is complete for VP_e. For VP, we do not know a nice complete polynomial, but the first natural families have been found in [31, 59, 25]. For instance, it is not known whether general iterated matrix multpliation or the determinant are complete for VP. We first prove some normal forms for circuits for VP. Then we look for a subclass of VP such that the determinant and iterated matrix multiplication are complete for it.

6.1 Homogeneous circuits

Recall that a polynomial is homogeneous if all its monomials have the same total degree. A circuit is called *homogeneous* if at every gate it computes a homogeneous polynomial. Of course, nonhomogeneous polynomials cannot be computed by homogeneous circuits. However, we have the following result.

Lemma 6.1. *If f is a polynomial of degree d that is computed by a circuit of size s, then there is a homogeneous circuit of size $O(d^2 s)$ computing the homogeneous parts of f. Furthermore, at every gate we only compute a polynomial of degree at most d.*

Proof. We replace every gate g by $d + 1$ gates. If g computes a polynomial f, then the new gates will compute the homogeneous components of f. We do this in a bottom up fashion. If g is an input gate, then there is nothing to do. We just have to add d dummy gates computing the zero polynomial in every degree other than degree zero or one. Let g be a gate with children h_1 and h_2 in the original circuit. Assume that h_1 and h_2 have been replaced by gates $h_{1,0}, \dots, h_{1,d}$

and $h_{2,0}, \dots, h_{2,d}$ computing polynomials $p_{1,0}, \dots, p_{1,d}$ and $p_{2,0}, \dots, p_{2,d}$, respectively. If g is an addition gate, then we will introduce new gates $g_0, \dots, g_d$ and g_i computes $p_{1,i} + p_{2,i}$. If g is a multiplication gate, then g_i computes $\sum_{j=0}^{i} p_{1,j} p_{2,i-j}$. We only need to compute the homogeneous parts of degree d, since potential higher degree parts in intermediate results will cancel each other in the end, as f has degree d. ◻

Corollary 6.2. *If f is a polynomial of degree d that is computed by an arithmetic circuit of size s, then there is a circuit C of size* $\mathrm{poly}(s, d)$ *computing f such that every node in C computes a polynomial of degree at most d. Furthermore, for every multiplication gate, at least one of the inputs is not a constant.*

Proof. We homogenize the given circuit as above. This immediately gives the upper bound on the degree. When two constants are multiplied, then either two degree zero components are multiplied or one of the higer degree homogeneous parts became zero. In the first case, we can replace the multiplication gate by an input gate labeled with the product of the two constants. (Remember that we can use every constant from $\mathbb{F}$.) In the second case, we simply can remove the gate that outputs 0. (Note that we do not have to construct the circuit, we just need to prove it existence.) ◻

6.2 Multiplicatively disjoint circuits

Families in VP are computed by polynomial size circuits. In this section, we show that we can assume that these circuits have an additional property: they are multiplicatively disjoint [61].

Definition 6.3. An arithmetic circuit is *multiplicatively disjoint* if for all multiplication gates, the subcircuits induced by its two children are disjoint.

Multiplicatively disjoint circuits are between circuits and formulas. In a formula, the subcircuits of addition gates are also disjoint. Note that in a multiplicative disjoint circuit, only the induced subcircuits are disjoint. Nodes of these subcircuits can be connected to arbitrary nodes outside these circuits.

Definition 6.4. Let C be an arithmetic circuit. The formal degree of a gate g is defined inductively: A leaf has formal degree 1 (even if it is labelled with a constant). If g is a multiplication gate, then its formal degree is the sum of the formal degrees of its two children. If g is an addition gate, then the formal degree of g is the maximum of the formal degrees of its children. The formal degree of C is the formal degree of its output gate.

The formal degree of a circuit disregards that the degree at gate might drop when there are cancellations. Multiplications with constants might also increase the formal degree.

If we have a bound on the formal degree of a circuit, we can turn the circuit into a multiplicatively disjoint one without increasing the size by too much. The naive approach would be to work from the inputs towards the output, and whenever a multiplication gate is encountered that does not have disjoint children, make a separate copy of the subcircuit rooted at one of the children to force them to be disjoint. In general, the increase in size will be exponential in the degree. Using depth reduction, one can thus get a multiplicatively disjoint circuit of quasi-polynomial size from any poly-size circuit with polynomial degree bound. But since we have a bound on the formal degree, we can do better using a clever indexing technique.

Lemma 6.5. *If a circuit has size s and formal degree d, then there is a multiplicatively disjoint circuit C' of size $\leq sd$ computing the same polynomial.*

Proof. Each gate g of formal degree $e \leq d$ will be replaced by $d+1-e$ copies $g_1, \dots, g_{d+1-e}$. Let g_i be one of these copies. We call i the index of the copy. We will make sure that all gates of the subcircuit with output g_i are copies with an index lying between i and $i+e-1$. In this way we ensure that we will get multiplicatively disjoint circuits.

Inductively, we construct a circuit C_e with the following property: For each gate g of formal degree $f \leq e$ in C, there are copies of the gates $g_1, \dots, g_{d+1-f}$ in C_e computing the same function as g and all the gates of the subcircuit with root g_i have indices lying between i and $i+f-1$.

The nodes of formal degree one are all input nodes and sums of formal degree one nodes. C_1 consists of d copies of the formal degree 1 nodes. Since C is acyclic, we can order the addition gates in such a way, that whenever we deal with a gate g, all its predecessors have been processed. For each addition gate g of formal degree one, we add copies $g_1, \dots, g_d$. Let g' and g'' be the children of g in C with formal degrees one. We connect g_i with the copy g'_i and g''_i. The restriction on the ranges is fulfilled by construction.

Assume that we constructed C_{e-1} (induction hypothesis). To obtain C_e, we now add copies of all gates g of formal degree e in C. Let g' and g'' be the children of such a gate g of formal degrees e' and e'', respectively.

We start with the multiplication gates. In this case $e = e' + e''$ with $e', e'' < e$. This means that the copies $g'_1, \dots, g'_{d+1-e'}$ and $g''_1, \dots, g''_{d+1-e''}$ were constructed in a previous step. We add the copies $g_1, \dots, g_{d+1-e}$ and connect g_i with g'_i and $g''_{i+e'}$. These copies exist, since $i \leq d+1-e \leq d+1-e'$ and $i+e' \leq d+1-e+e' = d+1-e''$. The indices of the copies of the subcircuit with root g'_i lie between i and $i+e'-1$, the indices of the copies in the subcircuit with root $g''_{i+e'}$ lie between $i+e'$ and $i+e'+e''-1 = i+e-1$. The two subcircuts of g_i are disjoint, because they contain gates with indices from two disjoint intervals. Therefore the condition on the indices of the subcircuits is fulfilled.

Next come the addition gates of formal degree e. Note that an addition gate of formal degree e might have a predecessor of formal degree e. As in the base case, we can order the addition gates in such a way, that whenever we deal with a gate g, all its predecessors have been processed. For each addition gate g of formal degree e, we add copies $g_1, \dots, g_{d+1-e}$. Let g' and g'' be the children of g in C with formal degrees $e' \leq e$ and $e'' \leq e$, respectively. We connect g_i with the copy g'_i and g''_i. The indices of the copies in these subcircuits lie in the range from i to $i+e'-1 \leq i+e-1$ and $i+e''-1 \leq i+e-1$, respectively.

The circuit C_d is the circuit we are looking for. It contains a copy of the output gate of C. The circuit is multiplicatively disjoint by the way we chose the indices when connecting the copies of the children to the multiplication gate. ◻

Lemma 6.6. *Let f be a polynomial of degree d computed by a circuit C of size s. Then there is a circuit of size polynomial in d and s computing f such that its formal degree is bounded by $sd+1$.*

Proof. Let C be the given circuit and C' be the circuit constructed in Corollary 6.2. Recall that the circuit C' is a simulation of the circuit C. Every node is replaced by $d+1$ nodes, one for each homogeneneous component. Then every operation in C is simulated by several operations in C'. Let the depth of a gate in C be the length of a longest path from any leaf to this gate. The depth of the nodes in C' that compute the homogeneous components is defined as the depth of their corresponding node in C. We do not define depth for the other nodes in C'. We will now prove by induction on the depth that the formal degree of any gate g of C' of depth δ computing a homogeneous component of degree i is bounded by $\delta \cdot i + 1$. For the base case note that every leaf has formal degree 1. Now let g be a gate in C' of depth δ computing a homogeneous component of degree i. If $i = 0$, then note that g has formal degree 1 by construction. So we assume that $i \geq 1$. We first treat the case when g corresponds to an addition gate in C. In C', g is an addition gate, its two inputs are gates g' and g'' both computing homogeneous polynomials of degree i. The formal degree of these two gates is bounded by $\delta' \cdot i + 1$ and $\delta'' \cdot i + 1$ where δ' and δ'' are the depth of g' and g'', respectively. The formal degree of g is $\max\{\delta' \cdot i + 1, \delta'' \cdot i + 1\} \leq \delta \cdot i + 1$.

If g is a multiplication gate, then

$$g = \sum_{j=0}^{i} g'_j g''_{i-j}$$

where g'_j and g''_{i-j} are the homogeneous components of the predecessors of g. By the induction hypothesis, the formal degrees of g'_j and g''_{i-j} are bounded by $\delta' j + 1$ and $\delta''(i-j)+1$, respectively. The formal degree of $g'_j g''_{i-j}$ is bounded by $\delta' j + 1 + \delta''(i-j) + 1 \leq \delta i + 1$, when $0 < i < j$. Note for the upper bound that $\delta > \delta', \delta''$ and $i \geq 1$. The formal degree of $g'_0 g''_i$ is bounded by $1 + \delta'' i + 1 \leq \delta i + 1$. The same argument works for $g'_j g''_0$. This concludes the inductive step.

From the claim the bound on the formal degree of the new circuit follows immediately. □

Theorem 6.7 (Malod & Portier [61]). *A p-family (g_n) is in* VP *if and only if there is a family of polynomial size multiplicative disjoint circuits (C_n) computing (g_n).*

Proof. If $(g_n) \in$ VP, then by Lemma 6.6, there is a sequence of circuits (C_n) of size poly(n) computing (g_n) such that the formal degree of C_n is polynomially bounded. Now we can apply Lemma 6.5.

For the other direction, note that it can be easily proven by induction that the degree of a multiplicatively disjoint circuit of size s is bounded by s. □

Circuits for VP

The following models characterise VP:

- arithmetic circuits of polynomial size (and the degree of the family is polynomially bounded)
- multiplicative disjoint circuits of polynomial size
- homogeneous circuits of polynomial size (when the family is homogeneous of polynomial degree)

6.3 Combinatorial interpretation of the determinant

Interpretation via cycle covers: Let $M = (m_{i,j})$ be an $n \times n$ matrix. We can interpret M as the weighted adjacency matrix of some directed graph over the node set $\{1, \ldots, n\}$. For every (i, j), there is an edge (i, j) of weight $m_{i,j}$. A *cycle cover* in a directed graph is a collection of node-disjoint directed cycles such that every node is contained in exactly one cycle. Permutations in $\mathfrak{S}_n$ stand in a one-to-one correspondence with cycle covers. Every permutation σ yields a cycle cover consisting of the edges $(i, \sigma(i))$. On the other hand, the edges of a cycle cover encode a permutation of the nodes with the intepretation that an edge (i, j) means that i is mapped to j. Note that this is precisely the cycle decomposition of a permutation. The sign of the permutation is -1 if the number of even-length cycles is odd, and 1 if it is even. The weight $w(C)$ of a cycle cover C is the product of the weights of the edges in it. Therefore,

$$\det(M) = \sum_{\text{cycle covers } C} (-1)^{n+\text{number of cycles in } C} w(C)$$

Conceptually, it is often easier to think of an edge of weight zero as not being present in the graph. Since the weight of a cycle cover is the product of its edge weights, this does not make any difference in the above equation for $\det(M)$.

Interpretation via bipartite matchings: Instead of interpreting M as the adjacency matrix of some directed graph, we can also interpret it as the adjacency matrix of some bipartite graph. We have nodes $\{1, \ldots, n\}$ on the lefthand side and "copies" $\{1', \ldots, n'\}$ on the other side. For every (i, j), there is an edge $\{i, j'\}$ with weight $m_{i,j}$. A *matching* N in a graph is a set of edges such that every node is incident with at most one edge from N. It is called *perfect*, if every node is incident with exactly one edge from N. Permutations in $\mathfrak{S}_n$ stand in a one-to-one correspondence with perfect matchings in bipartite graphs: Every permutation σ yields a perfect matching consisting of the edges $\{i, \sigma(i)'\}$. This construction can be reversed. If we set the sign of a perfect matching in a bipartite graph to be the sign of the corresponding permutation, we get the following expression:

$$\det(M) = \sum_{\text{perfect matchings } N} \operatorname{sgn}(N) w(N).$$

The weight $w(N)$ is the product of the weights of the edges in N.

6.4 Weakly skew circuits and algebraic branching programs

A *weakly connected component* in a digraph is defined as a connected component in the corresponding undirected graph that is obtained from forgetting the edge directions.

Definition 6.8. A circuit is called *weakly skew* if every multiplication gate g has at least one child g' such that after removing the edge (g', g), the graph consists of two weakly connected components.

In a formula, this is true for every child of a gate. In a formula, no intermediate result is reusable, that is, the output of every gate can only be used as the input of exactly one other gate. In a weakly skew circuit, one child of every gate can be reused, but not both. Weakly skew is however stronger than multiplicatively disjoint, since in the latter case, while the subcircuits need to be disjoint, they can be connected to the rest of the circuit.

Definition 6.9. Let $\mathbb{F}$ be a field and $X_1, \dots, X_n$ be indeterminates.

1. An *algebraic branching program A* is an acyclic graph with two distinguished nodes s and t and an edge labeling with labels from $\mathbb{F} \cup \{X_1, \dots, X_n\}$.[6]
2. The weight $w(P)$ of a path P from s to t is the product of the labels of the edges in the path.
3. The polynomial computed by A is
$$\sum_{s\text{-}t \text{ path } P} w(P).$$
4. The size of an algebraic branching program is the number of edges in it.
5. A is called *layered* if for every node v in A, all s-v paths have the same length.

If A is layered, then we can think of the nodes of A being grouped into layers: two nodes u and v are in the same layer i if the lengths of all paths from s to u and from s to v is i. In a layered branching program, edges only go from one layer to the next.

Lemma 6.10. *Let A be a branching program of size s. Then there is a layered branching program of size $O(s^2)$ computing the same function.*

Proof. For a node v in the branching program, let $d(v)$ be the length of a longest path from s to v. The node v will be put into layer $d(v)$. By construction, for every edge $\{u, v\}$, we have $d(u) < d(v)$. Therefore, we only have edges from layers with smaller index to larger index. If there is an edge e from layer i going to layer j with $i + 1 < j$, then we replace this edge by a path of length $j - i$ and put the nodes of this path into the layers inbetween. One (arbitrary) edge of the path gets the weight of e and all other edges get the weight 1. □

[6] Some authors allow affine linear forms, but this will not make any difference.

We formalize the notion of being *reusable*. Intuitively, a gate in a weakly skew circuit is reusable if it is not in the subcircuit of a multiplication gate that is not connected to the rest of the circuit.

Definition 6.11. Let C be a weakly skew arithmetic circuit. The set of *reuseable* gates in C is inductively defined as follows: Every gate of outdegree zero is reusable. (We consider circuits with multiple output gates to simplify some proofs in the following.) We remove every gate g of outdegree zero from C and for each such multiplication gate, we also remove the subcircuit of that child g' that is only connected to the rest of the circuit via the edge (g', g). If both children of g have this property, we only remove one child. (It does not matter which one.) Let C' be the resulting circuit. Every gate that is reusable in C' is reusable in C, too.

Theorem 6.12. *Let $f \in \mathbb{F}[X_1, \dots, X_n]$ with $\deg f = \mathrm{poly}(n)$. The following statements are equivalent:*

1. *f is computed by a weakly skew circuit of size $\mathrm{poly}(n)$.*
2. *f is computed by an algebraic branching program of size $\mathrm{poly}(n)$.*
3. *f is a projection of $\mathrm{imm}_{p(n)}$ for some polynomially bounded function p.*
4. *f is a projection of $\det_{p(n)}$ for some polynomially bounded function p.*

Proof. (1) $\Rightarrow$ (2): Assume that f is computed by a weakly skew circuit C of size m. We now prove by induction on m that there is a algebraic branching program computing A of size $\leq 2m$ such that for every reusable gate g in C there is a node v_g such that the sum of the weights of all paths from s to v_g is the same polynomial as computed at g. The construction is illustrated in Figure 4.

Let g be some output node. If g is also an input node, then A consists of a single edge. (This is the induction basis.)

For the induction hypothesis, assume that g is not an input gate. If g is an addition gate, then we remove g from C, let C' be the resulting circuit. By the induction hypothesis, there is an algebraic branching program A' such that for every gate g' that is reusable in C', there is a node $v_{g'}$ in C' such that the sum of the weights of all path from s to $v_{g'}$ equals the polynomial computed at g'. Let h and h' be the children of g. They are both reusable. We add a new node v_g and connect the nodes v_h and $v_{h'}$ to it. Both edges get weight one. If $h = h'$, then we add only one edge with weight two. By construction, the sum of the weights of all paths from s to v_g is the sum of the polynomials computed at h and h'. The resulting algebraic branching program has two more edges than A'. For all reusable nodes g' of C', the node $v_{g'}$ is still present in A and the sum of the weights of all path from s to $v_{g'}$ equals the polynomial computed at g'.

If g is a multiplication gate, then after removal of g, we get two separate circuits C_1 and C_2. Let g_1 and g_2 be the children of g. Only the gates of one of them, say C_2, can be reusable in C. Let m_1 and m_2 be the sizes of C_1 and C_2. From the induction hypotheses, we get corresponding algebraic branching programs A_1 and A_2 with sources s_1 and s_2. In A_1, there are vertices s_1 and v_{g_1} such that the sum of the weights of all path from s_1 to v_{g_1} equals the polynomial computed at g_1. We connect the node v_{g_2} in A_2 with the node s_1 of A_1 by an edge of weight 1. Then the sum of the weights of all path from s_2 to v_{g_1} is the product computed at g. For all gates h in C_2, the sum of the weights of all paths from s_2 to v_h paths equals the polynomial computed at h. The size of the new branching program is $2m_1 + 2m_2 \leq 2m$.[7]

(2) $\Rightarrow$ (3): Let A be an algebraic branching program computing f. By Lemma 6.10 we can assume that A is layered. Let ℓ be the maximum size of a layer and let m be the number of layers. We order the nodes in each layer arbitrarily. We will inductively construct $\ell \times \ell$-matrices $M_1, \dots, M_m$ with entries from $\mathbb{F} \cup \{X_1, \dots, X_n\}$ such that the first row of $M_1 \cdots M_i$ are the polynomials computed at the nodes in the ith layer, that is, the sum of the weights of all path from s to each node in this layer. M_1 has a 1 in position $(1, 1)$ and zeros everywhere else. This single 1 corresponds the the source node s. Assume we constructed $M_1, \dots, M_i$. Let $(a_1, \dots, a_\ell)$ be the first row of $M_1 \cdots M_i$. A node v in the $(i + 1)$th layer receives edges from the nodes of the ith layer. Let $(b_1, \dots, b_\ell)$ be the labels of these edges (if an edge is not present, the corresponding $b_j = 0$.) The polynomial computed at v is given by

$$(a_1, \dots, a_\ell) \cdot \begin{pmatrix} b_1 \\ \vdots \\ b_\ell \end{pmatrix}.$$

The matrix M_{i+1} simply consists of the corresponding columns $(b_1, \dots, b_\ell)^T$. If the $(i + 1)$th layer has less than ℓ nodes, we append zero rows to M_{i+1}.

Since we can embed a product of m $\ell \times \ell$-matrices into a product of d $d \times d$-matrices with $d = \max\{m, \ell\}$, we get that f is a projection of $\mathrm{imm}_{\mathrm{poly}(n)}$.

(3) $\Rightarrow$ (4): Note that an iterated matrix product can be easily computed by a layered algebraic branching program, you just have to "reverse" the construction of the previous step. Therefore it suffices to prove that every polynomial that is computed by a layered algebraic branching program A is a projection of a determinant of polynomial size. We modify A as follows: add an edge of weight one from t to s and add a self loop of weight one to every node except s and t. Let M be the weighted adjacency matrix of this modified program A'. $\det(M)$ is the sum of the weights of all cycle covers in A'. All cycle covers in A' consist of one big cycle through s and t and the remaining nodes are covered by self-loops. Since the program is layered, all cycle covers have the same number of cycles and therefore the same sign. The weight of a cycle cover equals the weight of the corresponding path from s to t, potentially with an opposite sign (but this sign is the same for all cycle covers). Therefore, f is a projection of a determinant of a polynomially large matrix.

(4) $\Rightarrow$ (1): One way to evaluate the determinant by a weakly skew circuit is known as Csanky's algorithm [30]. Another one is due to Mahajan and Vinay [60]. □

Remark 6.13. For a polynomial f, the smallest n such that f can be written as a projection of $\det_n$ is called the *determinantal complexity* $\mathrm{dc}(f)$ of f.

[7] This construction does not work if the circuit is only multiplicatively disjoint, since in this case, while the subcircuits of every multiplication gate are disjoint, they might both be connected to the rest of the circuit. However, the nodes of A_1 cannot be used any more, once s_1 is identified with v_{g_2}.

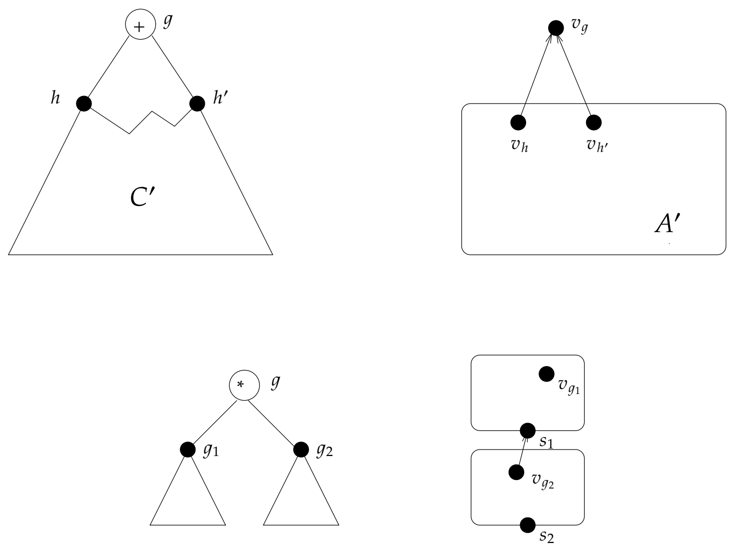

Figure 4: Transforming weakly skew circuits into algebraic branching programs. (Top: addition gate, bottom: multiplication gate)

Definition 6.14. A p-family (f_n) is in VP_{ws} if it is computed by weakly skew circuits of polynomial size.

Theorem 6.12 gives us further, equivalent definitions of VP_{ws}. In particular, a p-family (f_n) is in VP_{ws} if it is a p-projection of the determinant family. Note that imm can be computed by very restricted weakly skew circuits, namely for every multiplication gate, one of the inputs is a variable or a constant. We call such circuits *skew*. This is achieved by sequentially multiplying the matrices using the trivial methods. Since by Theorem 6.12, every polynomial that is computed by a weakly skew circuits of polynomial size is a p-projection of imm, we get the following corollary.

Corollary 6.15. *If a polynomial is computed by a weakly skew circuit of size s, then it is computed by a skew circuit of size* $\mathrm{poly}(s)$.

The determinant

VP_{ws} describes the complexity of the determinant.
Equivalent models are: algebraic branching programs and projections of iterated matrix multiplication.

7 The permanent

In this chapter we define the class VNP and prove the VNP-completeness of the permanent polynomial.

7.1 VNP and formulas

A language L is in NP if there is a deterministic polynomial time relation R such that for all x, $x \in L$ iff there is a polynomially long bit string y such that $R(x, y) = 1$. Think of x being a formula in 3-CNF and y being an assignment. $R(x, y) = 1$ means that y satisfies x. While NP can be defined as the class of functions that assign to each x the Boolean value

$$\bigvee_{y} R(x, y),$$

the class #P is the class of functions that assign to each x the number of y such that $R(x, y) = 1$, that is, we compute

$$\sum_{y} [R(x, y) = 1].$$

Here, the bracket is the Iverson bracket, which is one if the Boolean expression is true and zero otherwise. So in our example, we want to count the number of satisfying assignments.

Definition 7.1.
1. A p-family (f_n) is in VNP, if there are polynomially bounded functions p and q and a sequence $(g_n) \in \mathrm{VP}$ of polynomials $g_n \in \mathbb{F}[X_1, \dots, X_{p(n)}, Y_1, \dots, Y_{q(n)}]$ such that
$$f_n = \sum_{e \in \{0,1\}^{q(n)}} g_n(X_1, \dots, X_{p(n)}, e_1, \dots, e_{q(n)}).$$
2. A family of polynomials f_n is in VNP_e if in the definition of VNP, the family (g_n) is in VP_e.

You can think of the X-variables representing the input and the Y-variables the witness. With this interpretation, VNP is more like #P. In particular, we will see that the permanent polynomial

$$\mathrm{per}_n = \sum_{\sigma \in \mathfrak{S}_n} X_{1,\sigma(1)} \cdots X_{n,\sigma(n)}$$

is complete for VNP.

With the help of so-called parse trees we will now show $\mathrm{VNP} = \mathrm{VNP}_e$.

Definition 7.2. Let C be an arithmetic circuit.

1. A *parse tree* of C is defined recursively as follows: Every circuit consisting of one node is a parse tree. If the size of C is larger than one, let g be the output gate and g_1 and g_2 be its children. Let C_1 and C_2 be the subcircuits rooted at g_1 and g_2, respectively. If g is an addition gate, then we get the set of all parse trees by either taking a parse tree of C_1 or a parse tree of C_2 and connecting it to g. If g is a multiplication gate, then we get the set of all parse trees by taking a parse tree of C_1 and a parse tree of C_2 and connecting both to g.

2. The set of all parse trees of C is denoted by $\mathrm{pt}(C)$.

3. The weight $w(T)$ of a parse tree T is the product of the labels of its leaves.

For every multiplication gate, we have to include both children in the parse tree, for every addition gate we have to choose one of them. Note that a gate may occur several times in a parse tree, since it is reused in the circuit several times. For each occurrence in the parse tree, we introduce a new copy. (Otherwise, it would not be a tree.)

Exercise 7.3. Let C be a circuit and p be the polynomial computed by C. Prove (for instance by structural induction) that

$$p = \sum_{T \in \mathrm{pt}(C)} w(T).$$

Lemma 7.4. *A circuit C is multiplicatively disjoint if every parse tree of C is a subcircuit of C.*

Proof. Assume that C is not multiplicatively disjoint. Then there is a node v in C such that there are two node disjoint paths to some multiplication gate g. Since g is a multiplication gate, these two paths can be extended to a parse tree.

Conversely, if there is a parse tree T that is not a subcircuit of C, then there are gates g and h in C such that there a two node disjoint paths from g to h. Since T is a parse tree, h is a multiplication gate. Thus, C is not multiplicatively disjoint. □

Lemma 7.5. *Let C be a multiplicatively disjoint circuit with edge set E. For each edge $e \in E$, let X_e be an indeterminate. There is a formula F in the X_e's of size polynomial in the size of C such that for every $a \in \{0,1\}^{|E|}$, $F(a)$ is the weight of the parse tree, if the edges "selected" by the vector a form a parse tree in C, and zero otherwise.*

Proof. Since by Lemma 7.4, every parse tree is a subcircuit of C, it is sufficient to consider subtrees of the given circuit. For every node v in C, we introduce an additional variable Y_v. Note that for $\{0,1\}$ valued variables X and Y, we can simulate Boolean AND by XY and Boolean NOT by $1 - X$. We can write the fact that a given vector encodes a parse tree by the following Boolean expressions:

$$\bigwedge_{(i,j)\in E} X_{(i,j)} \Rightarrow Y_i \wedge Y_j$$

ensures that whenever an edge is selected, its end points are selected, too. Let g be the output gate of C. Then

$$Y_g$$

ensures that the output gate is selected. For a gate g, let $\ell(g)$ and $r(g)$ be its children. The following expression ensures that for every multiplication gate g that is selected, both incoming edges are selected, too.

$$\bigwedge_{\text{multiplication gate } g} Y_g \Rightarrow X_{(\ell(g),g)} \wedge X_{(r(g),g)}.$$

If we replace the Boolean AND on the righthand side by a Boolean XOR, we get an expression that checks for every selected addition gate whether exactly one of the incoming edges is chosen. Finally, we have to check that every selected gate has at least one outgoing edge. This is done by the following expression:

$$\bigwedge_{v \in V} \left(Y_v \Rightarrow \bigvee_{(v,u) \in E} X_{(v,u)} \right).$$

We can eliminate all occurences of the newly introduced variables by replacing Y_v by the expression

$$\bigvee_{(v,u) \in E} X_{(v,u)}$$

and Y_g by 1. The Boolean AND of these expressions is a Boolean formula that is true iff the vector a encodes a parse tree. By the considerations above, it can be replaced by an arithmetic formula.

If a encodes a parsetree, we can get the corresponding weight by the following expression:

$$\prod_{v \in V} (Y_v \cdot w_v + 1 - Y_v).$$

Here w_v is the label of v if it is an input gate and 1 otherwise. Again, we can eliminate the Y_v's as above. The product of the two expressions, one for checking whether a is a parse tree and one for computing its weight, is the formula F. □

Corollary 7.6. *Let f be a polynomial computed by an arithmetic circuit of size s. Then there is an arithmetic formula F of size polynomial in s and a polynomially bounded p such that*

$$f(X) = \sum_{a \in \{0,1\}^{p(s)}} F(X, a).$$

Proof. This follows from combining Theorem 6.7, Exercise 7.3, and Lemma 7.5. □

Theorem 7.7. $\mathrm{VNP} = \mathrm{VNP}_e$.

Proof. Let (f_n) be in VNP and $(g_n) \in \mathrm{VP}$ such that

$$f_n(X) = \sum_{e \in \{0,1\}^{q(n)}} g_n(X, e).$$

Using Corollary 7.6, there is a formula F_n of polynomial size such that

$$g_n(X, Y) = \sum_{a \in \{0,1\}^{p(n)}} F_n(X, Y, a).$$

Therefore,

$$f_n(X) = \sum_{e \in \{0,1\}^{p(n)},\, a \in \{0,1\}^{q(n)}} F_n(X, e, a).$$ □

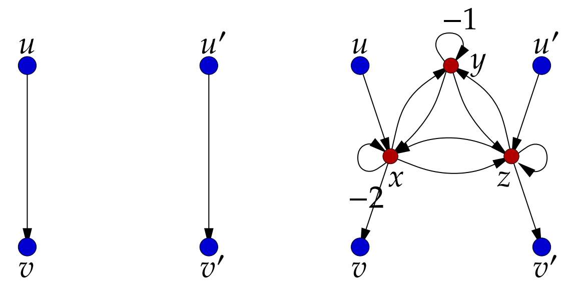

Figure 5: The equality gadget. The pair of edges (u, v) and (u', v') of the left-hand side is connected as shown on the right-hand side.

While the statement of the theorem sounds astonishing at a first glance, it just uses the fact that we can write the result of a polynomially large circuit by an exponential sum over a polynomially large formula and then combines the two exponential sums into one.

On a very high level, Theorem 7.7 is similar in nature to the different but equivalent definitions of NP via a verifier from P or from NC^1.

7.2 Hardness of the permanent

In this section, we prove that the permanent family is complete for VNP. Let $G = (V, E)$ be an edge weighted graph. Recall that a cycle cover C of G is a selection of node disjoint directed cycles such that every node is contained in exactly one cycle. The weight $w(C)$ of C is the product of the weight of the edges in C. Cycle covers can be viewed as the graph of a permutation. The cycles in the cycle cover correspond to the cycles in the cycle decomposition of a permutation. If we also write G for the weighted adjacency matrix of G (by abuse of notation), then

$$\mathrm{per}(G) = \sum_{\text{cycle cover } C \text{ of } G} w(C).$$

Let G be a graph and $e = (u, v)$ and $e' = (u', v')$ be two edges in G. As a first step, we want to replace G by a graph $\hat{G}$ such that $\mathrm{per}(\hat{G})$ is the sum over all $w(C)$ such that C is a cycle cover of G that either contains both e and e' or none of them. This is achieved by subdividing the edges and connecting them by an equality gadget as depicted in Figure 5.

Let C be a cycle cover of G that takes both edges. Then there is one way to extend this to a cycle cover of $\hat{G}$. The only new edges and nodes in $\hat{G}$ are the internal ones of the equality gadget. When both e and e' are taken, then we only need to cover the internal node. The weight of this new cycle cover is $2 \cdot w(C)$, see Figure 6. When C does not take any of the two edges, then there are six ways to extend C. These six ways sum up to weight $2 \cdot w(C)$.

If C is a cycle cover of G that takes only one edge of e and e', say e, then there are two ways to extend C to $\hat{G}$, see Figure 7. The weight of these covers is the same, but they differ in sign, therefore the contributions of these two covers cancel each other.

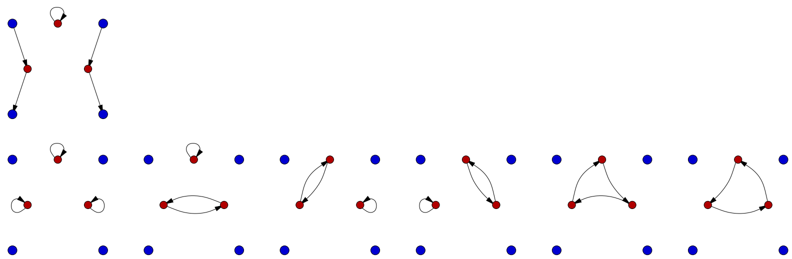

Figure 6: First row: The one possible configuration if both edges are taken. It has weight $-1 \cdot (-2) = 2$. Second row: The six possible configurations if none of the edges is taken. They have weights $-1, -1, 1, 1, 1$, and 1, respectively.

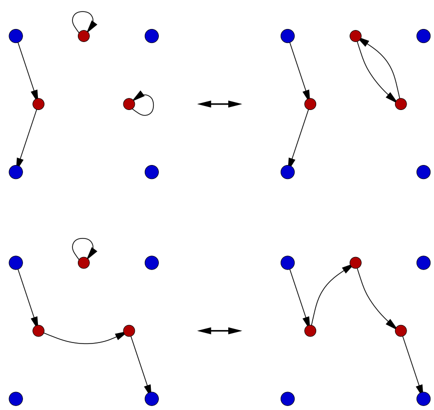

Figure 7: First row: The two covers of the equality gadget when only one edge is taken. Second row: Inconsistent covers of the equality gadget. (In both rows, there is a corresponding symmetric case).

Finally, there are inconsistent ways to cover the equality gadget in $\hat{G}$, that is, covers of $\hat{G}$ that do not correspond to any cover in G, see Figure 7. Again, we can form pairs of these covers such that the contribution of these covers cancel each other.

This construction proves the following lemma.

Lemma 7.8. *Let $\mathbb{F}$ be a field of characteristic distinct from 2. Let G be a graph and e and e' be edges in G. Then there is a graph $\hat{G}$ such that*

$$\frac{1}{2}\mathrm{per}(\hat{G}) = \sum_{C} w(C),$$

where the sum is taken over all cycle covers C of G that either use both of e and e' or none of them.

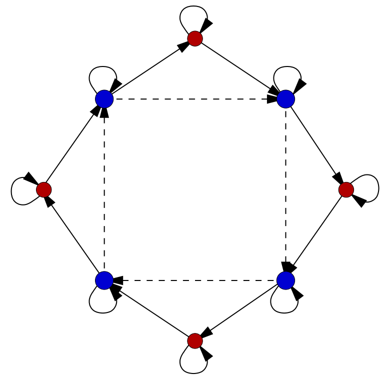

Figure 8: The rosette graph of size four. Connector edges are drawn dashed.

Let $(f_n) \in \mathrm{VNP}$ and let $(g_n) \in \mathrm{VP}$ such that

$$f_n(X_1, \ldots, X_{p(n)}) = \sum_{e \in \{0,1\}^{q(n)}} g_n(X_1, \ldots, X_{p(n)}, e_1, \ldots, e_{q(n)}).$$

By Theorem 7.7 we may assume that $(g_n) \in \mathrm{VP}_e$. We proved that every polynomial that is computed by a formula of size s is a projection of a determinant of polynomial size. The same proof yields that it is also a projection of a polynomially large permanent, since the cycle covers of the arithmetic branchning program occuring in the proof all had the same sign. It follows that we can write f_n as an exponential sums of permanents. The permanent itself is an exponential sum. So we are done if we can "squeeze" the outer exponential sum into the inner one.

The *rosette graph* of size t consists of a directed cycle of size t. The edges $c_1, \ldots, c_t$ of this cycle are called connector edges. The head and the tail of each connector edge are connected by a path of length two. Every node has a self-loop. All edges have weight one in the rosette graph. The following fact is easily verified:

Lemma 7.9. *Let S be a subset of the connector edges.*

1. *If S is nonempty, then there is exactly one cycle cover of the rosette graph containing the edges in S and no other connector edges.*

2. *There are two cycle covers containing no connector edges.*

g_n is a projection of a polynomially large permanent. This means that there is an edge weighted graph G (with the weights being field elements and variables) such that

$$g_n(X_1, \ldots, X_{p(n)}, Y_1, \ldots, Y_{q(n)}) = \sum_{\text{cycle cover } C} w(C).$$

Assume that the variable Y_i occurs ℓ_i times in G. We add a rosette graph of size ℓ_i and connect every edge labeled with Y_i with one of the connector edges of the rosette using an equality gadget. All edges inherit their weights from the corresponding subgraphs, that is, the edges from G get the weights they have in G, the edges in the equality gadgets keep their weights, and the edges in the rosette graph all have weight one. The only exception are the edges carrying a weight Y_i in G, they get the weight 1 instead. We do this for each i. Assume, we introduced t equality gadgets altogether. We will add one isolated self loop with weight $1/2^t$ to compensate for the 2 that is introduced by every equality gadget. (The characteristic of k should be distinct from 2 for this!) Let H be the resulting graph.

Let C be a cycle cover of G. $w(C)$ is a monomial $m(X_1, \dots, X_n, Y_1, \dots, Y_{q(n)})$. Let I be the set of indices such that Y_i appears in $w(C)$. What is the contribution of C in

$$\sum_e g_n(X_1, \dots, X_{p(n)}, e_1, \dots, e_{q(n)})?$$

If Y_i appears in $w(C)$, then we have to set $e_i = 1$, otherwise, the constribution to the exponential sum will be zero. If Y_i does not appear in $w(C)$, then we can set e_i to 0 or 1. Therefore, the contribution of C is

$$2^{q(n)-|I|} m(X_1, \dots, X_{p(n)}, 1, \dots, 1).$$

We call a cycle cover D of H consistent if for every equality gadget, either both edges it connects are chosen or none of them is chosen. A cycle cover C of G can be extended to a consistent cycle cover of H. If an edge with label Y_i appears in C, then we can extend it in one possible way in the corresponding rosette. If no such edge appears in C then there are two ways. In total, there are $2^{q(n)-|I|}$ extensions. By Lemma 7.8, we know that

$$\operatorname{per}(H) = \sum_{\text{consistent D}} w(D).$$

Therefore,

$$\operatorname{per}(H) = \sum_e g_n(X_1, \dots, X_{p(n)}, e_1, \dots, e_{q(n)}).$$

Theorem 7.10. *Over fields of characteristic distinct from* 2, per *is* VNP-*complete.*

Proof. It remains to show that per $\in$ VNP. It is quite easy to write a Boolean expression $E(Y)$ of polynomial size which checks whether a given matrix $Y \in \{0,1\}^{n \times n}$ is a permutation matrix. As done before, we can write this as an equivalent arithmetic formula $\hat{E}(Y)$. Now it is easy to check that

$$\operatorname{per}(X) = \sum_{Y \in \{0,1\}^{n \times n}} \hat{E}(Y) \prod_{i,j} (X_{i,j} Y_{i,j} + 1 - Y_{i,j}).$$

□

Over fields of characteristic 2, the permanent can only be VNP-hard, if VNP = VP, since it coincides with the determinant in this case. But there are other VNP-complete polynomials that are also hard over fields of characteristic two like the Hamiltonian cycle polynomial, see, e. g., [17].

7.3 Valiant's conjecture

Valiant's conjecture is the algebraic counterpart of the P versus NP conjecture.

Conjecture 7.11 (Valiant)**.** VP $\subsetneq$ VNP.

Since the permanent is VNP-complete, we can rephrase this conjecture as

$$\mathrm{per} \notin \mathrm{VP}.$$

Since $\mathrm{VP}_{ws} \subseteq \mathrm{VP}$, we can formulate a weaker version of Valiant's conjecture, namely, $\mathrm{VP}_{ws} \subsetneq \mathrm{VNP}$. Since VP_{ws} has a nice complete family, this version can be reformulated as

$$\mathrm{per} \not\leq_p \det.$$

It is easy to check that VP_{ws} is closed under interpolation and substitutions. Therefore, by Lemma 5.25, the conjecture $\mathrm{VP}_{ws} \subsetneq \mathrm{VP}$ can also be restated as

$$\mathrm{per} \not\leq_{end} \det.$$

Here, geometric complexity theory starts. As in the case of Waring rank, we will replace arbitrary endomorphisms by invertible ones.

Valiant's conjecture

- $\mathrm{VNP} = \mathrm{VNP}_e$
- per is complete for VNP over fields of characteristic $\neq 2$.
- Valiant's conjecture: VP $\subsetneq$ VNP
- weaker variant: $\mathrm{VP}_{ws} \subsetneq \mathrm{VNP}$ (equivalent to $\mathrm{per} \not\leq_{end} \det$)

8 Border complexity and group orbit closures

In this chapter we explain why in Section 3.5 we went from monoid orbits to group orbit closures: We phrase the questions from algebraic complexity theory in terms of group orbit closures.

Recall that on the space of polynomials $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_M]_d$ we have seen several ways to measure complexity:

- minimal size of an arithmetic formula in which the leafs are constants or variables,
- minimal size of an arithmetic formula in which the leafs are affine linear forms,
- $\min\{n \mid \ell^{n-d} h \in \mathsf{End}_{3n^2} \mathrm{imm}_n^{(3)}\}$,

where ℓ denotes the padding variable (which could be one of the existing variables). For a p-family either all or none of these measures are polynomially bounded, that is, all measures can be used to characterize the class VP_e. Recall that VP_e consists of all p-families where these measures are polynomially bounded.

The class VP_{ws} is characterized by the following measures:

- minimal weakly skew circuit size,
- minimal skew circuit size,
- minimal algebraic branching program size,
- $\min\{n \mid \ell^{n-d}h \in \mathsf{End}_{n^3}\mathrm{imm}_n\}$,
- $\min\{n \mid \ell^{n-d}h \in \mathsf{End}_{n^2}\det_n\}$ (i. e., determinantal complexity).

For a p-family either all or none of these measures are polynomially bounded. The class VP_{ws} consists of the p-families where these measures are polynomially bounded.

Both classes have a characterization in terms of an endomorphism orbit. Since group orbits are much easier to handle than monoid orbits, we replace each orbit by its closure. Then we can replace End_m by GL_m (like we did for Waring rank). Furthermore, in algebraic geometry, understanding the closure of constructible sets is often much easier than understanding the set itself. We refer to [41, App. B] for an extended discussion of this and how it relates to complexity.

We obtain new ways of measuring the complexity of polynomials. We define $\overline{\mathrm{VP}_e}$ to be the class of p-families where $\min\{n \mid \ell^{n-d}h \in \overline{\mathsf{GL}_{9n}\mathrm{imm}_n^{(3)}}\}$ is polynomially bounded. We define $\overline{\mathrm{VP}_{ws}}$ to be the class of p-families where $\min\{n \mid \ell^{n-d}h \in \overline{\mathsf{GL}_{n^2}\det_n}\}$ is polynomially bounded. It is easy to see that taking the closure in the other endomorphism description yields the same class.

In general for a set C of sequences of polynomials we define its closure $\overline{C}$ as follows: The sequence $(f_n)_n$ is in $\overline{C}$ iff there exist polynomials $f_{n,i}$ such that

- for all i, the sequences $(f_{n,i})_{n\in\mathbb{N}}$ are in C
- for all n, the sequences $(f_{n,i})_{i\in\mathbb{N}}$ converge to f_n.

Clearly $C \subseteq \overline{C}$, in particular $\mathrm{VP}_e \subseteq \overline{\mathrm{VP}_e}$, $\mathrm{VP}_{ws} \subseteq \overline{\mathrm{VP}_{ws}}$, and $\mathrm{VNP} \subseteq \overline{\mathrm{VNP}}$. But the relationship between C and $\overline{C}$ is unknown in most cases. It is not even known whether $\overline{\mathrm{VP}_e} \subseteq \mathrm{VNP}$. In particular we could have that $\mathrm{VP}_e \neq \mathrm{VNP}$ but their closures are the same. In this case, the main methods of this article would fail. In fact, no method that could only work with the closure could work. Instead, we would have to consider the Zariski-constructible sets themselves and try to separate them with methods of algebraic geometry.

The classical group orbit closure studied in geometric complexity theory is $\overline{\mathsf{GL}_{n^2}\det_n}$. For fixed m and n we search for ways to prove $\ell^{n-m}\mathrm{per}_m \notin \overline{\mathsf{GL}_{n^2}\det_n}$. Since (per_m) is VNP-complete, proving superpolynomial lower bounds is equivalent to separating $\mathrm{VNP} \not\subseteq \overline{\mathrm{VP}_{ws}}$.

It is a challenging problem to understand the closures of algebraic complexity classes. Some recent results can be found in [13, 15, 42, 55].

Border complexity and group orbit closures

The lower bound questions in algebraic complexity theory can be stated in terms of border complexity.
Proving border complexity lower bounds is a special case of the problem of separating a point from a group orbit closure.

9 Representation Theory

Recall the group actions from Chapter 3 and in particular the lifting of the action to the function space. In this chapter we lift group actions again until in Section 10.2 we obtain new significant search space restrictions for obstructions that come from representation theory. Our ground field is the complex numbers.

9.1 Key example: Lifting the group action

Let $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_N]_d$. Recall that $\mathbb{C}[\mathbb{A}]_\delta$ is the vector space of homogeneous degree δ polynomials on $\mathbb{A}$, i.e., $\mathbb{C}[\mathbb{A}]_\delta$ consists of homogeneous degree δ polynomials in the coefficients of a degree-d polynomial. Let $G := \mathsf{GL}_N$. Since G acts on $\mathbb{A}$, we know that G also acts linearly on $\mathbb{C}[\mathbb{A}]$ as follows: for every $g \in G$ and every $f \in \mathbb{C}[\mathbb{A}]$ we define the polynomial $gf \in \mathbb{C}[\mathbb{A}]$ via:

$$\text{for every } h \in \mathbb{A} \text{ we have } (gf)(h) := f(g^T h). \tag{9.1}$$

The following two lemmas are special cases of Lemmas 3.5 and 3.6, since $\mathsf{GL}_N \subseteq \mathsf{End}_N$. We state them explicitly, since we refer to them frequently.

Lemma 9.1. *Let $f, f' \in \mathbb{C}[\mathbb{A}]$ and let $g \in G$. For all complex numbers α, α' we have*

$$g(\alpha f + \alpha' f') = \alpha(gf) + \alpha'(gf').$$

By induction Lemma 9.1 holds for arbitrary finite linear combinations.

Lemma 9.2. *Let $f, f' \in \mathbb{C}[\mathbb{A}]$ and let $g \in G$. Then*

$$g(f \cdot f') = (gf) \cdot (gf').$$

Again, as in Chapter 3, this means that we only need to understand the action on single variables.

Example 9.3. Let $N = 2$, $d = 2$, $\delta = 2$, $\mathbb{A} = \mathbb{C}[X, Y]_2$, and $\mathbb{C}[\mathbb{A}]_2 = \mathbb{C}[T_1, T_2, T_3]_2$ be the homogeneous degree 2 functions on $\mathbb{A}$. For the sake of readability, we here reuse the letter d differently (with a different meaning): Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $g^T X = aX + bY$ and $g^T Y = cX + dY$

as in Chapter 3. Thus,

$$\begin{aligned} g^T X^2 &= (aX + bY)^2 = a^2X^2 + 2abXY + b^2Y^2, \\ g^T XY &= acX^2 + (bc + ad)XY + bdY^2, \\ g^T Y^2 &= c^2X^2 + 2cdXY + d^2Y^2. \end{aligned}$$

For every polynomial $f \in \mathbb{C}[\mathbb{A}]_1$, $f = \alpha T_1 + \beta T_2 + \gamma T_3$, we have $f(X^2) = \alpha$, $f(XY) = \beta$, and $f(Y^2) = \gamma$. Thus $f = f(X^2) \cdot T_1 + f(XY) \cdot T_2 + f(Y^2) \cdot T_3$. Using $(gT_i)(h) = T_i(g^T h)$ we obtain:

$$\begin{aligned} gT_1 &= (gT_1)(X^2) \cdot T_1 + (gT_1)(XY) \cdot T_2 + (gT_1)(Y^2) \cdot T_3 = a^2T_1 + acT_2 + c^2T_3, \\ gT_2 &= 2abT_1 + (bc + ad)T_2 + 2cdT_3, \\ gT_3 &= b^2T_1 + bdT_2 + d^2T_3, \\ gT_2^2 &= 4a^2b^2T_1^2 + 4ab(bc + ad)T_1T_2 + 8abcdT_1T_3 + (bc + ad)^2T_2^2 + 4(bc + ad)cdT_2T_3 + 4c^2d^2T_3^2, \\ gT_1T_3 &= a^2b^2T_1^2 + (a^2bd + ab^2c)T_1T_2 + (a^2d^2 + b^2c^2)T_1T_3 + abcdT_2^2 + (acd^2 + bc^2d)T_2T_3 + c^2d^2T_3. \end{aligned}$$

For the discriminant, we obtain

$$\begin{aligned} g(T_2^2 - 4T_1T_3) &= gT_2^2 - 4gT_1T_3 \\ &= 8abcdT_1T_3 + (bc + ad)^2T_2^2 - 4((a^2d^2 + b^2c^2)T_1T_3 + abcdT_2^2) \\ &= 8abcdT_1T_3 + (b^2c^2 + 2abcd + a^2d^2)T_2^2 - (4a^2d^2 + 4b^2c^2)T_1T_3 - 4abcdT_2^2 \\ &= (8abcd - 4a^2d^2 - 4b^2c^2)T_1T_3 + (b^2c^2 + 2abcd + a^2d^2 - 4abcd)T_2^2 \\ &= 4(2abcd - a^2d^2 - b^2c^2)T_1T_3 + (b^2c^2 - 2abcd + a^2d^2)T_2^2 \\ &= (ad - bc)^2T_2^2 - 4(ad - bc)^2T_1T_3 \\ &= (ad - bc)^2(T_2^2 - 4T_1T_3) \\ &= \det(g)^2(T_2^2 - 4T_1T_3). \end{aligned}$$

Thus the discriminant is fixed under the group action (up to the prefactor). Indeed, the discriminant is fixed under the group action of the special linear group (i. e., matrices with determinant 1).

Example 9.4. Let $g = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$. Then $gT_1^2 = a^4T_1^2$. Thus the line $\mathbb{C}T_1^2$ is fixed under the action of upper triangular matrices. However, it is not fixed under the action of lower triangular matrices: Let $g = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$. Then $gT_1^2 = T_1^2 + 2cT_1T_2 + 2c^2T_1T_3 + c^2T_2^2 + 2c^3T_2T_3 + c^4T_3^2$. The linear span of these gT_1^2 is of dimension at least 5: Their coefficient vectors are $(1, 2c, 2c^2, c^2, 2c^3, c^4)$, so putting $c = -2, -1, 0, 1, 2$ yields the rank 5 matrix

$$\begin{pmatrix} 1 & -4 & 8 & 4 & -16 & 16 \\ 1 & -2 & 2 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 2 & 1 & 2 & 1 \\ 1 & 4 & 8 & 4 & 16 & 16 \end{pmatrix}.$$

Remark 9.5. It is a small calculation to verify that for $\mathbb{A} = \mathbb{C}[X, Y]_2$ we have $\mathbb{C}[\mathbb{A}]_2 = \mathbb{C}(T_2^2 - 4T_1T_3) \oplus \langle \mathsf{GL}_2 T_1^2 \rangle$.

9.2 General representation theory

Remark 9.5 gives an interesting example of a decomposition of $\mathbb{C}[\mathbb{A}]_\delta$. It is called the decomposition into *irreducible representations*. To understand this decomposition and how it can be used to restrict the search space for obstructions we now study some basic representation theory. This section and Chapter 10 are based on [16].

Let G be a group and V be a finite dimensional complex vector space. Recall the definition of a linear monoid action from Chapter 3. Now we restrict our attention to monoids that are groups: A group homomorphism $\varrho : G \to \mathsf{GL}(V)$ is called a linear group action or a *representation* of G. If the action is understood, then we just say that V is a representation.

Example 9.6. Let $G = \mathfrak{S}_n$. Let $V = \mathbb{C}^n$ and $\varrho(\pi)(e_i) = e_{\pi(i)}$, that is, the symmetric group on n letters acts on $\mathbb{C}^n$ by permuting the n coordinates. Then $(\pi\sigma)(e_i) = e_{\pi(\sigma(i))} = \pi(\sigma e_i)$ for all $\pi, \sigma \in \mathfrak{S}_n$. This is called the *defining representation of* $\mathfrak{S}_n$. The matrices $\varrho(\pi)$ are called *permutation matrices*. They consist of a single 1 in each row and each column, and the rest is filled with zeros.

Example 9.7. Let $C_n := \mathbb{Z}/n\mathbb{Z}$ denote the cyclic group of order n. We can think of $C_n \subseteq \mathbb{C}$ as the group of n-th roots of unity. The identity element in C_n is denoted by 1_{C_n}. Let C_n be generated by the element g, i. e., $C_n = \langle g \rangle$. Let $\mathbb{C}^* := \mathbb{C} \setminus \{0\} = \mathsf{GL}_1$.

Let $\varrho : C_n \to \mathbb{C}^*$ be a 1-dimensional representation of C_n. Since $g^n = 1_{C_n}$ we have $1 = \varrho(1_{C_n}) = \varrho(g^n) = \varrho(g)^n$, thus $\varrho(g) = \zeta^k$ for some $k \in \mathbb{Z}$ and $\zeta := \mathrm{e}^{\frac{2\pi i}{n}}$. Indeed, each k gives a representation.

Example 9.8. As we saw in Section 9.1, for $\mathbb{A} = \mathbb{C}[X_1, \dots, X_N]_d$ the vector space $\mathbb{C}[\mathbb{A}]_\delta$ is a GL_N representation.

Remark 9.9. Representations are sometimes called *modules*. Representations of G are precisely the finite dimensional modules where the underlying ring is the group algebra of G. The group algebra $\mathbb{C}[G]$ is defined as the vector space of formal linear combinations of finitely many group elements. Two elements of $\mathbb{C}[G]$ are multiplied via the obvious convolution: $(\sum_i \alpha_i g_i) \cdot (\sum_j \beta_j g_j) = \sum_{i,j} \alpha_i \beta_j (g_i g_j)$, where $\alpha_i, \beta_j \in \mathbb{C}$ and $g_i, g_j \in G$.

Definition 9.10. A linear subspace $W \subseteq V$ of a representation V is called a *subrepresentation* if it is closed under the action of G, i. e.,

$$\forall g \in G \, \forall w \in W : gw \in W.$$

The zero vector space and V itself are always subrepresentations. These are called the trivial subrepresentations.

Example 9.11. Let $\mathbb{C}^n$ denote the defining representation of $\mathfrak{S}_n$. Then $w := e_1 + e_2 + \cdots + e_n$ is fixed under the action of $\mathfrak{S}_n$ and hence the line $\mathbb{C}w$ is a subrepresentation.

It is easy to verify that in general, for every element $v \in V$, the linear span $\langle Gv \rangle$ of the orbit Gv is a subrepresentation.

Example 9.12. According to Example 9.3, for the discriminant, $\langle \mathsf{GL}_2(b^2 - 4ac) \rangle$ is a 1-dimensional subrepresentation of $\mathbb{C}[\mathbb{A}]_2$, where $\mathbb{A} = \mathbb{C}[X, Y]_2$.

Example 9.13. Let $Z \subseteq \mathbb{A}$ be a set that is closed under the action of GL_N. Then the vanishing ideal $I(Z)_\delta$ is a subrepresentation of $\mathbb{C}[\mathbb{A}]_\delta$. This can be seen as follows: Let $g \in \mathsf{GL}_N$ and $z \in Z$. If f vanishes on Z, then $(gf)(z) = f(g^T z) = 0$ for all $g \in \mathsf{GL}_N$, because $g^T z \in Z$.

Definition 9.14. If a representation V only has the two trivial subrepresentations, then V is called *irreducible*.

From the definition it is clear that every 1-dimensional representation is irreducible.

Lemma 9.15. *Let $C_n := \mathbb{Z}/n\mathbb{Z}$. Every irreducible C_n-representation is 1-dimensional.*

Proof. Let $C_n = \langle g \rangle$. Let V be a representation of C_n. Consider $\varrho(g) \in \mathsf{GL}(V)$ and let $v \in V \setminus \{0\}$ be an eigenvector of $\varrho(g)$ to some eigenvalue $\beta \in \mathbb{C}$. Then the line $\mathbb{C}v$ is a subrepresentation of V, because:

$$\forall \alpha \in \mathbb{C}: \; g(\alpha v) = \varrho(g)(\alpha v) = \beta \alpha v \in \mathbb{C}v$$

and thus $g^k(\alpha v) = g(g \cdots (g(g\alpha v)) \cdots) = \beta^k \alpha v \in \mathbb{C}v$. Thus if V is irreducible, V is 1-dimensional. □

We will see in Corollary 10.10 that for separating points h from orbit closures $\overline{Gc}$ it suffices to consider polynomials that lie in irreducible representations. This significantly strengthens Proposition 4.17.

Representations

A group homomorphism $G \to \mathsf{GL}(V)$ is a representation. If the action is understood, we simply call the representation V.
The vanishing ideal $I(Z)_\delta$ of a set $Z \subset \mathbb{A}$ that is closed under the action of G is a subrepresentation of $\mathbb{C}[\mathbb{A}]_\delta$.

10 Representation theory of finite groups and Maschke's theorem

Recall from Proposition 4.17 that in order to prove that a certain homogeneous polynomial h of degree δ is not contained in some orbit closure $\overline{\mathsf{GL}_N c}$, we want to find a homogenous polynomial $f \in \mathbb{C}[\mathbb{A}]$ with $f(\mathsf{GL}_N c) = \{0\}$ and $f(h) \neq 0$. Since GL_N is *linearly reductive* (see below), we can find such an f in an irreducible subrepresentation of $I(\overline{Gc})_\delta$. While the linear reductivity of GL_N is beyond the scope of these lecture notes, we will prove it for finite groups.

10.1 Maschke's theorem

Definition 10.1. Let U and W be linear subspaces of V. We say that V is the *direct sum* of U and W if for every $v \in V$ there is a unique $u \in U$ and $w \in W$ such that $v = u + w$. We write $V = U \oplus W$.

If V is a G-representation and U and W are subrepresentations such that $V = U \oplus W$, then we say that U and W are *representation complements*.

Example 10.2. Let $\mathbb{C}^n$ denote the defining representation of $\mathfrak{S}_n$. Then $\mathbb{C}^n = \langle e_1 + \cdots + e_n \rangle \oplus W$, where $W = \{w \in \mathbb{C}^n \mid w_1 + \cdots + w_n = 0\}$. We have $\dim U = 1$, $\dim W = n - 1$, $U \cap W = 0$, thus $U \oplus W = \mathbb{C}^n$.

Does every subrepresentation have a complement? In this chapter we will see that the answer is yes, provided that G is finite.

Definition 10.3. An *inner product* on a finite dimensional complex vector space V is a map

$$\langle .,. \rangle : V \times V \to \mathbb{C}$$

with

- $\langle \alpha_1 v_1 + \alpha_2 v_2, w \rangle = \alpha_1 \langle v_1, w \rangle + \alpha_2 \langle v_2, w \rangle$ for all $\alpha_i \in \mathbb{C}$, $v_i, w \in V$,
- $\langle v, w \rangle = \overline{\langle w, v \rangle}$, where the bar denotes complex conjugation,
- $\langle v, v \rangle > 0$ if $v \neq 0$.

For a linear subspace $U \subseteq V$ the *orthogonal complement* $U^\perp$ is defined as

$$U^\perp := \{v \in V \mid \forall u \in U : \langle v, u \rangle = 0\}.$$

Lemma 10.4. *If $U \subseteq V$ is a linear subspace, then $U^\perp \subseteq V$ is a linear subspace and $V = U \oplus U^\perp$.*

Proof. If $v_1, v_2 \in U^\perp$ and $\alpha_1, \alpha_2 \in \mathbb{C}$, then let $u \in U$ and calculate

$$\langle \alpha_1 v_1 + \alpha_2 v_2, u \rangle = \alpha_1 \langle v_1, u \rangle + \alpha_2 \langle v_2, u \rangle = 0 + 0 = 0,$$

thus $U^\perp \subseteq V$ is a linear subspace. If $u \in U$ and $u \in U^\perp$, then $\langle u, u \rangle = 0$ and hence $u = 0$, thus $U \cap U^\perp = 0$. Since $\dim U^\perp = n - \dim U$ ($U^\perp$ is the vanishing set of $\dim U$ linearly independent linear constraints: $\langle v, u_i \rangle = 0$ for all basis vectors $u_i \in U$), we have $V = U \oplus U^\perp$. □

For $\mathbb{C}^n$, we can define the inner product $\langle v, w \rangle := \sum_{k=1}^n v_k \overline{w_k}$, thus every finite dimensional complex vector space has an inner product.

Definition 10.5. $\langle .,. \rangle$ is called G-invariant, if for all $g \in G$, $v, w \in V$:

$$\langle gv, gw \rangle = \langle v, w \rangle.$$

Lemma 10.6. *Let $\langle .,. \rangle$ be G-invariant and let $U \subseteq V$ be a subrepresentation. Then $U^\perp \subseteq V$ is also a subrepresentation.*

Proof. Let $v \in U^\perp$, $g \in G$. We have to show that $gv \in U^\perp$. Let $u \in U$ be arbitrary. Then

$$\langle gv, u \rangle = \langle gv, gg^{-1}u \rangle \overset{(*)}{=} \langle v, \underbrace{g^{-1}u}_{\in U} \rangle = 0.$$

where $(*)$ holds because $\langle .,. \rangle$ is G-invariant. □

Lemma 10.7. *Let G be finite and let V be a G-representation. Then V has a G-invariant inner product.*

Proof. Let $\langle .,. \rangle'$ be an inner product on V. For $v, w \in V$, we define

$$\langle v, w \rangle := \frac{1}{|G|} \sum_{g \in G} \langle gv, gw \rangle'$$

It is straighforward to verify that $\langle .,. \rangle$ is an inner product on V. We show that $\langle .,. \rangle$ is G-invariant: Let $g' \in G$.

$$\langle g'v, g'w \rangle = \frac{1}{|G|} \sum_{g \in G} \langle gg'v, gg'w \rangle' = \frac{1}{|G|} \sum_{x \in G} \langle xv, xw \rangle' = \langle v, w \rangle.$$

□

Of course, dividing by $|G|$ in the proof of Lemma 10.7 is optional, but it makes the construction idempotent: if $\langle .,. \rangle'$ is already G-invariant, then $\langle .,. \rangle' = \langle .,. \rangle$.

Theorem 10.8 (Maschke's theorem)**.** *Let G be finite and V be a G-representation. Then V decomposes into a direct sum $V = U_1 \oplus U_2 \oplus \cdots \oplus U_t$ of irreducible G-representations U_i.*

Proof. We proceed by induction on $\dim V =: d$. If $d = 0$, then $t = 0$. For the induction step we make a case distinction. If V is irreducible, then we are done. If V is not irreducible, then let $U \in V$ be a nontrivial subrepresentation, i. e., $U \neq 0$ and $U \neq V$. Let $\langle .,. \rangle$ be a G-invariant inner product on V, which exists by the previous lemma, and let $U^\perp$ be the orthogonal complement of U with respect to this inner product.

Using Lemma 10.4 and Lemma 10.6 we see that $V = U \oplus U^\perp$ with the G-representations U and $U^\perp$. Using the induction hypothesis on U and $U^\perp$ we see that both decompose into a direct sum of irreducibles. Summing up this sum finishes the proof. □

10.2 Search space restrictions

Groups for which every representation decomposes into a direct sum of irreducibles are called *linearly reductive* or just *reductive* (which is the same over fields of characteristic 0). We just showed that finite groups are reductive. For us it will be important to know that GL_N is reductive.

Theorem 10.9. *Every finite dimensional polynomial/rational GL_N-representation V decomposes into a direct sum $V = U_1 \oplus U_2 \oplus \cdots \oplus U_t$ of irreducible polynomial/rational GL_N-representations U_i.*

The proof of Theorem 10.9 uses the same idea as Maschke's theorem, but the invariant scalar product is created by using the compact Zariski-dense subgroup $\mathsf{SU}_N \subseteq \mathsf{GL}_N$ and normalizing using the so-called Haar measure. We omit the details here, because they require some measure theory.

The next corollary strengthens Proposition 4.17 by putting another significant restriction on the search space for our obstructions that we search to find complexity lower bounds.

Corollary 10.10. *Let $Z \subsetneq \mathbb{A}$ be a Zariski-closed cone that is closed under the action of GL_N. We have seen that $I(Z)_\delta$ is a GL_N-representation and hence by Theorem 10.9 $I(Z)_\delta$ decomposes into a sum of irreducibles. Let $h \notin Z$. Then there exists an irreducible GL_N-representation $U \subseteq I(Z)_\delta$ and an $f \in U$ such that $f(h) \neq 0$.*

Proof. Let $0 \neq I(Z)_\delta = U_1 \oplus \cdots \oplus U_k$. Pick $0 \neq f \in I(Z)_\delta$ with $f(h) \neq 0$. Write $f = f_1 + \cdots + f_k$ with $f_i \in U_i$, so in particular $f_i \in I(Z)_\delta$. Since $f(h) \neq 0$ there exists i with $f_i(h) \neq 0$. □

Search space restrictions

GL_N is a reductive group, that is, every GL_N-representation decomposes into a direct sum of irreducible GL_N-representations.
If $h \notin \overline{\mathsf{GL}_N c}$, then for some degree δ a separating polynomial f can be found in an irreducible representation of $I(\mathsf{GL}_N c)_\delta$.

11 The irreducible representations of the general linear group: First properties

Since by Corollary 10.10 we can find obstructions in irreducible GL_n-representations, we want to understand the structure of irreducible GL_n-representations better. In this chapter we will prove the existence of so-called highest weight vectors in irreducible representations of GL_n. These will be sufficient to separate points from orbit closures.

We follow [51, III.1.3–III.1.4].

11.1 Equivariant maps and isomorphisms

Given two representations (V, ϱ) and (V', ϱ') of a group G, a linear map $\varphi : V \to V'$ is called *equivariant* or a *G-morphism* if

$$\forall g \in G, v \in V : \ g\varphi(v) = \varphi(gv),$$

or in other words, $\varrho'(g)\varphi(v) = \varphi(\varrho(g)v)$. If φ is an equivariant vector space isomorphism, then we say that V and V' are *isomorphic representations*.

Definition 11.1. Let $G \leq \mathsf{GL}_n$ be a subgroup (we will take G to be GL_n or the group of diagonal matrices). A representation $\varrho : G \to \mathsf{GL}(V)$ is called *polynomial* if the $\dim(V)^2$ coordinate functions are multivariate polynomials in the n^2 coordinate functions of GL_n.

Example 11.2. $\mathbb{C}[\mathbb{A}]_\delta$ is a polynomial representation, see Proposition 3.21. Subrepresentations of polynomial representations are polynomial.

Example 11.3. Let $\mathbb{C}^2 = \langle v_1, v_2 \rangle$. Let $G = \mathsf{GL}_1 \times \mathsf{GL}_1$ act on $\mathbb{C}^2$ via $(g_1, g_2)(v_1, v_2) = (g_1^2.v_1, g_1 g_2^2.v_2)$, where we used lower dots for the product of a scalar and a vector. We see that $\mathbb{C}^2$ is a polynomial representation of G that decomposes into a direct sum $\mathbb{C}^2 = \mathbb{C}v_1 \oplus \mathbb{C}v_2$ of two polynomial subrepresentations.

Our goal is to classify the classes of isomorphic polynomial irreducible representations of GL_n. In order to do that, we will generalize Example 11.3.

11.2 The algebraic torus and the weight decomposition

Before we study the irreducible representations of GL_n, in this section we study the subgroup of invertible diagonal matrices in GL_n.

Definition 11.4. $T_n := (\mathbb{C}^\times)^n \subseteq \mathsf{GL}_n$ denotes the group of invertible diagonal matrices, also called the *algebraic torus*.

We prove that T_n is linearly reductive and we fully describe its irreducible representations.

Recall that a matrix g is called *diagonalizable* if there exists an invertible matrix P such that $P^{-1}gP$ is a diagonal matrix. Matrices with pairwise distinct eigenvalues are diagonalizable, in particular the set of diagonalizable matrices lies dense in the set of all matrices, i. e., every matrix can be approximated arbitrarily closely by diagonalizable matrices via slight perturbations of the entries.

We will use the following lemma for subgroups H which consist of representation matrices $\varrho(t) \in \mathsf{GL}(V)$, where $t \in T_n$.

Lemma 11.5 (Simultaneous diagonalizability)**.** *Let $H \leq \mathsf{GL}(V)$ be an abelian subgroup and each $g \in H$ diagonalizable. Then H is simultaneously diagonalizable, i. e., there exists $P \in \mathsf{GL}(V)$ such that for all $g \in H$ we have that $P^{-1}gP$ is a diagonal matrix.*

Proof. The proof is by induction on the size of the matrices. The base case is when all matrices of H have only one eigenvalue. This in particular includes the case for which the matrices have size 1×1.

If $g = S^{-1}DS$ is diagonalizable (with a diagonal matrix D) and has only one eigenvalue λ, then $g = \mathrm{diag}(\lambda, \ldots, \lambda)$, because $S^{-1}\mathrm{diag}(\lambda, \ldots, \lambda)S = \lambda S^{-1}S = \mathrm{diag}(\lambda, \ldots, \lambda)$. If all $g \in H$ have only one eigenvalue, then there is nothing to show, because all g are diagonal.

Let $g \in H$ with at least 2 eigenvalues. Then find $S^{-1}gS = \mathrm{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots)$. Note that since $bg = gb$ for all $b \in H$ we have $S^{-1}bSS^{-1}gS = S^{-1}gSS^{-1}bS$. Therefore

$$\mathrm{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots)S^{-1}bS = S^{-1}bS\mathrm{diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots)$$

and hence all matrices in $S^{-1}HS$ are block diagonal, where the block sizes depend only on the multiplicities of the eigenvalues of g.

Then by the induction hypothesis the single blocks can be simultaneously diagonalised by matrices $S_1, S_2 \ldots$. Then $\mathrm{diag}(S_1, S_2, \ldots)$ simultaneously diagonalizes $S^{-1}HS$ and thus $P := S \cdot \mathrm{diag}(S_1, S_2, \ldots)$ simultaneously diagonalizes H. □

Let (V, ϱ) be a polynomial representation of T_n. Since the elements of T_n commute, all elements $\varrho(t)$, $t \in T_n$, commute. Let $H := \{\varrho(t) \mid t \in T_n\}$. To apply Lemma 11.5 we need that each $\varrho(t)$ is diagonalizable. This is achieved in the following lemma.

Lemma 11.6. *Let (V, ϱ) be a polynomial representation of T_n. Then $\varrho(t)$ is diagonalizable for every $t \in T_n$.*

Proof. We start with a fact on multivariate interpolation. A multivariate polynomial f of degree d in n variables is uniquely defined by its $(d+1)^n$ evaluations $f(x_1, \ldots, x_n)$ at points $(x_1, \ldots, x_n) \in \mathbb{C}^n$, where we put $d+1$ different values for each of the x_i, as can be seen by multivariate interpolation.

Let $\tilde{T}_n \leq T_n$ denote the subgroup of elements $\mathrm{diag}(t_1, \ldots, t_n)$ for which each t_i has finite order (i. e., t_i is a root of unity). By definition $\tilde{T}_n = (\tilde{T}_1)^n$. For every k, the primitive k-th roots of unity are in $\tilde{T}_1$, in particular $\tilde{T}_1$ has infinitely many elements. By multivariate interpolation we conclude that if f vanishes on $(\tilde{T}_1)^n$, then $f = 0$. We say that $(\tilde{T}_1)^n = \tilde{T}_n$ lies *Zariski-dense* in $\mathbb{C}^n$.[8]

All elements in $\tilde{T}_n$ commute. Thus all elements in $\varrho(\tilde{T}_n)$ commute. Given $s \in \tilde{T}_n$, let $\langle s \rangle$ be the cyclic group generated by s. Since $\langle s \rangle$ is a finite cyclic group, it is linearly reductive and its irreducible representations are 1-dimensional (Theorem 10.8 and Lemma 9.15). Thus we can decompose V into $\langle s \rangle$-irreducibles, each spanned by a single vector v_i. Now $P^{-1}\varrho(s)P$ is diagonal, where the columns of P are given by the v_i: For standard basis vectors e_i we have $(P^{-1}\varrho(s)P)e_i = P^{-1}\varrho(s)v_i = P^{-1}\alpha v_i = \alpha e_i$ for some $\alpha \in \mathbb{C}$. Therefore $\varrho(s)$ is diagonalizable and using Lemma 11.5 we see that $\varrho(\tilde{T}_n)$ is simultaneously diagonalizable: There exists P such that $P^{-1}\varrho(\tilde{T}_n)P$ are all diagonal.

Define $f_{i,j} : \mathbb{C}^n \to \mathbb{C}$, $(t_1, \ldots, t_n) \mapsto (P^{-1}\varrho(\mathrm{diag}(t_1, \ldots, t_n))P)_{i,j}$. We just saw that $f_{i,j}(s) = 0$ for all $s \in \tilde{T}_n$, $i \neq j$. Since $\tilde{T}_n$ lies Zariski-dense in $\mathbb{C}^n$ it follows that $f_{i,j}(t) = 0$ for all $t \in \mathbb{C}^n$, $i \neq j$. Thus $P^{-1}\varrho(t)P$ is diagonal for all $t \in T_n$. □

Lemma 11.7. *Given a nonzero multivariate polynomial κ in n variables $t = (t_1, \ldots, t_n)$ with $\kappa(t^2) = (\kappa(t))^2$, where $t^2 := (t_1^2, \ldots, t_n^2)$. Then κ is a monomial.*

Proof. For natural numbers $k_1, \ldots, k_n$ we have that $\kappa(\alpha^{k_1}, \alpha^{k_2}, \ldots, \alpha^{k_n})$ is a univariate nonzero polynomial $\zeta(\alpha)$. Moreover, $\zeta(\alpha)^2 = \zeta(\alpha^2)$. The idea is that if $k_1 \gg k_2 \gg \cdots \gg k_n$, then there is a 1:1 correspondence between the nonzero homogeneous parts ζ_i of ζ—which are just single monomials, since ζ is univariate—and the monomials in κ with nonzero coefficient. (This kind of substitution is also called *Kronecker substition* and has been used in polynomial identity

[8] Remark: $\tilde{T}_1$ is an example of a set for which Zariski and Euclidean closure are not the same. The Zariski closure is $\mathbb{C}$, the Euclidean closure is the unit circle. $\tilde{T}_1$ is not constructible.

testing.) Because of this correspondence it suffices to show that ζ is homogeneous, since this imples that κ is a single monomial or zero.

For the sake of contradiction assume that ζ is not homogeneous, so assume that $\zeta = \zeta_i + \zeta_j + \zeta'$ with $\zeta_i \neq 0$ being homogeneous of degree i and $\zeta_j \neq 0$ being homogeneous of degree j, $i > j$, and ζ' being of degree less than j. Then $\zeta(\alpha^2) = \alpha^{2i}\zeta_i(1) + \alpha^{2j}\zeta_j(1) + O(\alpha^{2j-2})$ and $\zeta(\alpha)^2 = \alpha^{2i}\zeta_i(1)^2 + \alpha^{2j}\zeta_j(1)^2 + 2\alpha^{i+j}\zeta_i(1)\zeta_j(1) + O(\alpha^{i+j-1})$. Comparing the coefficient of degree $i+j$ we see that $2\alpha^{i+j}\zeta_i(1)\zeta_j(1) \neq 0$, in contradiction to $\zeta(\alpha^2) = \zeta(\alpha)^2$. □

Theorem 11.8. *For $t = \operatorname{diag}(t_1, \ldots, t_n) \in T_n$ and $\lambda \in \mathbb{N}^n$ we write $t^\lambda := \prod_{i=1}^n t_i^{\lambda_i} \in \mathbb{C}$.*

For every polynomial representation $\varrho : T_n \to \mathsf{GL}(V)$ we have that

$$V = \bigoplus_{\lambda \in \mathbb{N}^n} V_\lambda,$$

where

$$V_\lambda := \{v \in V \mid \varrho(t)v = t^\lambda v \text{ for all } t \in T_n\}.$$

Proof. Using Lemma 11.6, $\{\varrho(t) \mid t \in T_n\}$ is simultaneously diagonalizable, so there is $P \in \mathsf{GL}(V)$ such that $P^{-1}\varrho(t)P$ is diagonal for every $t \in T_n$. Therefore the i-th diagonal entry of $P^{-1}\varrho(t)P$ is given by a function $\kappa(t)$. Since ϱ is a polynomial representation, each κ is a multivariate polynomial in n variables. Since $P^{-1}\varrho(t)P \leq \mathsf{GL}(V)$ is a subgroup of diagonal matrices, $\kappa(tt') = \kappa(t)\kappa(t')$, where the product tt' is defined componentwise. Using Lemma 11.7 it follows that $\kappa(t) = t^\lambda$ for some λ. □

The decomposition in Theorem 11.8 is called the *weight decomposition*. V_λ is called the *weight space of weight λ* and a vector in V_λ is called a *weight vector of weight λ*.

Corollary 11.9. *The polynomial irreducible representations of T_n are 1-dimensional and indexed by lists in $\mathbb{N}^n$.*

Proof. Given a polynomial irreducible representation V of T_n, by Theorem 11.8, V decomposes into a direct sum of weight spaces, each of which decomposes (arbitrarily) into a direct sum of 1-dimensional irreducible T_n-representations. Since V is irreducible, there can be only one summand. It follows that each polynomial irreducible representation is 1-dimensional and a weight space for some weight $\lambda \in \mathbb{N}^n$.

Moreover, let $\lambda \neq \mu$ and V and W be 1-dimensional with weight λ and μ, respectively. For the sake of readability we use lower dots for scalar multiplication: If $\varphi : V \to W$ is T_n-equivariant, then $\varphi(tv) = \varphi(t^\lambda.v) = t^\lambda.\varphi(v) \neq t^\mu.\varphi(v) = t\varphi(v)$ and thus V and W are not isomorphic. The other direction works analogously: If $\lambda = \mu$, then V and W are isomorphic. □

GL_n versus T_n

Every polynomial T_n-representation decomposes into a direct sum of weight spaces, indexed by $\lambda \in \mathbb{N}^n$. This is called the weight decomposition.
Since a polynomial irreducible GL_n-representation is also a polynomial T_n-representation, it also has a weight decomposition.

11.3 Highest weight vectors

As seen in the last section, every polynomial irreducible GL_n-representation has a weight decomposition. This is a first structural result about irreducible GL_n-representations. In this section we fully classify the irreducible GL_n-representations.

Embed $\gamma : \mathfrak{S}_n \hookrightarrow \mathsf{GL}_n$ via permutation matrices, i. e., the permutation π is mapped to the matrix that has entries 1 at positions $(i, \pi(i))$ and zeros everywhere else. One can readily verify that γ is a group homomorphism. $\mathfrak{S}_n$ acts on $\mathbb{N}^n$ in the natural way by permuting the positions, so $\pi(\lambda) \coloneqq (\lambda_{\pi^{-1}(1)}, \ldots, \lambda_{\pi^{-1}(n)})$.

Lemma 11.10. *Given a GL_n-representation V and let V_λ denote its λ weight space. Let $\mathfrak{S}_n$ act on V via $\gamma : \mathfrak{S}_n \hookrightarrow \mathsf{GL}_n$. Then $\pi V_\lambda = V_{\pi(\lambda)}$.*

Proof. Let $v \in V_\lambda$, $t = (t_1, \ldots, t_n) \in T_n$ and $\pi \in \mathfrak{S}_n$.

For the sake of readability let the lower dot denote the multiplication with a scalar from the left. We calculate: $\gamma(\pi^{-1})\mathrm{diag}(t_1, \ldots, t_n)\gamma(\pi) = \mathrm{diag}(t_{\pi(1)}, \ldots, t_{\pi(n)}) =: t_\pi$. Since

$$t_\pi v = \mathrm{diag}(t_{\pi(1)}, \ldots, t_{\pi(n)})v = t_{\pi(1)}^{\lambda_1} \cdots t_{\pi(n)}^{\lambda_n}.v = t_1^{\lambda_{\pi^{-1}(1)}} \cdots t_n^{\lambda_{\pi^{-1}(n)}}.v = t^{\pi(\lambda)}.v,$$

we have

$$t(\pi v) = t\pi v = (\pi\pi^{-1})t\pi v = \pi(\pi^{-1}t\pi)v = \pi(t_\pi v) = \pi(t^{\pi(\lambda)}.v) = t^{\pi(\lambda)}.(\pi v),$$

and therefore $\pi v \in V_{\pi(\lambda)}$. We conclude $\pi V_\lambda \subseteq V_{\pi(\lambda)}$ and by symmetry $\pi V_\lambda = V_{\pi(\lambda)}$. ☐

Definition 11.11. A finite list of natural numbers $\lambda \in \mathbb{N}^n$ is called a *partition* if it is nonincreasing, i. e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. We define $|\lambda| \coloneqq \sum_{i=1}^n \lambda_i$. More generally, we define $|\lambda| \coloneqq \sum_{i=1}^n \lambda_i$ for all $\lambda \in \mathbb{Z}^n$. We say that λ is *a partition of* N if $|\lambda| = N$.

On $\mathbb{Z}^n$ we define the following partial order, the so-called *dominance order*. Two lists $\lambda, \mu \in \mathbb{Z}^n$ satisfy $\lambda \trianglerighteq \mu$ iff

- for all $1 \leq i \leq n$: $\sum_{j=1}^i \lambda_j \geq \sum_{j=1}^i \mu_j$.

In this situation we say that λ *dominates* μ. Usually when comparing λ and μ we have $|\lambda| = |\mu|$. We write $\lambda \triangleright \mu$ to denote that both $\lambda \trianglerighteq \mu$ and $\lambda \neq \mu$ hold.

Example 11.12. We have $(6, 3, 3) \trianglerighteq (6, 2, 2, 2)$, because $6 \geq 6$, $6 + 3 \geq 6 + 2$, $6 + 3 + 3 \geq 6 + 2 + 2$, and $6 + 3 + 3 + 0 \geq 6 + 2 + 2 + 2$.

Dominance is a partial order: $(6, 3, 1, 1) \ntrianglerighteq (5, 3, 3)$ and $(6, 3, 1, 1) \ntrianglelefteq (5, 3, 3)$.

Let $U_n \leq \mathsf{GL}_n$ denote the subgroup of upper triangular matrices with 1s on the diagonal. Analogously, let $U_n^- \leq \mathsf{GL}_n$ denote the subgroup of lower triangular matrices with 1s on the diagonal.

Lemma 11.13. *Let v be a weight vector of weight λ and let $g \in U_n$. Then $gv = v + w$, where $w \in \bigoplus_{\mu \triangleright \lambda} V_\mu$.*

If $g \in U_n^-$ instead, then $gv = v + w$, where $w \in \bigoplus_{\mu \triangleleft \lambda} V_\mu$.

Proof. We only prove the first part. The second part is completely analogous. Again, for the sake of readability we sometimes use the lower dot to denote the multiplication with a scalar from the left.

Define $x_{ij}(\alpha)$, $i \neq j$, to be the identity matrix plus a single $\alpha \in \mathbb{C}$ in row i, column j. We prove the result for $g = x_{ij}(\alpha)$ with $i < j$. This is without loss of generality, because U_n is generated as a group by these $x_{ij}(\alpha)$.

For $t = \operatorname{diag}(t_1, \dots, t_n) \in T_n$ we have $t x_{ij}(\alpha) t^{-1} = x_{ij}(t_i \cdot t_j^{-1} \cdot \alpha)$.

Let $\{w_1, \dots, w_\eta\}$ be a basis of V. Since ϱ is a polynomial representation, each coordinate function of $x_{ij}(\alpha)v$ is a univariate polynomial in α:

$$x_{ij}(\alpha)v = \sum_{s=1}^{\eta} \Big(\sum_{h \geq 0} c_{h,s}\alpha^h\Big) w_s = \sum_{h \geq 0} \alpha^h . v_h$$

with $c_{h,s} \in \mathbb{C}$ and $v_h := \sum_{s=1}^{\eta} c_{h,s} w_s$.

Since $x_{ij}(0) = \mathrm{Id}_n$, it follows $x_{ij}(0)v = v$, thus we get that the constant term $v_0 = v$. We have

$$\begin{aligned} t x_{ij}(\alpha) v &= t x_{ij}(\alpha) t^{-1} t v = (x_{ij}(t_i t_j^{-1} \alpha)) t v = (x_{ij}(t_i t_j^{-1}\alpha))(t^\lambda . v) \\ &= t^\lambda . (x_{ij}(t_i t_j^{-1}\alpha)) v = t^\lambda . \sum_{h \geq 0} t_i^h . t_j^{-h} . \alpha^h . v_h = \sum_{h \geq 0} \alpha^h . (t^{\lambda + h\zeta_{ij}} . v_h), \end{aligned}$$

where $\zeta_{ij} := (0, 0, \dots, 0, 1, 0, \dots, 0, -1, 0, \dots, 0)$ with the 1 at position i and the -1 at position j.

On the other hand

$$t x_{ij}(\alpha) v = t\Big(\sum_{h \geq 0} \alpha^h . v_h\Big) = \sum_{h \geq 0} \alpha^h . t v_h$$

Comparing coefficients we see that

$$t v_h = t^{\lambda + h\zeta_{ij}} . v_h,$$

thus each v_h is an element of $V_{\lambda + h\zeta_{ij}}$.

The proof is finished by observing that $\lambda \lhd (\lambda + h\zeta_{ij})$ for $i < j$, $0 < h$. $\square$

Let $B_n \leq \mathsf{GL}_n$ denote the subgroup of upper triangular matrices. A 1-dimensional linear subspace $\mathbb{C}v$ of a GL_n-representation V is called a *B_n-stable line*, if it is closed under the action of B_n. Since $T_n \leq B_n$ is a subgroup, in a polynomial GL_n-representation every B_n-stable line is also a 1-dimensional T_n-representation and hence every B_n-stable has a weight $\lambda \in \mathbb{N}^n$.

Corollary 11.14. *Let V be a GL_n-representation and $v \in V_\lambda$ for some $\lambda \in \mathbb{N}^n$ such that $\mathbb{C}v$ is a B_n-stable line. Then $\langle \mathsf{GL}_n v \rangle \subseteq \mathbb{C}v \oplus \sum_{\mu \lhd \lambda} V_\mu$.*

Proof. The set $U_n^- T_n U_n \subseteq \mathsf{GL}_n$ is dense, because LU factorization of matrices almost always (on a Zariski open subset) works without pivoting, in other words $\mathsf{GL}_n = \overline{U_n^- T_n U_n}$, where the closure is taken in GL_n. Thus $\mathsf{GL}_n v = \overline{U_n^- T_n U_n} v \subseteq \overline{U_n^- T_n U_n v} \subseteq \overline{\mathbb{C} \cdot U_n^- v}$. Lemma 11.13 yields $\mathbb{C} \cdot U_n^- v \subseteq \mathbb{C}v + \sum_{\mu \lhd \lambda} V_\mu$. The right hand side is closed, since it is a finite-dimensional vector space, thus $\overline{\mathbb{C} \cdot U_n^- v} \subseteq \mathbb{C}v + \sum_{\mu \lhd \lambda} V_\mu$. Therefore $\mathsf{GL}_n v \subseteq \mathbb{C}v + \sum_{\mu \lhd \lambda} V_\mu$. Since the right hand side is a vector space, it follows that $\langle \mathsf{GL}_n v \rangle \subseteq \mathbb{C}v + \sum_{\mu \lhd \lambda} V_\mu$. $\square$

The following theorem completely classifies the irreducible GL_n-representations by the weight of their unique B_n-stable line.

Theorem 11.15. *a') Let V be a polynomial GL_n-representation with a B_n-stable line $\mathbb{C}v \subseteq V$. Then $\langle \mathsf{GL}_n v\rangle$ is irreducible.*

a) For each irreducible polynomial GL_n-representation V there exists exactly one B_n-stable line $\mathbb{C}v \subseteq V$. Let λ be the weight of v, called the highest weight *of V. Then the λ-weight space $V_\lambda = \mathbb{C}v$ is 1-dimensional. Furthermore we have that λ is a partition and for all weights μ that appear in V (i. e., $V_\mu \neq 0$) we have $\mu \trianglelefteq \lambda$.*

b) Two irreducible polynomial representations V and V' are isomorphic, iff their highest weights λ and λ' are equal.

c) Let $\lambda \in \mathbb{N}^n$ be a partition. Then there exists an irreducible polynomial representation of GL_n with highest weight λ.

Remark 11.16. In the situation a) we call v a *highest weight vector (HWV)*.

Proof. (c) For every partition λ we can explicitly construct an irreducible representation. The construction is slightly technical and we postpone it until Section 17.

(a′) Let $\mathbb{C}v$ be a B_n-stable line of weight λ and let $W := \langle \mathsf{GL}_n v\rangle$. Decompose $W = \bigoplus_i W_i$ into irreducible GL_n-representations W_i. Decompose the W_i further into their weight spaces spanned by weight vectors v_j, so that the v_j form a basis of W. Since $v \in W$ has weight λ, one of the v_j must have weight λ. Let $v_j \in W_i =: W'$. By Cor. 11.14 the λ weight space W_λ of W is 1-dimensional and thus the λ weight space W'_λ is also 1-dimensional, in fact $W'_\lambda = W_\lambda$. Thus $v \in W'_\lambda$. Thus $W = \langle \mathsf{GL}_n v\rangle \subseteq W'$. Hence W is irreducible.

(a) Let $\mathbb{C}v \subseteq V$ be a B_n-stable line. The orbit span $\langle \mathsf{GL}_n v\rangle \subseteq V$ is a subrepresentation, but since V is irreducible, actually $V = \langle \mathsf{GL}_n v\rangle$. Using Cor. 11.14 we see that $\langle \mathsf{GL}_n v\rangle \subseteq \mathbb{C}v + \sum_{\mu \triangleleft \lambda} V_\mu$. Therefore:

- The poset (with respect to the dominance order) of weights that occur in V has a maximum: λ
- In V there is a unique line of weight λ.

We now see that the B_n-stable line in V is unique. A second B_n-stable $\mathbb{C}w$ line would have a weight $\mu \triangleleft \lambda$, but then $\langle \mathsf{GL}_n w\rangle \subseteq \mathbb{C}w + \sum_{\nu \triangleleft \mu} V_\nu$ would not contain $\mathbb{C}v$, which is a contradiction to V being irreducible. Thus the B_n-stable line in V is unique.

If λ is not a partition, then $\pi(\lambda)$ is a partition for some $\pi \in \mathfrak{S}_n$. By Lemma 11.10, $\pi V_\lambda = V_{\pi(\lambda)}$. But $\pi(\lambda) \triangleright \lambda$ (easy exercise for the reader), a contradiction to λ dominating all weights in V.

It remains to show that there exists a B_n-stable line. Take all weights μ for which $V_\mu \neq 0$ and take a maximal element λ with respect to the dominance order. Take $0 \neq v \in V_\lambda$. Use Lemma 11.13 to see that $U_n v = v$. Thus V contains at least the U_n-stable line $\mathbb{C}v$. Since $B_n = T_n U_n$ and since $v \in V_\lambda$, it follows that $\mathbb{C}v$ is a B_n-stable line.

(b) Isomorphic representations clearly have equal highest weights: If $\varphi : V \to V'$ is an isomorphism, then v is B_n-stable iff $\varphi(v)$ is B_n-stable. Moreover, $t\varphi(v) = \varphi(tv) = \varphi(t^\lambda v) = t^\lambda \varphi(v)$.

To see the other direction, let $\lambda = \lambda'$, where $\mathbb{C}v \subseteq V$ and $\mathbb{C}v' \subseteq V'$ are the B_n-stable lines in V and V', respectively. Consider the GL_n-representation $V \oplus V'$, $g(u, u') := (gu, gu')$ for all $u \in V$, $u' \in V'$, $g \in \mathsf{GL}_n$. Then $w := (v, v') \in V \oplus V'$. Let $W := \langle \mathsf{GL}_n \cdot w \rangle \subseteq V \oplus V'$. Since $\lambda = \lambda'$ it follows that w has weight λ and $\mathbb{C}w \subseteq W$ is a B_n-stable line. By part (a') we have that W is irreducible.

We write $V = V \oplus \{0\}$ and $V' = \{0\} \oplus V'$. Then $W \cap V' \subseteq W$ is a subrepresentation, but W is irreducible, so $W \cap V' = W$ (i. e., $W \subseteq V'$) or $W \cap V' = \{0\}$. Since $w \in W$, but $w \notin V'$, we have $W \cap V' = \{0\}$. But $W \cap V'$ is the kernel of the linear projection map $\mathrm{pr} : W \to V$, $(v, v') \mapsto v$. Thus pr is injective. Moreover, pr is equivariant. Thus $\tilde{\mathrm{pr}} : W \to \mathrm{pr}(W)$, $w \mapsto \mathrm{pr}(w)$ is a GL_n-isomorphism. Its image is thus a subrepresentation of V, isomorphic to W. But V is irreducible and $W \neq \{0\}$, thus V and W are isomorphic GL_n-representations ($V \cong W$). Analogously we show that $W \cong V'$ and thus $V \cong V'$. ◻

The previous theorem gives a complete characterization of irreducible polynomial GL_n-representations V: Using part (a) we see that V has a unique B_n-stable line $\mathbb{C}v$ of some weight λ. Using part (a') we see that $V = \langle \mathsf{GL}_n v \rangle$. Using parts (b) and (c) we see that there is a 1:1 correspondence between partitions into at most n parts and isomorphism types of irreducible polynomial GL_n-representations.

Example 11.17. Let $\mathbb{A} = \mathbb{C}[X, Y]_2$, $V = \mathbb{C}[\mathbb{A}]_2$. The calculation from Example 9.3 shows that the discriminant $f := T_2^2 - 4T_1T_3$ satisfies $gf = \det(g)^2 f$ and thus $\mathrm{diag}(\alpha_1, \alpha_2)f = \alpha_1^2\alpha_2^2 f$, so f is a weight vector of weight $(2, 2)$. Moreover, $gf = f$ if $\det(g) = 1$, in particular f is fixed under U_n. Thus $\mathbb{C}f$ is a B_n-stable line. Hence f is a highest weight vector of weight $(2, 2)$.

Example 9.4 shows analogously that the polynomial T_1^2 is a highest weight vector of weight $(4, 0)$.

Each orbit span of a HWV is an irreducible subrepresentation. Here the orbit span of the discriminant is 1-dimensional, while the orbit span of T_1^2 is 5-dimensional. Since V is 6-dimensional this concludes the decomposition into irreducibles, as already pointed out in Remark 9.5: the 6-dimensional GL_2-representation V decomposes into a direct sum of two irreducibles: One of type $(4, 0)$ and one of type $(2, 2)$.

11.4 Highest weight vector obstructions

In this section we will see how Corollary 10.10 can be strenghtened even further with an additional significant search space restriction for obstructions: we only need to consider HWVs, see Corollary 11.19.

Proposition 11.18. *Let $\mathbb{A} = \mathbb{C}[X_1, \dots, X_N]_d$. Then $\mathbb{C}[\mathbb{A}]_\delta$ decomposes into irreducibles as*

$$\mathbb{C}[\mathbb{A}]_\delta = \bigoplus_i V_i,$$

where the type of each V_i is a partition of $d\delta$.

Proof. We consider the action of $t := \mathrm{diag}(\alpha, \ldots, \alpha) \in T_n$ on $\mathbb{C}[\mathbb{A}]_\delta$. Indeed, $tf = \alpha^{d\delta} f$ for every $f \in \mathbb{C}[\mathbb{A}]_\delta$. Thus each weight vector $f \in \mathbb{C}[\mathbb{A}]_\delta$ of weight λ (i. e., $tf = t^\lambda f$) must satisfy

$$\alpha^{d\delta} f = tf = t^\lambda f = t_1^{\lambda_1} t_2^{\lambda_2} \cdot t_n^{\lambda_n} f = \alpha^{\lambda_1 + \cdots + \lambda_n} f,$$

thus $\lambda_1 + \cdots + \lambda_n = d\delta$. In particular this is true for highest weight vectors. The statement follows with the classification in Theorem 11.15. □

Corollary 11.19. *Let $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_N]_d$. Let $c \in \mathbb{A}$, $G = \mathsf{GL}_N$. If $h \notin \overline{Gc}$, then there exists $\delta \in \mathbb{N}$, $\lambda \in \mathbb{N}^N$ a partition of δd, and a highest weight vector $f \in \mathbb{C}[\mathbb{A}]_\delta$ of weight λ, such that $f(\overline{Gc}) = \{0\}$ and for almost all group elements $g \in G$ we have $f(gh) \neq 0$. ("Almost all" means that the set of g for which $f(gh) = 0$ is a Zariski-closed proper subset of G.)*

Proof. From Corollary 10.10 we already know that an f exists that is contained in the homogeneous degree δ part $I(\overline{Gc})_\delta$ of the vanishing ideal, but that also satisfies $f(h) \neq 0$. Moreover, $I(\overline{Gc})_\delta$ is a subrepresentation of $\mathbb{C}[\mathbb{A}]_\delta$, so we can decompose it into irreducibles

$$I(\overline{Gc})_\delta = \bigoplus_{j \in \Omega} V_j$$

for a finite index set Ω, where by Prop. 11.18 the type of each V_j is a partition of $d\delta$. Now we can write $f = \sum_{j \in \Omega} f_j$, where $f_j \in V_j$. By Theorem 11.15(a),(a') it follows that we can write the finite sum $f_j = \sum_i g_{j,i} f_{j,i}$, where $g_{j,i} \in G$ and $f_{j,i}$ is an HWV (scalars in the linear combination can be merged with the HWVs, so they do not appear in the sum).

Since $f(h) \neq 0$, we have that $(g_{j,i} f_{j,i})(h) \neq 0$ for some j, i. This means $f_{j,i}(g_{j,i}^T h) \neq 0$, which proves the first part of the corollary, choosing $g = g_{j,i}^T$. For the second part we have to analyze the subset of group elements $\tilde{g} \in G$ that satisfy $f_{j,i}(\tilde{g} h) \neq 0$. But $f_{j,i}(\tilde{g} h)$ is a polynomial in the entries of $\tilde{g}$. This finishes the proof. □

The following calculation gives a feel that looking at HWVs should be useful.

For $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_M]_d$ we have $\dim \mathbb{C}[\mathbb{A}]_\delta = \binom{\delta + \binom{d+M-1}{d} - 1}{\delta}$. Thus if $M = d = \delta = 6$ we have $\dim \mathbb{C}[\mathbb{A}]_\delta = 13\,949\,678\,575\,756$. But one can compute that the dimension of the vector space of highest weight vectors is only 31 781 and the highest dimension of the highest weight subspace in a V_λ is 105.

More crucially, the dimensions of the highest weight vector spaces do not change when increasing M, but $\dim \mathbb{C}[\mathbb{A}]_\delta$ increases significantly! For $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_M]_2$ we have $\dim \mathbb{C}[\mathbb{A}]_2 = \binom{1+\binom{1+M}{2}}{2} = \frac{M^4}{8} + \frac{M^3}{4} + \frac{3M^2}{8} + \frac{M}{4}$, but we will see that the dimension of the space of HWVs is just 2, independent of M, provided $M \geq 2$.

Irreducible representations of GL_n and HWVs

The irreducible polynomial representations of GL_n are indexed by partitions $\lambda \in \mathbb{N}^n$. Each irreducible polynomial representation V has a unique highest weight vector (up to scale): A weight vector that is B_n-stable. Its weight determines the isomorphism type of V.
For proving lower bounds $h \notin \overline{Gc}$ we can restrict our search for obstructions f to highest weight vectors. This greatly reduces the dimension of the search space.

12 Schur's lemma, multiplicities, and isotypic decompositions

So far we restricted the search space for obstructions further and further. In this chapter we want to present a *sufficient* criterion for obstructions that is not known to be a necessary criterion: Comparing representation-theoretic multiplicities, see Section 12.4. This strategy for proving $h \notin \overline{Gc}$ is mathematically beautiful, but still bears many open research questions.

Again we follow Bürgisser's lecture notes.

12.1 Schur's lemma

For a group G and two G-representations V and W we define

$$\mathrm{Hom}_G(V, W) := \{\varphi \mid \varphi : V \to W \text{ a } G\text{-morphism}\}.$$

$\mathrm{Hom}_G(V, W)$ is a vector space and a linear subspace of $\mathrm{Hom}(V, W)$. Moreover, define $\mathsf{End}_G(V) := \mathrm{Hom}_G(V, V)$.

Lemma 12.1. *Let $\varphi \in \mathrm{Hom}_G(V, W)$. Then*

1. $\ker \varphi := \{v \in V \mid \varphi(v) = 0\}$ *is a subrepresentation of V*
2. $\mathrm{im}\varphi := \{\varphi(v) \mid v \in V\}$ *is a subrepresentation of W*

Proof. It is clear that kernel and image are linear subspaces. We have to verify that both are closed under the group action.

If $v \in \ker \varphi$ and $g \in G$, then $\varphi(gv) = g\varphi(v) = g0 = 0$ and thus $gv \in \ker \varphi$.

If $w \in \mathrm{im}\varphi$ and $g \in G$, then choose $v \in V$ such that $\varphi(v) = w$. Then $gw = g\varphi(v) = \varphi(gv) \in \mathrm{im}\varphi$. □

We write $V \cong W$ to denote that V and W are isomorphic representations, and $V \ncong W$ otherwise.

Lemma 12.2 (Schur's lemma)**.** *Let V and W be irreducible G-representations. Then*

1. $V \ncong W \Rightarrow \mathrm{Hom}_G(V, W) = 0$

2. $V \cong W \Rightarrow \dim \mathrm{Hom}_G(V, W) = 1$

Proof. 1.: We show the contraposition and thus assume the existence of a G-morphism $\varphi : V \to W$, $\varphi \neq 0$.

- $\ker \varphi \subsetneq V$ is a subrepresentation. Since V is irreducible, it follows that $\ker \varphi = 0$, thus φ is injective.
- $0 \neq \mathrm{im}\varphi \subseteq W$ is a subrepresentation. Since W is irreducible, $\mathrm{im}\varphi = W$, thus φ is surjective.

Putting both bullet points together we see that φ is bijective. Thus $V \cong W$.

2.: We first treat the case $V = W$. Let $\varphi \in \mathsf{End}_G(V)$ be arbitrary. Let v be an eigenvector of φ to the eigenvalue α. Then $\varphi - \alpha\mathrm{id} \in \mathsf{End}_G(V)$ and $v \in \ker(\varphi - \alpha\mathrm{id})$.

$$0 \neq v \in \underbrace{\ker(\varphi - \alpha\mathrm{id})}_{\text{subrepresentation of } V} \overset{V \text{ irred}}{\Rightarrow} \ker(\varphi - \alpha\mathrm{id}) = V,$$

thus $\varphi - \alpha\mathrm{id} = 0$, therefore $\varphi = \alpha\mathrm{id}$.

For the more general case $V \cong W$ let $\psi \in \mathrm{Hom}_G(V, W)$ be a G-isomorphism. Let $\varphi \in \mathrm{Hom}_G(V, W)$ be arbitrary. Then $\psi^{-1} \circ \varphi \in \mathsf{End}_G(V) = \mathbb{C}\mathrm{id}_V$. Thus there exists $\alpha \in \mathbb{C}$ with $\psi^{-1} \circ \varphi = \alpha\mathrm{id}_V$. Therefore $\varphi = \alpha\psi$. We conclude that $\mathrm{Hom}_G(V, W) = \mathbb{C}\psi$. ◻

12.2 Multiplicities

In this section we present the definition of representation-theoretic multiplicities. We will use this to define special types of obstructions, see Section 12.4.

Corollary 12.3. *Let V be a G-representation. Let $V = U_1 \oplus \cdots \oplus U_t$ be a decomposition into irreducibles. Let W be an irreducible G-representation. Then $|\{i \mid U_i \cong W\}| = \dim \mathrm{Hom}_G(W, V) = \dim \mathrm{Hom}_G(V, W)$.*

Proof. Let $\mathrm{pr}_i : V \to U_i$ denote the i-th canonical projection. The following is an isomorphism of vector spaces:

$$\begin{array}{rcl} \bigoplus_{i=1}^{t} \mathrm{Hom}_G(W, U_i) & \to & \mathrm{Hom}_G(W, V) \\ (\varphi_1, \ldots, \varphi_t) & \mapsto & \big(w \mapsto \varphi_1(w) + \ldots + \varphi_t(w)\big) \\ (\mathrm{pr}_1 \circ \psi, \ldots, \mathrm{pr}_t \circ \psi) & \leftarrow\!\shortmid & \psi \end{array}$$

Schur's lemma implies $\dim \bigoplus_{i=1}^{t} \mathrm{Hom}_G(W, U_i) = \sum_{i=1}^{t} \dim \mathrm{Hom}_G(W, U_i) = |\{i \mid U_i \cong W\}|$, which finishes the proof of the first equality. For the second part we proceed analogously with an isomorphism of vector spaces.

$$\begin{array}{rcl} \bigoplus_{i=1}^{t} \mathrm{Hom}_G(U_i, W) & \to & \mathrm{Hom}_G(V, W) \\ (\varphi_1, \dots, \varphi_t) & \mapsto & \varphi_1 \circ \mathrm{pr}_1 + \cdots + \varphi_t \circ \mathrm{pr}_t \\ (\psi|_{U_1}, \dots, \psi|_{U_t}) & \leftarrow\!\shortmid & \psi \end{array}$$

Schur's lemma implies $\dim \bigoplus_{i=1}^{t} \mathrm{Hom}_G(U_i, W) = \sum_{i=1}^{t} \dim \mathrm{Hom}_G(U_i, W) = |\{i \mid U_i \cong W\}|$, which finishes the proof of the second equality. □

From this corollary we see that $|\{i \mid U_i \cong W\}|$ is *independent* of the decomposition. This justifies the name "multiplicity" in the following definition.

Definition 12.4. For a G-representation V and an irreducible G-representation W the *multiplicity* $\mathrm{mult}_W(V)$ of W in V is defined as

$$\mathrm{mult}_W(V) := \dim \mathrm{Hom}_G(W, V).$$

Corollary 12.5. *If $U \subseteq V$ is a subrepresentation, then $\mathrm{mult}_W(U) \le \mathrm{mult}_W(V)$.*

Proof. If $U \subseteq V$, then $\mathrm{Hom}_G(W, U)$ is a linear subspace of $\mathrm{Hom}_G(W, V)$. □

Corollary 12.6. *If $U \twoheadrightarrow V$ is a G-equivariant surjection of representations, then $\mathrm{mult}_W(U) \ge \mathrm{mult}_W(V)$.*

Proof. Let $\varphi : U \twoheadrightarrow V$ be a G-equivariant surjection. Define the linear map $\kappa : \mathrm{Hom}_G(V, W) \to \mathrm{Hom}_G(U, W)$ by $\kappa(\psi) = \psi \circ \varphi$. It remains to show that κ is injective. For this we assume that $\kappa(\psi) = 0$, i. e., $\psi \circ \varphi = 0$. Since φ is surjective, it follows $\psi = 0$. □

In the case where $G = \mathsf{GL}_N$ we have the following very useful way of determining multiplicities:

Proposition 12.7. *If V is a GL_N-representation, then $\mathrm{mult}_\lambda(V) = \dim \mathsf{HWV}_\lambda(V)$, where $\mathsf{HWV}_\lambda(V)$ is the linear subspace of highest weight vectors of weight λ in V.*

Proof. Fix an irreducible GL_N-representation W_λ and fix a nonzero vector h from the 1-dimensional linear subspace of HWVs in W_λ. By Theorem 11.15(a'), if $0 \neq v \in \mathsf{HWV}_\lambda(V)$, then $\langle \mathsf{GL}_N v \rangle$ is irreducible. By Lemma 12.2 it follows that $\dim \mathrm{Hom}_G(W_\lambda, \langle \mathsf{GL}_N v \rangle) = 1$. Since every equivariant map maps HWVs of weight λ to HWVs of weight λ or to 0, every element in $\mathrm{Hom}_G(W_\lambda, \langle \mathsf{GL}_N v \rangle)$ has $\varphi(h) = \alpha v$ for some $\alpha \in \mathbb{C}$. Moreover, for each α there exists such a G-homomorphism.

Now we have the following isomorphism of vector spaces $\mathsf{HWV}_\lambda(V) \to \mathrm{Hom}_G(W_\lambda, V)$:

$$v \mapsto \Big(\varphi \in \mathrm{Hom}_G(W_\lambda, \langle \mathsf{GL}_N v \rangle),\ \varphi(h) = v\Big)$$

with inverse map $\varphi \mapsto \varphi(h)$. □

12.3 Isotypic components

In this section we assume that our group G is linearly reductive. The decomposition into irreducible representations might not be unique as soon as the multiplicity of some isomorphism type λ exceeds 1. In this section we group together isomorphic copies of the same irreducible representation to obtain the unique *isotypic decomposition*.

A representation V is called *isotypic* if V is a (not necessarily direct or finite) sum of irreducible representations that are all isomorphic.

Definition 12.8. Let G be a group and let V be a G-representation (in particular finite dimensional). Let W be an irreducible G-representation and define λ to to be its isomorphism type. The isotypic component V_λ of type λ is defined as the (possibly infinite) sum $\sum_i V_i$ of all irreducible subrepresentations of type λ.

For example the weight spaces in Section 11 are isotypic components, where the group G is the algebraic torus.

Lemma 12.9. *An isotypic representation of type λ decomposes into a direct sum of irreducibles of type λ.*

Proof. Let V be isotypic and write $V = E_1 + \cdots + E_t$ with E_i irreducible of type λ and t minimal. Clearly t is finite because $\dim V$ is finite and $\dim E_i \geq 1$. For the sake of contradiction assume that the sum is not direct: There exists $x_i \in E_i$ such that $x_1 + \cdots + x_t = 0$ and w.l.o.g. $x_t \neq 0$. Thus $x_t = -(x_1 + \cdots + x_{t-1})$ and hence $(E_1 + \cdots + E_{t-1}) \cap E_t \neq 0$. Since E_t is irreducible: $E_t \subseteq E_1 + \cdots + E_{t-1}$, which is a contradiction to t being minimal. $\square$

Proposition 12.10. *Every representation V decomposes into a direct sum of isotypic representations $V = \bigoplus_\lambda V_\lambda$, where λ runs over all types of irreducible representations.*

Proof. Let $V = M_1 \oplus \cdots \oplus M_k$ be a decomposition into irreducibles. Then $\mathrm{mult}_\lambda(V)$ equals the number of times for which M_i is of type λ. Define the direct sum

$$M := \bigoplus_{i \text{ with } M_i \text{ of type } \lambda} M_i,$$

so $\dim M = \mathrm{mult}_\lambda(V) \cdot \dim \lambda$, where $\dim \lambda$ denotes the dimension of the irreducible representation of type λ. It remains to show that $M = V_\lambda$, because then we see that the direct sum of isotypic components results from adding up isomorphic copies of irreducible representations. Clearly $M \subseteq V_\lambda$.

Since V_λ decomposes into a direct sum of irreducibles of type λ (Lemma 12.9), the number of summands in this decomposition is $\mathrm{mult}_\lambda(V_\lambda)$ by Cor. 12.3. Therefore $\dim V_\lambda = \mathrm{mult}_\lambda(V_\lambda) \cdot \dim \lambda$. Since $V_\lambda \subseteq V$ it follows $\mathrm{mult}_\lambda(V_\lambda) \leq \mathrm{mult}_\lambda(V)$ and thus $\dim V_\lambda \leq \mathrm{mult}_\lambda(V) \cdot \dim \lambda = \dim M$.

Since $\dim V_\lambda \leq \dim M$ and $M \subseteq V_\lambda$, we conclude $M = V_\lambda$. $\square$

12.4 Using multiplicities or occurrences as obstructions

Let $\mathbb{A} = \mathbb{C}^\eta$. For a Zariski-closed set $Z \subseteq \mathbb{A}$ we define the *coordinate ring*

$$\mathbb{C}[Z] := \mathbb{C}[\mathbb{A}]/I(Z)$$

If Z is a cone, then $\mathbb{C}[Z]$ is graded via $\mathbb{C}[Z]_\delta = \mathbb{C}[\mathbb{A}]_\delta / I(Z)_\delta$.

Lemma 12.11. *Let G be a group, V be a G-representation, and let there exist a G-invariant inner product on V. If $U \subseteq V$ is a subrepresentation, then the quotient V/U is also a G-representation. More precisely, $V \cong U \oplus V/U$.*

Proof. Since U is a G-representation, for $g \in G$ we have $gU = U$ as a set. Thus if $v \in V$, then $v + U \in V/U$ and $g(v+U) = gv + gU = gv + U \in V/U$.

If we have a G-invariant inner product, then $V = U \oplus U^\perp$. We now show that $U^\perp \simeq V/U$. Let $p : V \twoheadrightarrow U^\perp$ be the projection that sends U to 0. The equivariant isomorphism $V/U \to U^\perp$ is given by $v + U \mapsto p(v+U) = p(v)$ with inverse map $w \mapsto w + U$. □

From the lemma we conclude that if $\mathbb{C}[\mathbb{A}]_\delta$ is a G-representation and $I(Z)_\delta$ is a G-representation, then $\mathbb{C}[Z]_\delta$ is a G-representation.

An approach towards proving complexity lower bounds goes as follows. Let $Z' \subseteq Z$ be a Zariski-closed cone that is closed under the action of GL_M. Think of $\overline{\mathsf{GL}_{n^2+1} T^{n-m} \mathrm{per}_m} \subseteq \overline{\mathsf{GL}_{n^2+1} \det_n}$ or of $\overline{\mathsf{GL}_{n^2} X_{1,1}^{n-m} \mathrm{per}_m} \subseteq \overline{\mathsf{GL}_{n^2} \det_n}$ for some fixed values of n and m. Then $I(Z)_\delta \subseteq I(Z')_\delta$ and thus we obtain a canonical GL_{n^2}-equivariant surjection $\mathbb{C}[Z]_\delta \twoheadrightarrow \mathbb{C}[Z']_\delta$. By Schur's lemma (Cor. 12.6) this implies $\mathrm{mult}_\lambda(\mathbb{C}[Z]) \geq \mathrm{mult}_\lambda(\mathbb{C}[Z'])$.

Thus if we want to prove $Z' \not\subseteq Z$, it is sufficient to show the existence of some λ that satisfies $\mathrm{mult}_\lambda(\mathbb{C}[Z]) < \mathrm{mult}_\lambda(\mathbb{C}[Z'])$. Such λ are called *representation theoretic multiplicity obstructions*. If $\mathrm{mult}_\lambda(\mathbb{C}[Z]) = 0 < \mathrm{mult}_\lambda(\mathbb{C}[Z'])$, then these λ a called *occurrence obstructions*.

Mulmuley and Sohoni conjectured that one could separate $\mathrm{VNP} \not\subseteq \overline{\mathrm{VP}_{ws}}$ by using occurrence obstructions, but this was recently rejected:

Conjecture 12.12. *For every polynomial p there exist infinitely many m and $n \geq p(m)$ with: If $Z' := \overline{\mathsf{GL}_{n^2} X_{1,1}^{n-m} \mathrm{per}_m}$ and $Z := \overline{\mathsf{GL}_{n^2} \det_n}$, then there exists λ with $\mathrm{mult}_\lambda(\mathbb{C}[Z']) > 0 = \mathrm{mult}_\lambda(\mathbb{C}[Z])$.*

Theorem 12.13 ([23])**.** *Let $n \geq m^{25}$ and let $Z' := \overline{\mathsf{GL}_{n^2} X_{1,1}^{n-m} \mathrm{per}_m}$ and $Z := \overline{\mathsf{GL}_{n^2} \det_n}$. If $\mathrm{mult}_\lambda(\mathbb{C}[Z']) > 0$, then $\mathrm{mult}_\lambda(\mathbb{C}[Z]) > 0$.*

It is an open problem if multiplicities can be used to separate orbit closures. More specifically, it is open if $\mathrm{VNP} \not\subseteq \overline{\mathrm{VP}_{ws}}$ can be proved using representation theoretic multiplicity obstructions.

12.4.1 Plethysm coefficients

When we try to find λ such that $\mathrm{mult}_\lambda(\mathbb{C}[Z']) > \mathrm{mult}_\lambda(\mathbb{C}[Z]) > 0$, then a necessary condition is that the plethysm coefficient exceeds $\mathrm{mult}_\lambda(\mathbb{C}[Z])$, which we explain in this section (see Lemma 12.14).

Let $\mathbb{A} = \mathbb{C}[X_1, \dots, X_M]_d$. Fix $\delta \in \mathbb{N}$. Let λ be a partition of $d\delta$. Define the *plethysm coefficient* as

$$a_\lambda(\delta, d) := \mathrm{mult}_\lambda(\mathbb{C}[\mathbb{A}]_\delta).$$

We will see later (Proposition 19.8) that $a_\lambda(\delta, d)$ basically does not depend on M: Define $\ell(\lambda) := \max\{i \mid \lambda_i > 0\}$. If $\ell(\lambda) > M$, then $a_\lambda(\delta, d) = 0$. On the other hand $a_\lambda(\delta, d)$ has the same value for all $M \geq \ell(\lambda)$. Therefore we define $a_\lambda(\delta, d)$ to be the value for large M.

Finding a combinatorial description for $a_\lambda(\delta, d)$ is a major open problem in algebraic combinatorics. It is problem 9 on Stanley's problem list from 2000 [74]. In terms of theoretical computer science, this quesion can be phrased as: Is the function $(\lambda, \delta, d) \mapsto a_\lambda(\delta, d)$ in the complexity class $\#P$? Here we are allowed to encode the partition λ in unary.

The SCHUR software and the LIE software can compute plethysm coefficients.

Lemma 12.14. *There exists λ with $\mathrm{mult}_\lambda(\mathbb{C}[\overline{Gv}]_\delta) < a_\lambda(\delta, d)$ iff the type λ occurs in the vanishing ideal $I(\overline{Gv})_\delta$.*

Proof. $\mathbb{C}[\mathbb{A}]_\delta = I(\overline{Gv})_\delta \oplus \mathbb{C}[\overline{Gv}]_\delta$ and thus $a_\lambda(\delta, d) = \mathrm{mult}_\lambda(\mathbb{C}[\overline{Gv}]_\delta) + \mathrm{mult}_\lambda(I(\overline{Gv})_\delta)$. □

Multiplicities

Representation-theoretic multiplicities count how often an irreducible representation occurs in a decomposition into irreducibles.
The vanishing ideal and the coordinate ring are dual notions. Their multiplicities add up to the plethysm coefficient.
An attack route towards finding obstructions goes via comparing multiplicities in coordinate rings of orbit closures. Occurrence obstructions are known not to separate VP_{ws} from VNP.

13 Tensors for computer scientists

In this chapter we discuss tensors. This will serve mainly two purposes: To discuss the computational complexity of bilinear maps using geometric complexity theory, and to explicitly construct the irreducible representations of GL_n and their highest weight vectors.

13.1 Bilinear forms

Let U and V be vector spaces over $\mathbb{F}$. All vector spaces are assumed to be finite dimensional. Let $f : U \times V \to \mathbb{F}$ be a bilinear form, that is, a form which is linear in both components. We denote the set of all bilinear forms by $\mathsf{Bil}(U, V; \mathbb{F})$. A linear form $\ell : U \to \mathbb{F}$ is uniquely determined when we know its values at any basis $u_1, \dots, u_m$ of U. How about f?

Lemma 13.1. *Let $u_1, \dots, u_m$ and $v_1, \dots, v_n$ be bases of U and V, respectively. Then f is uniquely determined by the values $f_{i,j} := f(u_i, v_j)$, $1 \leq i \leq m$, $1 \leq j \leq n$.*

Proof. Let $g : U \times V \to \mathbb{F}$ be another bilinear form with $g(u_i, v_j) = f_{i,j}$, $1 \le i \le m$, $1 \le j \le n$. Let $u = \sum_{i=1}^{m} \alpha_i u_i$ and $v = \sum_{j=1}^{n} \beta_j v_j$ be arbitrary. We have

$$\begin{aligned} g(u,v) &= g(\sum_{i=1}^{m} \alpha_i u_i, \sum_{j=1}^{n} \beta_j v_j) \\ &= \sum_{i=1}^{m} \alpha_i g(u_i, \sum_{j=1}^{n} \beta_j v_j) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \beta_j g(u_i, v_j) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \beta_j f_{i,j} \\ &= f(u,v). \end{aligned}$$

Note that to get the last line, we used bilinearity again. □

By choosing the bases, we identiy U with $\mathbb{F}^m$ and V with $\mathbb{F}^n$. Now, we can write f even more concretely as

$$f(x,y) = \sum_{i=1}^{m} \sum_{j=1}^{n} f_{i,j} x_i y_j.$$

(As a golden rule, you should avoid specifying a basis unless it is really neccessary. However, it is at first more intuitive to think in terms of bases.) You usually think of x_i as an indeterminate, and to evaluate f we substitute the value α_i for x_i. But you can also think of x_i being a linear form mapping (by substitution) a vector $\sum_{i=1}^{m} \alpha_i u_i$ to α_i, that is, $x_1, \dots, x_m$ is a dual basis to $u_1, \dots, u_m$. The same is true for $y_1, \dots, y_n$. The products of linear forms $x_i y_j$ form a basis of the linear space of bilinear forms $U \times V \to \mathbb{F}$. Recall that the set of all linear forms on U or V are denoted by U^* of V^*, respectively.

Definition 13.2. The space of all bilinear forms $U \times V \to \mathbb{F}$ is called the tensor product of U^* and V^* and is denoted by $U^* \otimes V^*$.

Let $x = \sum_{i=1}^{m} \alpha_i x_i \in U^*$ and $y = \sum_{j=1}^{n} \beta_j y_j \in V^*$. We have

$$xy = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \beta_j x_i y_j.$$

How do you get the tensor product $U \otimes V$? You simply start with bilinear forms $U^* \times V^* \to \mathbb{F}$. While this looks complicated at a first glance—bilinear forms mapping pairs of linear forms to scalars—also U^* is just a vector space and once you choose a basis, everything is isomorphic to some $\mathbb{F}^m$.

Exercise 13.3. Prove that $\mathsf{Hom}(U, V) \cong U^* \otimes V$. While this is absolutely clear to every mathematican, computer scientists tend to forget about this pretty soon. (Or even never learned it this way.)

The previous exercise identifies linear maps $U \to V$ with bilinear forms on $U \times V^*$. Both objects are specified by a two-dimensional array of field elements and we interpret this data in two different ways. So $\mathsf{Hom}(U, V) \cong \mathsf{Bil}(U, V^*; \mathbb{F})$ essentially says nothing. However, by re-interpreting objects in the right way, one can often prove astonishing facts very quickly. You should get used to this!

13.2 Universal property

We can define tensor products also in terms of a universal property. A tensor product of two spaces U and V is a vector space, denoted by $U \otimes V$, together with a bilinear map $\phi : U \times V \to U \otimes V$ such that for any bilinear map $b : U \times V \to W$, there is a unique linear map $\ell : U \otimes V \to W$ such that $b = \ell \circ \phi$. Given a tensor product, we set $u \otimes v := \phi(u, v)$ for every $u \in V$ and $v \in V$.

Theorem 13.4. *Let U and V be (finite-dimensional) vector spaces.*

1. *U and V have a tensor product.*
2. *Any two tensor products of U and V are isomorphic.*
3. *If $u_1, \dots, u_m$ is a basis of U and $v_1, \dots, v_n$ is a basis of V, then $u_i \otimes v_j$, $1 \le i \le m$, $1 \le j \le n$ is a basis of $U \otimes V$.*

Proof. To prove the first item, we construct an explicit tensor product. It will be the construction of the previous section. We set $U \otimes V = \mathsf{Bil}(U^*, V^*; \mathbb{F})$ and $\phi(u, v)(x, y) = x(u) \cdot y(v)$. Then the third item immediately follows from the discussion right after Lemma 13.1.

Let $b : U \times V \to W$ be a bilinear map. We choose a basis $u_1, \dots, u_m$ of U and a basis $v_1, \dots, v_n$ of V. To finish the proof of the first item, we define the linear map $\ell : U \otimes V \to W$ by $\ell(u_i \otimes v_j) := b(u_i, v_j)$, $1 \le i \le m$, $1 \le j \le n$. Note that for $u = \sum_{i=1}^m \alpha_i u_i$ and $\sum_{j=1}^n \beta_j v_j$, we have

$$\begin{aligned}
\ell(u \otimes v) &= \ell(\sum_{i=1}^m \alpha_i u_i \otimes \sum_{j=1}^n \beta_j v_j) \\
&= \ell(\sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j u_i \otimes v_j) \\
&= \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j \ell(u_i \otimes v_j) \\
&= \sum_{i=1}^m \sum_{j=1}^n \alpha_i \beta_j b(u_i, v_j) \\
&= b(u, v),
\end{aligned}$$

so $\ell \circ \phi$ and b are equal. ℓ is unique, since it is defined on a basis of $U \times V$.

It remains to prove the second item. Assume we have two tensor products $U \otimes V$, ϕ and $U \otimes' V$, ϕ'. We apply the definition of tensor product to $U \otimes V$ and ϕ and let the bilinear map $b = \phi'$ and the vector space $W = U \otimes' V$. We get a linear map $\ell : U \otimes W \to U \otimes' V$ such that $\ell \circ \phi = \phi'$. In the same way, by interchanging the roles of the two tensor products, we get a linear map $\ell' : U \otimes' V \to U \otimes V$. The situation is depicted below:

$$\begin{tikzcd}
& U \otimes V \arrow[dd, bend left, "\ell"] \\
U \times V \arrow[ur, "\phi"] \arrow[dr, "\phi'"'] & \\
& U \otimes' V \arrow[uu, bend left, "\ell'"]
\end{tikzcd}$$

We have $\ell' \circ \ell \circ \phi = \ell' \circ \phi' = \phi$. We apply the definition of tensor product to $U \otimes V$ and ϕ and let the bilinear map be ϕ and the vector space $W = U \times V$. Then the linear map can be the identity and it can be $\ell' \circ \ell$ by the equation above. By the uniqueness of the linear map, we get that $\ell' \circ \ell$ is the identity (on $U \otimes V$). In the same way, we get that $\ell \circ \ell'$ is the identity (on $U \otimes' V$). Thus ℓ and ℓ' are isomorphisms. ◻

Exercise 13.5. Let U, V, and W be vector spaces. Prove the following (the isomorphisms can be proved using the univesal property):

1. $U \otimes V \cong V \otimes U$.
2. $U \otimes (V \otimes W) \cong (U \otimes V) \otimes W$.
3. $\dim U \otimes V = \dim U \cdot \dim V$.

The second item says that the tensor product is associative (up to isomorphisms), therefore, we simply can write $U \otimes V \otimes W$. One could also define a threefold tensor product directly by defining it as the vector space of trilinear form $U^* \times V^* \times W^* \to \mathbb{F}$. In the same way, we can build the tensor product of an arbitrary number of vector spaces.

13.3 Tensor rank

Elements of the form $u \otimes v \in U \otimes V$ are called *elementary* or *decomposable* or *rank-one tensors*. Not all elements are elementary, for instance $u_1 \otimes v_1 + u_2 \otimes v_2$ is not elementary when u_1 and u_2 as well as v_1 and v_2 are linearly independent. In general, if we have a tensor product $V_1 \otimes \cdots \otimes V_k$, we call elements of the form $v_1 \otimes \cdots \otimes v_k$ with $v_i \in V_i$ elementary (or decomposable or rank-one tensors).

The rank of a matrix M can be defined as the minimum number r of rank one matrices $S_1, \dots, S_r$ such that $M = S_1 + \cdots + S_r$. In the same way, we define the rank of a tensor $t \in V_1 \otimes \cdots \otimes V_k$ as the minimum number of rank-one tensors $s_1 \otimes \cdots \otimes s_r \in V_1 \otimes \cdots \otimes V_k$ such that

$$t = s_1 + \cdots + s_r.$$

We denote the rank of a tensor by $R(t)$.

Note that this generalizes the rank of a matrix. Any matrix M can be interpreted as an element of $U^* \otimes V$. A rank-one matrix S can be written as $S = a \cdot b$ where a is a column vector and b is a row vector. Then for any column vector x,

$$S \cdot x = (a \cdot b) \cdot x = (b \cdot x) \cdot a,$$

because $b \cdot x$ is a 1×1 matrix. In this way, we can interpret b as a linear form on U.

Note that for matrices, we have further equivalent definitions of rank. In particular, there are efficient algorithms for computing the rank. This is not true for tensors in a threefold (or higher) tensor product. Here the problem is NP-hard [43], even complete for the existential theory over the underlying ground field $\mathbb{F}$ [71, 72], and also hard to approximate [12, 78, 73].

13.4 Actions on tensor products

Let $V_1, \ldots, V_k$ and $U_1, \ldots, U_k$ be vector spaces and let $A_i \in \mathsf{Hom}(V_i, U_i)$, $1 \le i \le k$. We can extend the A_i to a homomorphism

$$A_1 \otimes \cdots \otimes A_k : V_1 \otimes \cdots \otimes V_k \to U_1 \otimes \cdots \otimes U_k$$

in the following way: Let $v_1 \otimes \cdots \otimes v_k \in V_1 \otimes \cdots \otimes V_k$. We set

$$A_1 \otimes \cdots \otimes A_k(v_1 \otimes \cdots \otimes v_k) = A_1(v_1) \otimes \cdots \otimes A_k(v_k)$$

and extend $A_1 \otimes \cdots \otimes A_k$ to $V_1 \otimes \cdots \otimes V_k$ by linearity.

Exercise 13.6. Prove that $A_1 \otimes \cdots \otimes A_k$ is well-defined, that is, if we decompose a tensor t in two different ways into rank-one tensors, then we get the same result.

Definition 13.7. Let $t \in V_1 \otimes \cdots \otimes V_k$ and $s \in U_1 \otimes \cdots \otimes U_k$. We call s a *restriction* of t and write $s \le t$ if there are $A_i \in \mathsf{Hom}(V_i, U_i)$, $1 \le i \le k$, such that $A_1 \otimes \cdots \otimes A_k(t) = s$.

The proof of the following lemma is obvious.

Lemma 13.8. *If s is a restriction of t, then $R(s) \le R(t)$.*

We can let $\mathsf{End}(V_1) \times \cdots \times \mathsf{End}(V_k)$ act on $V_1 \otimes \cdots \otimes V_k$ by

$$(A_1, \ldots, A_k)t = A_1 \otimes \cdots \otimes A_k(t).$$

If U_i is a subspace of V_i, then we can write the fact that s is a restriction of t as a monoid orbit problem, namely, $s \le t$ iff

$$s \in (\mathsf{End}(V_1) \times \cdots \times \mathsf{End}(V_k))t.$$

Note that by Lemma 13.8, this means that $R(s) \le R(t)$. In the next chapter, we will see how we can interpret this in terms of complexity.

The language of tensors

The language of tensors is a natural way of describing multilinear maps.
This will help us in the study of the complexity of bilinear maps.
Moreover, tensor products are fundamental building blocks in the representation theory of GL_n.

14 Complexity of bilinear maps

The following two chapters give a brief introduction to the tensor rank problem and its relation to fast matrix multiplication. Many results have been taken from [10], nevertheless we decided to restate them explicitly for the reader's convenience. For even more details, the reader is referred to [10] and the references given there.

14.1 Strassen's algorithm

Given a $k \times m$-matrix $x = (x_{hi})$ and and $m \times n$-matrix $y = (y_{ij})$ whose entries are indeterminates over some field $\mathbb{F}$, we want to compute their product $xy = (z_{hj})$. The entries z_{hj} are given by

$$z_{hj} = \sum_{i=1}^{m} x_{hi} y_{ij}, \qquad 1 \leq h \leq k, \;\; 1 \leq j \leq m. \tag{14.1}$$

In 1969, Strassen [75] found a way to multiply 2×2-matrices with only 7 multiplications but 18 additions.

Let z_{ij}, $1 \leq i, j \leq 2$, be given by

$$\begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}.$$

We compute the seven products

$$\begin{aligned}
p_1 &= (x_{11} + x_{22})(y_{11} + y_{22}), \\
p_2 &= (x_{21} + x_{22}) y_{11}, \\
p_3 &= x_{11}(y_{12} - y_{22}), \\
p_4 &= x_{22}(-y_{11} + y_{21}), \\
p_5 &= (x_{11} + x_{12}) y_{22}, \\
p_6 &= (-x_{11} + x_{21})(y_{11} + y_{12}), \\
p_7 &= (x_{12} - x_{22})(y_{21} + y_{22}).
\end{aligned}$$

We can express each of the z_{ij} as a linear combination of these seven products, namely,

$$\begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} = \begin{pmatrix} p_1 + p_4 - p_5 + p_7 & p_3 + p_5 \\ p_2 + p_4 & p_1 + p_3 - p_2 + p_6 \end{pmatrix}.$$

By applying this construction recursively, we get the well-known algorithm which multiplies matrices in time $O(n^{\log_2 7})$. To make the recursion work, it is crucial that the entries are bilinear products.

14.2 Relation to tensor rank

Assume that in general, we have k bilinear forms

$$z_h = \sum_{i=1}^{m} \sum_{j=1}^{n} t_{i,j,h} x_i y_j, \qquad h = 1, \dots, k$$

and we have r bilinear products

$$p_\rho = (u_{\rho,1} x_1 + \cdots + u_{\rho,m} x_m)(v_{\rho,1} y_1 + \cdots + v_{\rho,m} y_m)$$

such that we can write each z_h as a linear combination of them, that is,

$$z_h = w_{1,h} p_1 + \cdots + w_{r,h} p_r.$$

We can view the "array" $t = (t_{h,i,j})$ as a tensor in $\mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$. The products $p_1, \dots, p_r$ correspond to a decomposition of t into rank-one tensors: Namely, let $w_\rho = (w_{\rho,1}, \dots, w_{\rho,k})$, $1 \leq \rho \leq r$ and define u_ρ and v_ρ accordingly. Then

$$w_\rho \otimes u_\rho \otimes v_\rho = (w_{\rho,h} u_{\rho,i} v_{\rho,j})$$

and by comparing coefficients, we get that

$$t = \sum_{\rho=1}^{r} w_\rho \otimes u_\rho \otimes v_\rho.$$

In the same way, if we have a decomposition of t into r rank-one tensors, then we can obtain r bilinear products such that each z_h is contained in their linear span. Therefore, the minimal number of such products is precisely $R(t)$.

14.3 The exponent of matrix multiplication

We denote the tensor of the multiplication of $k \times m$-matrices with $m \times n$-matrices by $\langle k, m, n \rangle$. The corresponding tensor lives in $\mathbb{F}^{k \times m} \otimes \mathbb{F}^{m \times n} \otimes \mathbb{F}^{n \times k}$. We here transpose the matrices in the last component for symmetry reasons. Note that every component is indexed by double-indices. We have

$$z_{j',h} = \sum_{i=1}^{m} x_{h,i} y_{i,j'} = \sum_{h'=1}^{k} \sum_{i=1}^{m} \sum_{i'=1}^{m} \sum_{j=1}^{n} \delta_{h,h'} \delta_{i,i'} \delta_{j,j'} x_{h',i} y_{i',j},$$

Thus $\langle k, m, n \rangle = (\delta_{h,h'} \delta_{i,i'} \delta_{j,j'})$. Figure 9 contains an explicit description of the tensor of 2×2-matrix multiplication.

	$x_{1,1}$	$x_{1,2}$	$x_{2,1}$	$x_{2,2}$
$y_{1,1}$	$(1,1)$		$(1,2)$	
$y_{2,1}$		$(1,1)$		$(1,2)$
$y_{1,2}$	$(2,1)$		$(2,2)$	
$y_{2,2}$		$(2,1)$		$(2,2)$

Figure 9: The tensor of 2×2-matrix multiplication. It is $\{0,1\}$-valued. An entry (h,j) in the row (h,i) and column (i,j) means that $x_{h,i}y_{i,j}$ appears in $z_{j,i}$. Recall that we transposed the third component.

Definition 14.1. $\omega = \inf\{\beta \mid R(\langle n,n,n\rangle) \leq O(n^{\beta})\}$ is called the *exponent of matrix multiplication*.

In the definition of ω above, we only count bilinear products. For the asymptotic growth, it does not matter whether we count all operations or only bilinear products [7], see also [19, 10]. Let $\tilde{\omega}$ be the infimum over all β such that there is a family of arithmetic circuits of size $O(n^{\beta})$ computing the product of two $n \times n$-matrices. Since these circuits compute forms of degree two, we can make these circuits homogeneous such that the only nonscalar multiplications are products of linear forms.

Theorem 14.2. $\omega = \tilde{\omega}$.

Proof. We first prove $\omega \leq \tilde{\omega}$: Consider an arbitrary circuit computing the product of two matrices. Let r be number of nonscalar multiplications in it. As in the transformation of arbitrary circuits into homogeneous ones, we now compute with each homogeneous component separately. Note that since the output of each circuit is homogeneous of degree two, we only need to keep the components of degree up to two. The only nonscalar multiplications that we need to perform are the multiplications between the degree-one-terms, which is a product of linear forms. Therefore, we can modify the circuit as follows: We first compute several linear forms, then we perform r multiplications in them and then we compute linear combinations of the r products.

Does this prove that the rank is bounded by r, too? Not quite. The linear forms can be linear forms in the entries of both matrices. Consider such a product $u(X,Y)v(X,Y)$. We can write $u(X,Y) = u'(X) + u''(Y)$. We do the same for v. Then

$$u(X,Y)v(X,Y) = u'(X)v'(X) + u'(X)v''(Y) + u''(Y)v'(X) + u''(Y)v''(Y).$$

Since the outputs are all bilinear forms, the contribution of all $u'(X)v'(X)$ and of all $u''(Y)v''(Y)$ cancel. Therefore, we can replace the product above by two bilinear products $u'(X)v''(Y) + u''(Y)v'(X)$. Therefore, the rank is bounded by $2r$ and $\omega \leq \tilde{\omega}$.

For the other inequality, note that from the definition of ω, it follows that

$$\forall \epsilon > 0 : \exists \alpha \text{ and } m_0 > 1 : \forall m \geq m_0 : R(\langle m,m,m\rangle) \leq \alpha \cdot m^{\omega+\epsilon}.$$

Let $\epsilon > 0$ be given and choose m large enough. Let $r = R(\langle m,m,m\rangle)$.

To multiply $m^i \times m^i$-matrices we decompose them into blocks of $m^{i-1} \times m^{i-1}$-matrices and apply recursion. To multiply matrices of arbitrary sizes, we can pad with 0 to the next power of m. Let $A(n)$ be the number of arithmetic operations for the multiplication of $n \times n$-matrices with this approach. We obtain

$$A(n) \leq rA(n/m) + c(n/m)^2$$

where c is the number of additions and scalar multiplications that are performed by the chosen bilinear algorithm for $\langle m, m, m\rangle$ with r bilinear multiplications. Solving the recursion using the master theorem [29], we get $A(n) = O(n^{\log_m r})$. (Note that $r > m^2$ in general, so $\log_m r > 2$ and we are in the first case of the master theorem.)

Since $r \leq \alpha \cdot m^{\omega+\epsilon}$, we have $\log_m r \leq \omega + \epsilon + \log_m \alpha$. With $\epsilon' = \epsilon + \log_m \alpha$,

$$L(\langle n, n, n\rangle) = O(n^{\log_m r}) = O(n^{\omega+\epsilon'}).$$

Thus

$$\tilde{\omega} \leq \omega + \epsilon \qquad \text{for all } \epsilon > 0,$$

since $\log_m \alpha \to 0$ if $m \to \infty$. This means $\tilde{\omega} = \omega$, since $\tilde{\omega}$ is an infimum. □

Remark 14.3. The lower bound $R(\langle n, n, n\rangle) \geq \frac{5}{2}n^2 - 3n$ was obtained in [8]. [52] improved this lower bound to $R(\langle n, n, n\rangle) \geq 3n^2 - o(n^2)$.

14.4 Rank and restrictions

In the following, $\langle r\rangle$ denotes the tensor in $\mathbb{F}^r \otimes \mathbb{F}^r \otimes \mathbb{F}^r$ that has a 1 in the positions (ρ, ρ, ρ), $1 \leq \rho \leq r$, and 0s elsewhere (a "diagonal", the three-dimensional analogue of the identity matrix). This tensor corresponds to the r bilinear forms $x_\rho y_\rho$, $1 \leq \rho \leq r$ (r independent products) and is called the *unit tensor*.

Lemma 14.4. *$R(t) \leq r \Leftrightarrow t \leq \langle r\rangle$.*

Proof. "$\Leftarrow$": follows immediately from the observations that $s \leq s'$ implies $R(s) \leq R(s')$.

"$\Rightarrow$": $\langle r\rangle = \sum_{\rho=1}^{r} e_\rho \otimes e_\rho \otimes e_\rho$, where e_ρ is the ρth unit vector. If the rank of t is $\leq r$, then we can write t as the sum of r triads,

$$t = \sum_{\rho=1}^{r} u_\rho \otimes v_\rho \otimes w_\rho.$$

We define three homomorphisms

$$\begin{aligned} \alpha &: e_\rho \mapsto u_\rho, \;\; 1 \leq \rho \leq r, \\ \beta &: e_\rho \mapsto v_\rho, \;\; 1 \leq \rho \leq r, \\ \gamma &: e_\rho \mapsto w_\rho, \;\; 1 \leq \rho \leq r. \end{aligned}$$

By construction,

$$(\alpha \otimes \beta \otimes \gamma)\langle r\rangle = \sum_{\rho=1}^{r} \underbrace{\alpha(e_\rho)}_{=u_\rho} \otimes \underbrace{\beta(e_\rho)}_{=v_\rho} \otimes \underbrace{\gamma(e_\rho)}_{=w_\rho} = t.$$

□

Thus we can rephrase the question whether $R(\langle n,n,n\rangle) \le r$ as $\langle n,n,n\rangle \le \langle r\rangle$. Note that $\langle n,n,n\rangle$ and $\langle r\rangle$ live in general in different spaces, $(\mathbb{F}^{n\times n})^{\otimes 3}$ and $(\mathbb{F}^r)^{\otimes 3}$. For $r \ge n^2$ we can embed $\langle n,n,n\rangle$ into $(\mathbb{F}^r)^{\otimes 3}$ by padding the tensor with zeros. Therefore, the question $R(\langle n,n,n\rangle) \le r$ is equivalent whether the padded $\langle n,n,n\rangle$ is in the $\mathsf{End}(\mathbb{F}^r)^{\times 3}$-orbit of $\langle r\rangle$.

14.5 Permutations of matrix multiplication tensors

Let $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$ and $t = \sum_{j=1}^{r} t_j$ with rank-one tensors $t_j = a_{j1} \otimes a_{j2} \otimes a_{j3}$, $1 \le j \le r$. Let $\pi \in \mathfrak{S}_3$, where $\mathfrak{S}_3$ denotes the symmetric group on $\{1,2,3\}$. For a rank-one tensor t_j, let $\pi t_j = a_{j\pi^{-1}(1)} \otimes a_{j\pi^{-1}(2)} \otimes a_{j\pi^{-1}(3)}$ and $\pi t = \sum_{j=1}^{r} \pi t_j$. It is an easy exercise to prove that πt is well-defined. The proof of the following lemma is obvious.

Lemma 14.5. $R(t) = R(\pi t)$.

Let $t = (t_{h',i,i',j,j',h}) = \langle k,m,n\rangle$ and $\pi = (123)$. Then for $\pi t =: t' \in \mathbb{F}^{(n\times k)} \otimes \mathbb{F}^{(k\times m)} \otimes \mathbb{F}^{(m\times n)}$, we have

$$\begin{aligned} t'_{j',h,h',i,i',j} &= \delta_{j,j'}\delta_{h,h'}\delta_{i,i'} \\ &= \delta_{i,i'}\delta_{j,j'}\delta_{h,h'} \\ &= t_{h',i,i',j,j',i} \end{aligned}$$

Therefore,

$$R(\langle k,m,n\rangle) = R(\langle n,k,m\rangle) = R(\langle m,n,k\rangle).$$

Now, let $t'' = (t_{i,h',j,i',h,j'})$. We have $R(t) = R(t'')$, since permuting the "inner" indices corresponds to permuting the slices of the tensor.[9]

Next, let $\pi = (12)(3)$. Let $\pi t'' =: t''' \in \mathbb{F}^{(n\times m)} \otimes \mathbb{F}^{(m\times k)} \otimes \mathbb{F}^{(k\times n)}$. We have,

$$\begin{aligned} t'''_{j',i,i',h,h',j} &= \delta_{i,i'}\delta_{h,h'}\delta_{j,j'} \\ &= t_{h',i,i',j,j'h}. \end{aligned}$$

Therefore,

$$R(\langle k,m,n\rangle) = R(\langle n,m,k\rangle).$$

The second transformation corresponds to the well-known fact that $AB = C$ implies $B^T A^T = C^T$.

To summarize:

[9] We can think of a tensor $s = (s_{i,j,k}) \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$ as k matrices $s_i = (s_{i,j,k}) \in \mathbb{F}^{m\times n}$ stacked on top of each other. These are the 1-slices of s. In the same way, we define the 2- and 3-slices.

Lemma 14.6. *$R(\langle k, m, n\rangle) = R(\langle n, k, m\rangle) = R(\langle m, n, k\rangle) = R(\langle m, k, n\rangle) = R(\langle n, m, k\rangle) = R(\langle k, n, m\rangle)$.*

14.6 Products of matrix multiplication tensors

If we have two tensors $t \in U \otimes V \otimes W$ and $t' \in U' \otimes V' \otimes W'$, we can view their product $t \otimes t' \in (U \otimes V \otimes W) \otimes (U' \otimes V' \otimes W')$ as a tensor in $(U \otimes U') \otimes (V \otimes V') \otimes (W \otimes W')$ by using the natural isomorphism.

Lemma 14.7. *$R(t \otimes t') \leq R(t)R(t')$.*

Proof. Let $t = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$ and $t' = \sum_{i=1}^{r'} u'_i \otimes v'_i \otimes w'_i$. We have

$$\begin{aligned} t \otimes t' &= (\sum_{i=1}^{r} u_i \otimes v_i \otimes w_i) \otimes (\sum_{j=1}^{r'} u'_j \otimes v'_j \otimes w'_j) \\ &\sum_{i=1}^{r} \sum_{j=1}^{r'} (u_i \otimes v_i \otimes w_i) \otimes (u'_j \otimes v'_j \otimes w'_j) \\ &\sum_{i=1}^{r} \sum_{j=1}^{r'} (u_i \otimes u'_j) \otimes (v_i \otimes v'_j) \otimes (w_i \otimes w'_j). \end{aligned}$$

□

Note that for the rank, it can make a difference whether we view $t \otimes t'$ as a tensor in $(U \otimes U') \otimes (V \otimes V') \otimes (W \otimes W')$ or $U \otimes U' \otimes V \otimes V' \otimes W \otimes W'$. In the first case the number of inputs stays the same, we still compute bilinear forms. But the size of each input increases. In the second case, we would have five inputs, but their size stays the same. For complexity applications, we choose the first point of view. This is sometimes called a *vertical tensor product*.

Let $u_1, \dots, u_k$ be a basis of U, $v_1, \dots, v_m$ of V, and $w_1, \dots, w_k$ of W. Let $t_{h,i,j}$ be the coefficient of t of $u_h \otimes v_i \otimes w_j$. In the same way, choose bases for the other three spaces and let $t'_{h',i',j'}$ be the coefficient of t' of $u_{h'} \otimes v_{i'} \otimes w_{j'}$. Then the coefficient of $t \otimes t'$ of $(u_h \otimes u'_{h'}) \otimes (v_i \otimes v'_{i'}) \otimes (w_j \otimes w'_{j'})$ is $t_{h,i,j} t_{h',i',j'}$.

For the tensor product of matrix multiplications, we have

$$\begin{aligned} \langle k, m, n\rangle \otimes \langle k', m', n'\rangle &= (\delta_{\kappa\bar{\kappa}}\delta_{\mu\bar{\mu}}\delta_{\nu\bar{\nu}}\delta_{\kappa'\bar{\kappa}'}\delta_{\mu'\bar{\mu}'}\delta_{\nu'\bar{\nu}'}) \\ &= (\delta_{\kappa\bar{\kappa}}\delta_{\kappa'\bar{\kappa}'}\delta_{\mu\bar{\mu}}\delta_{\mu'\bar{\mu}'}\delta_{\nu\bar{\nu}}\delta_{\nu'\bar{\nu}'}) \\ &= \left(\delta_{(\kappa,\kappa'),(\bar{\kappa},\bar{\kappa}')}\delta_{(\mu,\mu'),(\bar{\mu},\bar{\mu}')}\delta_{(\nu,\nu'),(\bar{\nu},\bar{\nu}')}\right) \\ &= \langle kk', mm', nn'\rangle \end{aligned}$$

Thus, the tensor product of two matrix multiplication tensors is a bigger matrix multiplication tensor. This corresponds to the well known identity $(A \otimes B)(A' \otimes B') = (AA' \otimes BB')$ for the

Kronecker product of matrices. (Note that we use quadruple indices to address the entries of the Kronecker products and also of the slices of $\langle k, m, n\rangle \otimes \langle k', m', n'\rangle$.) It follows that the inequality in Lemma 14.7 can be strict. We have $R(\langle 2, 2, 2\rangle) = 7$, but there are faster ways to multiply matrices than Strassen's algorithm.

Using this machinery, we can show that whenever we can multiply matrices of a fixed format efficiently, then we get good bounds for ω.

Theorem 14.8. *If $R(\langle k, m, n\rangle) \leq r$, then $\omega \leq 3 \cdot \log_{kmn} r$.*

Proof. If $R(\langle k, m, n\rangle) \leq r$, then $R(\langle n, k, m\rangle) \leq r$ and $R(\langle m, n, k\rangle) \leq r$ by Lemma 14.6. Thus, by Lemma 14.7,

$$R(\underbrace{\langle k, m, n\rangle \otimes \langle n, k, m\rangle \otimes \langle m, n, k\rangle}_{=\langle kmn, kmn, kmn\rangle}) \leq r^3$$

and, with $N = kmn$,

$$R(\langle N^i, N^i, N^i\rangle \leq r^{3i} = (N^{3\log_N r})^i = (N^i)^{3\log_N r}$$

for all $i \geq 1$. Therefore, $\omega \leq 3 \log_N r$. $\square$

Thus, to get a fast matrix multiplication algorithm, it suffices to get a good upper bound on the rank of some fixed matrix multiplication tensor. However, Coppersmith and Winograd [28] prove that we cannot achieve $\omega = 2$ by starting with a fixed size matrix multiplication tensor.

Tensor rank and the exponent of matrix multiplication

The exponent of matrix multiplication ω can be expressed equivalently using arithmetic circuits and tensor rank.
Tensor rank can be studied via the restrictions of the unit tensor.

15 Border rank

15.1 Approximate computations

Over $\mathbb{R}$ or $\mathbb{C}$, the rank of matrices is semi-continuous. Let

$$\mathbb{C}^{n\times n} \ni A_j \to A = \lim_{j\to\infty} A_j$$

If for all j, $\mathrm{rk}(A_j) \leq r$, then $\mathrm{rk}(A) \leq r$ as $\mathrm{rk}(A_j) \leq r$ means all $(r+1)\times(r+1)$ minors vanish. But since minors are continuous functions, all $(r+1)\times(r+1)$ minors of A vanish, too.

The same is not true for 3-dimensional tensors. Consider the multiplication of univariate polynomials of degree one modulo X^2:

$$(a_0 + a_1 X)(b_0 + b_1 X) = a_0 b_0 + (a_1 b_0 + a_0 b_1)X + a_1 b_1 X^2$$

The tensor corresponding to the two bilinear forms $a_0 b_0$ and $a_1 b_0 + a_0 b_1$ consists of the two slices:

1	0
0	0

0	1
1	0

It has rank 3: To show the lower bound, we use a method called the substitution method. We first set $a_0 = 0$, $b_0 = 1$. Then we still compute a_1. Thus there is a product that depends on a_1, say one factor is $\alpha a_0 + \beta a_1$ with $\beta \neq 0$. When we replace a_1 by $-\frac{\alpha}{\beta} a_0$, we kill one product. We still compute $a_0 b_0$ and $-\frac{\alpha}{\beta} a_0 b_0 + a_0 b_1$. Next, set $a_0 = 1$, $b_0 = 0$. Then we still compute b_1. We can kill another product by substituting b_1 as above. After this, we still compute $a_0 b_0$, which needs one product.

However, we can approximate the tensor above by tensors of rank two. Let

$$t(\epsilon) = (1, \epsilon) \otimes (1, \epsilon) \otimes (0, \tfrac{1}{\epsilon}) + (1, 0) \otimes (1, 0) \otimes (1, -\tfrac{1}{\epsilon})$$

$t(\epsilon)$ obviously has rank two for every $\epsilon > 0$. The slices of $t(\epsilon)$ are

1	0
0	0

0	1
1	ϵ

Thus $t(\epsilon) \longrightarrow t$ if $\epsilon \longrightarrow 0$.

Bini, Capovani, Lotti and Romani [7] used this effect to design better matrix multiplication algorithms. They started with the following partial matrix multiplication tensor that we denote by $\{z_{11}, z_{12}, z_{21}\}$:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \left(\begin{array}{c|c} y_{11} & y_{12} \\ y_{21} & y_{22} \end{array} \right) = \left(\begin{array}{c|c} z_{11} & z_{12} \\ z_{21} & * \end{array} \right)$$

where we only want to compute three entries of the result. It can be shown using the substitution method that $R(\{z_{11}, z_{12}, z_{21}\}) = 6$, but we can approximate $\{z_{11}, z_{12}, z_{21}\}$ with only five products. Consider the following five products:

$$\begin{aligned}
p_1 &= (x_{12} + \epsilon x_{22}) y_{21}, \\
p_2 &= x_{11}(y_{11} + \epsilon y_{12}), \\
p_3 &= x_{12}(y_{11} + y_{21} + \epsilon y_{22}), \\
p_4 &= (x_{11} + x_{12} + \epsilon x_{21}) y_{11}, \\
p_5 &= (x_{12} + \epsilon x_{21})(y_{11} + \epsilon y_{22}).
\end{aligned}$$

We have

$$\begin{aligned}
\epsilon z_{11} &= \epsilon p_1 + \epsilon p_2 + O(\epsilon^2), \\
\epsilon z_{12} &= p_2 - p_4 + p_5 + O(\epsilon^2), \\
\epsilon z_{21} &= p_1 - p_3 + p_5 + O(\epsilon^2).
\end{aligned}$$

Here, $O(\epsilon^i)$ collects terms of degree i or higher in ϵ. Now we take a second copy of the partial matrix multiplication above, with new variables. With these two copies, we can multiply

2×2-matrices with 2×3-matrices (by identifying some of the variables in the copy). So we can approximate $\langle 2, 2, 3\rangle$ with 10 multiplications. If approximation would be as good as exact computation, then we would get $\omega \leq 2.78$ out of this, an improvement over Strassen's algorithm.

We will formalize the concept of approximation. We need to be more formal here, since we want to control the degree of approximation in order to turn approximate into exact computations. The role of the small quantity ϵ in the beginning of this chapter is now taken by the indeterminate ϵ.

Definition 15.1. Let $h \in \mathbb{N}$, $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$.

1. $R_h(t) = \min\{r \mid \exists u_\rho \in \mathbb{F}[\epsilon]^k, v_\rho \in \mathbb{F}[\epsilon]^m, w_\rho \in \mathbb{F}[\epsilon]^n : \sum_{\rho=1}^{r} u_\rho \otimes v_\rho \otimes w_\rho = \epsilon^h t + O(\epsilon^{h+1})\}$.

2. $\underline{R}(t) = \min_h R_h(t)$. $\underline{R}(t)$ is called the *border rank* of t.

Remark 15.2. 1. $R_0(t) = R(t)$.

2. $R_0(t) \geq R_1(t) \geq \ldots = \underline{R}(t)$.

3. For $R_h(t)$ it is sufficient to consider powers up to ϵ^h in u_ρ, v_ρ, w_ρ.

Above, we have used an algebraic definition of border rank. There is an equivalent geometric definition (see the end of this chapter), but the proof of equivalence is beyond the scope of this lecture.

Theorem 15.3 (Alder [3])**.** *Let U, V, and W be vector spaces over an algebraically closed field. The set of all tensors $t \in U \otimes V \otimes W$ with $\underline{R}(t) \leq r$ is the closure of the set of all tensors $s \in U \otimes V \otimes W$ with $R(s) \leq r$.*

15.2 Properties of border rank

Theorem 15.4. *Let $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$, $t' \in \mathbb{F}^{k'} \otimes \mathbb{F}^{m'} \otimes \mathbb{F}^{n'}$. We have*

1. *$\forall \pi \in \mathfrak{S}_3 : R_h(\pi t) = R_h(t)$.*

2. *$R_{h+h'}(t \otimes t') \leq R_h(t) \cdot R_{h'}(t')$.*

Proof. 1. Clear.

2. Let $t = (t_{i,j,l})$ and $t' = (t'_{i',j',l'})$. We have $t \otimes t' = (t_{i,j,l} \cdot t'_{i',j',l'}) \in \mathbb{F}^{kk'} \otimes \mathbb{F}^{mm'} \otimes \mathbb{F}^{nn'}$. Take two approximate computations for t and t' as above. Viewed as exact computations over $\mathbb{F}[[\epsilon]]$, their tensor product computes over the following:

$$T = \epsilon^h t + \epsilon^{h+1} s, \qquad T' = \epsilon^{h'} t' + \epsilon^{h'+1} s'$$

with $s \in \mathbb{F}[\epsilon]^k \otimes \mathbb{F}[\epsilon]^m \otimes \mathbb{F}[\epsilon]^n$ and $s' \in \mathbb{F}[\epsilon]^{k'} \otimes \mathbb{F}[\epsilon]^{m'} \otimes \mathbb{F}[\epsilon]^{n'}$. The tensor product of these two computations computes:

$$\begin{aligned} T \otimes T' &= (\epsilon^h t_{ijl} + \epsilon^{h+1} s_{ijl})(\epsilon^{h'} t'_{i'j'l'} + \epsilon^{h'+1} s'_{i'j'l'}) \\ &= (\epsilon^{h+h'} t_{ijl} t'_{i'j'l'} + O(\epsilon^{h+h'+1})) \\ &= \epsilon^{h+h'} t \otimes t' + O(\epsilon^{h+h'+1}) \end{aligned}$$

But this is an approximate computation for $t \otimes t'$.

□

15.3 From approximate to exact computations

The next lemma shows that we can turn approximate computations for matrix multiplication into exact ones. So for matrix multiplication, border rank is the right measure.

Lemma 15.5. *There is a constant c_h such that for all t: $R(t) \leq c_h R_h(t)$. c_h depends polynomially on h, in particular $c_h \leq \binom{h+2}{2}$.*

Remark 15.6. Over infinite fields, even $c_h = 1 + 2h$ works.

Proof. Let t be a tensor with border rank r and let

$$\sum_{\rho=1}^{r} \left(\sum_{\alpha=0}^{h} \epsilon^\alpha u_{\rho\alpha} \right) \otimes \left(\sum_{\beta=0}^{h} \epsilon^\beta v_{\rho\beta} \right) \otimes \left(\sum_{\gamma=0}^{h} \epsilon^\gamma w_{\rho\gamma} \right) = \epsilon^h t + O(\epsilon^{h+1})$$

The left-hand side of the equation can be rewritten as follows:

$$\sum_{\rho=1}^{r} \sum_{\alpha=0}^{h} \sum_{\beta=0}^{h} \sum_{\gamma=0}^{h} \epsilon^{\alpha+\beta+\gamma} u_{\rho\alpha} \otimes v_{\rho\beta} \otimes w_{\rho\gamma}$$

By comparing the coefficients of ϵ powers, we see that t is the sum of all $u_{\rho\alpha} \otimes v_{\rho\beta} \otimes w_{\rho\gamma}$ with $\alpha + \beta + \gamma = h$. Thus to compute t exactly, it is sufficient to compute $\binom{h+2}{2}$ products for each product in the approximate computation. □

The following theorem is the border rank version of Theorem 14.8.

Theorem 15.7. *If $\underline{R}(\langle k, m, n \rangle) \leq r$ then $\omega \leq 3 \log_{kmn} r$.*

Proof. Let $N = kmn$ and let $R_h(\langle k, m, n \rangle) \leq r$. By Theorem 15.4, we get $R_{3h}(\langle N, N, N \rangle) \leq r^3$ and $R_{3hs}(\langle N^s, N^s, N^s \rangle) \leq r^{3s}$ for all s. By Lemma 15.5, this yields $R(\langle N^s, N^s, N^s \rangle) \leq c_{3hs} r^{3s}$. Therefore,

$$\omega \leq \log_{N^s}(c_{3hs} r^{3s}) = 3s \log_{N^s}(r) + \log_{N^s}(c_{3hs}) = 3 \log_N(r) + \underbrace{\frac{1}{s} \log_N(\text{poly}(s))}_{\to 0}$$

for $s \to \infty$. Since ω is an infimum, we get $\omega \leq 3 \log_N(r)$. □

Corollary 15.8. $\omega \leq 2.78$.

Proof. Combine Theorem 15.7 with $\underline{R}(\langle 2,2,3\rangle) \leq 10$. □

15.4 Degeneration

Degenerations relate to border rank like restrictions relate to rank. Again, we only give an algebraic definition of degenerations and we simply state an equivalent topological definition. Furthermore, we choose coordinate right from the beginning, since it simplifies the notations somewhat.

Definition 15.9. Let $t \in \mathbb{F}^k \otimes \mathbb{F}^m \otimes \mathbb{F}^n$, $t' \in \mathbb{F}^{k'} \otimes \mathbb{F}^{m'} \otimes \mathbb{F}^{n'}$.

1. Let $t' = \sum_{\rho=1}^{r} u_\rho \otimes v_\rho \otimes w_\rho$ as well as $A(\epsilon) \in \mathbb{F}[\epsilon]^{k\times k'}$, $B(\epsilon) \in \mathbb{F}[\epsilon]^{m\times m'}$, and $C(\epsilon) \in \mathbb{F}[\epsilon]^{n\times n'}$. Define
$$(A(\epsilon) \otimes B(\epsilon) \otimes C(\epsilon))t' = \sum_{\rho=1}^{r} A(\epsilon)u_\rho \otimes B(\epsilon)v_\rho \otimes C(\epsilon)w_\rho.$$
(This is well-defined.)

2. t is a *degeneration* of t' if there are $A(\epsilon) \in \mathbb{F}[\epsilon]^{k\times k'}$, $B(\epsilon) \in \mathbb{F}[\epsilon]^{m\times m'}$, $C(\epsilon) \in \mathbb{F}[\epsilon]^{n\times n'}$, and $q \in \mathbb{N}$ such that
$$\epsilon^q t = (A(\epsilon) \otimes B(\epsilon) \otimes C(\epsilon))t' + O(\epsilon^{q+1}).$$
We will write $t \trianglelefteq_q t'$ or simply $t \trianglelefteq t'$.

As for the rank, it is very easy to prove the following lemma.

Lemma 15.10. *Let s and t be tensors.*

1. $t \trianglelefteq s \Rightarrow \underline{R}(t) \leq \underline{R}(s)$.
2. $\underline{R}(t) \leq r \Leftrightarrow t \trianglelefteq \langle r\rangle$.

The proof of the following theorem is beyond the scope of this lecture.

Theorem 15.11 (Strassen [77]). *Let $\mathbb{F}$ be algebraically closed. Let U, V, and W be vector spaces over $\mathbb{F}$. Let $t \in U \otimes V \otimes W$.*
$$\{s \in U \otimes V \otimes W \mid s \trianglelefteq t\} = \overline{\{s \in U \otimes V \otimes W \mid s \leq t\}}.$$

Since $\{s \in U \otimes V \otimes W \mid s \trianglelefteq t\}$ is constructible (it is the image of a constructible set under a polynomial map, see Theorem 4.12), the Zariski closure and the $\mathbb{C}$-closure coincide in Theorem 15.11 (see Theorem 4.15).

Let $t \in V \otimes V \otimes V$ and $\dim V = r$. $\underline{R}(t) \leq r$ is equivalent to $t \trianglelefteq \langle r\rangle$. If t lives in a smaller space $U \otimes U \otimes U$, we first embed it into $V \otimes V \otimes V$ by choosing a injective linear map $U \to V$. By the above theorem, $t \trianglelefteq \langle r\rangle$ is equivalent to
$$t \in \overline{\mathsf{End}(V)^{\times 3}\langle r\rangle},$$

so again, we have a (monoid) orbit closure problem. Since $\mathsf{GL}(V)$ lies dense in $\mathsf{End}(V)$, we can even just look at group orbits:

$$t \in \overline{\mathsf{GL}(V)^{\times 3}\langle r\rangle}.$$

Border rank: a geometric complexity measure

In the same way as defining the border determinantal complexity via the determinantal complexity, we define the border rank via the tensor rank.
As for the border determinantal complexity, border rank lower bounds can be defined via an orbit closure question.

16 Symmetric and alternating tensors

Let V be a vector space. Recall that the characteristic is zero. The symmetric group $\mathfrak{S}_d$ acts on $V^{\otimes d}$ by $(\pi, t = v_1 \otimes \cdots \otimes v_d) \mapsto \pi t = v_{\pi^{-1}(1)} \otimes \cdots \otimes v_{\pi^{-1}(d)}$ and linear extension to higher rank tensors. (In the previous chapters, we defined this for $d = 3$.)

16.1 S^2V and $\Lambda^2 V$

We start with the simplest examples. Let $v_1, \ldots, v_n$ be a basis of V. The space S^2V is defined as

$$S^2V = \langle v_i \otimes v_j + v_j \otimes v_i \mid 1 \leq i, j \leq n\rangle.$$

We call it the space of *symmetric 2-tensors of* V.

Proposition 16.1. *1.* $S^2V = \langle v \otimes v \mid v \in V\rangle$.

2. For $t \in V \otimes V$, $t \in S^2V$ *iff* $(1,2)t = t$.

Proof. We start with the first item. We have

$$(v_i + v_j) \otimes (v_i + v_j) - (v_i - v_j) \otimes (v_i - v_j) = 2(v_i \otimes v_j + v_j \otimes v_i).$$

Therefore, the left-hand side is contained in the right-hand side. On the other hand, if $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$, then

$$v \otimes v = \sum_{i<j} \alpha_i \alpha_j (v_i \otimes v_j + v_j \otimes v_i) + \sum_i \alpha_i^2 \tfrac{1}{2}(v_i \otimes v_i + v_i \otimes v_i).$$

Thus, the right-hand side is also contained in the left-hand side.

For the second item, notice that every tensor $t \in S^2V$ fulfills $(1,2)t = t$, since the basis does. For the other direction, let $t = \sum_{i,j} \alpha_{i,j} v_i \otimes v_j$. If $(1,2)t = t$, then $\alpha_{i,j} = \alpha_{j,i}$ for all i, j. Therefore, $t \in S^2V$. □

The *skew-symmetric* 2*-tensors of* V are defined as

$$\Lambda^2 V = \langle v_i \otimes v_j - v_j \otimes v_i \mid 1 \le i, j \le n \rangle.$$

Exercise 16.2. Prove the following:

1. $\Lambda^2 V = \langle v \otimes w - w \otimes v \mid v, w \in V \rangle$.
2. For all $t \in \Lambda^2 V$, $(1,2)t = -t$.

By the first item of Proposition 16.1 and Exercise 16.2, the spaces $S^2 V$ and $\Lambda^2 V$ are $\mathsf{GL}(V)$-invariant (where $\mathsf{GL}(V)$ acts simultaneously on both factors).

Proposition 16.3. $V \otimes V = S^2 V \oplus \Lambda^2 V$.

Proof. By the second item of Proposition 16.1 and Exercise 16.2, $S^2 V \cap \Lambda^2 V = \{0\}$. Furthermore, $v_i \otimes v_j = \frac{1}{2}(v_i \otimes v_j + v_j \otimes v_i) + \frac{1}{2}(v_i \otimes v_j - v_j \otimes v_i)$. □

Remark 16.4. Proposition 16.3 has a well-known interpretation for matrices: Every square matrix A can be written uniquely as the sum of a symmetric matrix and a skew-symmetric matrix. The decomposition is given by

$$A = \tfrac{1}{2}(A + A^T) + \tfrac{1}{2}(A - A^T)$$

16.2 Symmetric tensors

Let $\pi_S : V^{\otimes d} \to V^{\otimes d}$ be the map that is defined on rank-one tensors by

$$\pi_S(v_1 \otimes \cdots \otimes v_d) = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)}.$$

Definition 16.5. The *dth symmetric power of* V is defined as $S^d V := \pi_S(V^{\otimes d})$.

Note that this generalises the definition of $S^2 V$ in the previous section. We write $v_1 v_2 \ldots v_d := \pi_S(v_1 \otimes v_2 \otimes \cdots \otimes v_d)$.

Proposition 16.6. *For all* $t \in V^{\otimes d}$, $\pi_S(\pi_S(t)) = \pi_S(t)$, *that is,* π_S *is a projection.*

Proof. We have

$$\pi_S(\pi_S(v_1 \otimes \cdots \otimes v_d)) = \frac{1}{d!} \sum_{\tau \in \mathfrak{S}_d} \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} v_{\tau(\sigma(1))} \otimes \cdots \otimes v_{\tau(\sigma(d))}.$$

Since τ is a bijection, all $d!$ inner sums are the same. □

Proposition 16.7. *If* $w_1, \ldots, w_n$ *is a basis of* V, *then* $(w_{j_1} \cdots w_{j_d})_{1 \le j_1 \le \cdots \le j_d \le n}$, *is a basis of* $S^d V$.

Proof. If $\{j_1, \dots, j_d\} = \{i_1, \dots, i_d\}$, then

$$\pi_S(w_{j_1} \otimes \cdots \otimes w_{j_d}) = \pi_S(w_{i_1} \otimes \cdots \otimes w_{i_d}).$$

On the other hand, if $\{j_1, \dots, j_d\} \neq \{i_1, \dots, i_d\}$, then the terms appearing in the sums $\pi_S(w_{j_1} \otimes \cdots \otimes w_{j_d})$ and $\pi_S(w_{i_1} \otimes \cdots \otimes w_{i_d})$ are all distinct. Therefore, any linear dependency between $w_{j_1} \dots w_{j_d}, 1 \leq j_1 \leq \cdots \leq j_d \leq n$ would translate into a linear dependency between $w_{j_1} \otimes \cdots \otimes w_{j_d}$, $1 \leq j_1, \dots, j_d \leq n$. □

Corollary 16.8. $\dim S^d V = \binom{n+d-1}{d}$.

16.3 Alternating tensors

Let $\pi_\Lambda : V^{\otimes d} \longrightarrow V^{\otimes d}$ be the map that is defined on rank-one tensors by

$$\pi_\Lambda(v_1 \otimes \cdots \otimes v_d) = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} \operatorname{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)}.$$

Definition 16.9. The *dth alternating power of V* is defined as $\Lambda^d V := \pi_\Lambda(V^{\otimes d})$.

Exercise 16.10. π_Λ is a projection.

Again, this generalises the space $\Lambda^2 V$ of the first section. We write $v_1 \wedge v_2 \wedge \cdots \wedge v_d := \pi_\Lambda(v_1 \otimes v_2 \otimes \cdots \otimes v_d)$.

Proposition 16.11. $v_{\tau(1)} \wedge \cdots \wedge v_{\tau(d)} = \operatorname{sgn}(\tau) v_1 \wedge \cdots \wedge v_d$.

Proof. We have:

$$\begin{aligned}
v_{\tau(1)} \wedge \cdots \wedge v_{\tau(d)} &= \tfrac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} \operatorname{sgn}(\sigma) v_{\sigma(\tau(1))} \otimes \cdots \otimes v_{\sigma(\tau(d))} \\
&= \tfrac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} \operatorname{sgn}(\sigma \circ \tau^{-1}) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)} \\
&= \operatorname{sgn}(\tau^{-1}) \tfrac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} \operatorname{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)} \\
&= \operatorname{sgn}(\tau) v_1 \wedge \cdots \wedge v_d.
\end{aligned}$$

The third line follows from the fact that $\sigma \mapsto \sigma \circ \tau^{-1}$ is a bijection and the last line from the fact that $\operatorname{sgn}(\tau) = \operatorname{sgn}(\tau^{-1})$. □

Proposition 16.12. *We have $v = v_1 \wedge \cdots \wedge v_d = 0$ if and only if $v_1, \dots, v_d$ are linearly dependent.*

Proof. If two of the vectors are the same, say $v_1 = v_2$, then we can group the summands in $\pi_\Lambda(v)$ into pairs such that the two summands in the pair cancel. (Namely, if we switch the two identical vectors, we get the same tensor product but with opposite sign.)

In the general case, we can w.l.o.g. write $v_1 = \alpha_2 v_2 + \cdots + \alpha_d v_d$. Now by using linearity, we get a sum of tensors, each of which has two identical vectors.

For the other direction, assume that $v_1, \ldots, v_d$ are linearly independent. Enlarge the set $\{v_1, \ldots, v_d\}$ to a basis of V. Then in

$$v_1 \wedge \cdots \wedge v_d = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} \operatorname{sgn}(\sigma) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)},$$

all $v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)}$ are distinct basis vectors of $V^{\otimes d}$, hence $v_1 \wedge \cdots \wedge v_d$ cannot vanish. □

Proposition 16.13. *If $w_1, \ldots, w_n$ is a basis of V, then $w_{j_1} \wedge \cdots \wedge w_{j_d}$, $1 \leq j_1 < \cdots < j_d \leq n$, is a basis of $\Lambda^d V$.*

Proof. Given any tensor $v_1 \wedge \cdots \wedge v_d$, we can express each v_i as a linear combination of the basis vectors. Using linearity, we get a sum of alternating products of the basis vectors. Whenever two of the basis vectors are the same, the product vanishes by the previous proposition. Each product with pairwise distinct basis vectors can be brought into the form of the statement by permuting the vectors.

The vectors in the statement are obviously independent. □

Corollary 16.14. $\dim \Lambda^d V = \binom{n}{d}$.

Now let $d = n = \dim V$. Then $\Lambda^n V \cong \mathbb{C}$. $\mathsf{GL}(V)$ acts on $\Lambda^n V$ by

$$g(v_1 \wedge \cdots \wedge v_n) = gv_1 \wedge \cdots \wedge gv_n.$$

Let $gv_i = \sum_{j=1}^n \gamma_{j,i} v_j$. Then

$$\begin{aligned} g(v_1 \wedge \cdots \wedge v_n) &= (\sum_{j=1}^n \gamma_{j,1} v_j) \wedge \cdots \wedge (\sum_{j=1}^n \gamma_{j,n} v_j) \\ &= \sum_{j_1,\ldots,j_n} \gamma_{j_1,1} \ldots \gamma_{j_n,n} v_{j_1} \wedge \cdots \wedge v_{j_n}. \end{aligned}$$

In the last sum, only summands with pairwise distinct indices $j_1, \ldots, j_n$ are non-zero. Let σ be the permutation such that $\sigma(h) = j_h$ for all h. By Proposition 16.11, we have $v_{j_1} \wedge \cdots \wedge v_{j_n} = \operatorname{sgn}(\sigma) v_1 \wedge \cdots \wedge v_n$. Thus

$$g(v_1 \wedge \cdots \wedge v_n) = \det(g) v_1 \wedge \cdots \wedge v_n.$$

This is called the alternating representation.

Symmetric and alternating tensors

Two types of tensors of high importance are the symmetric and alternating tensors. They are defined by symmetrization and skew-symmetrization of tensors, respectively. These important tensors will serve as building blocks in the representation theory of GL_n.

17 The construction of the irreducible representations of the general linear group

We wish to understand much better the representation theory of coordinate rings of orbit closures. A first step into the right direction is to understand the building blocks: The irreducible representations of GL_n. They are completely understood in terms of combinatorial objects called Young tableaux. We follow the exposition in [36, Ch. 8] very closely.

17.1 Young tableaux

A Young diagram is a left-justified top-aligned array of boxes. To each partition λ we assign its Young diagram by interpreting λ_i as the number of boxes in row i. For example, the Young diagram corresponding to the partition $(5, 3, 3, 1)$ is

$$\ydiagram{5,3,3,1} \; .$$

We often identify partitions with their Young diagrams. The number of boxes in a Young diagram shall be denoted by $|\lambda| := \sum_i \lambda_i$.

If we fill the boxes of a Young diagram with numbers, we obtain a so-called *Young tableau*. For example,

$$\begin{ytableau} 6 & 2 & 4 & 4 & 8 \\ 2 & 1 & 6 \\ 4 & 3 & 1 \\ 3 \end{ytableau}$$

is a Young tableau. The partition corresponding to its Young diagram is called the *shape* of the Young diagram.

To simplify the notation, we define μ_i to be the number of boxes of the i-th column of λ. We call $\mu = (\mu_1, \mu_2, \ldots)$ the *transpose* of λ. The Young diagram of μ is obtained by transposing the Young diagram of λ.

We will need the notion of an *exchange*. This depends on a choice of two columns and a choice of k boxes in each column. For a Young tableau T of shape λ (with entries in any set) the corresponding exchange is the Young tableau S obtained from T by interchanging the entries in the two chosen sets of boxes, maintaining the vertical order in these; the entries outside these boxes are unchanged.

For example, if $\lambda = (4, 3, 3, 2)$ and the chosen boxes are the top two in the third column and the second and fourth in the second column, then the exchange takes

$$T = \begin{ytableau} 1 & 5 & 2 & 1 \\ 1 & 3 & 4 \\ 2 & 4 & 5 \\ 3 & 5 \end{ytableau} \quad \text{to } S = \begin{ytableau} 1 & 5 & 3 & 1 \\ 1 & 2 & 5 \\ 2 & 4 & 5 \\ 3 & 4 \end{ytableau} \; .$$

Sometimes we fix two columns and fix a subset of boxes in the right chosen column. The set of all corresponding exchanges are defined to have the same *exchange type*, i. e., an exchange type is a pair of columns together with a set of boxes from the right column.

17.2 Construction as a quotient space

Let $E = \mathbb{C}^n$ with the standard action of GL_n. All linear maps in the following constructions are equivariant, which defines the action of GL_n on the target space.

We write $E^{\times\lambda} = E \oplus E \oplus \cdots \oplus E$ ($|\lambda|$ times) and we associate each summand E with a position in the Young diagram of λ. In particular, if we write vectors in the boxes of λ, then we obtain an element of $E^{\times\lambda}$ and every element of $E^{\times\lambda}$ is obtained in this way. Next, we map $E^{\times\lambda}$ into $E^{\otimes|\lambda|}$. Then we mod out by some relations. These relations make the columns alternating, and between the columns we introduce the so-called exchange relations.

More formally: For a vector space $E^{\times k} = E \oplus \cdots \oplus E$ (k times) we define the linear map to $E \otimes E \otimes \cdots \otimes E$ via $(\ell_1, \ldots, \ell_k) \mapsto \ell_1 \otimes \cdots \otimes \ell_k$. We can compose this with an antisymmetrization map and obtain a linear map $E^{\times k} \to E^{\wedge k}$. We can tensor several of these maps to obtain the map $\psi : E^{\times\lambda} \to \bigotimes_{i=1}^{\lambda_1} \bigwedge^{\mu_i} E$. Now

$$E^{\lambda} := \big(\wedge^{\mu_1} E \otimes \cdots \otimes \wedge^{\mu_\ell} E \big) / Q^{\lambda}(E), \tag{17.1}$$

where $Q^{\lambda}(E)$ is the subrepresentation of $\bigotimes_{i=1}^{\lambda_1} \bigwedge^{\mu_i} E$ generated by all elements of the form $\psi(\vec{v}) - \sum \psi(\vec{w})$, where for some fixed exchange type t the sum is over all $\vec{w}$ obtained from $\vec{v}$ by an exchange of type t. E^{λ} is called a *Schur module*.

Suppose we have an ordered basis $\{e_1, \ldots, e_m\}$ of E. Then for any Young tableau of T of shape λ with elements in $\{1, \ldots, m\}$ we get an element of $E^{\times\lambda}$ by replacing every i in a box of T by the element e_i. We call this element $\hat{e}_T$. The image of this element in E^{λ} is denoted by e_T.

One can now easily verify that the map $\varphi : E^{\times\lambda} \to E^{\lambda}$ has the following three properties:

(1) φ is multilinear

(2) φ is alternating in the entries of any column of λ

(3) For any $\vec{v} \in E^{\times\lambda}$ and any exchange type t we have $\varphi(\vec{v}) = \sum \varphi(\vec{w})$, where the sum is over all $\vec{w}$ obtained from $\vec{v}$ by an exchange of the type t.

17.3 A more explicit quotient space

Lemma 17.1. *If $e_1, \ldots, e_m$ is a basis of E, then $E^{\lambda} \simeq F/Q$, where F is the vector space whose basis is the set $\hat{e}_T$ for all Young tableaux T of shape λ with entries from $\{1, \ldots, m\}$ and $Q \subseteq F$ is generated by the elements*

(i) $\hat{e}_T$ if T has two equal entries in a column,

(ii) $\hat{e}_T + \hat{e}_{T'}$ where T' is obtained from T by interchanging two entries in a column,

(iii) $\hat{e}_T - \sum_S \hat{e}_S$*, where for some exchange type* t *the sum is over all* S *obtained from* T *by an exchange of type* t*.*

Proof. For every Young tableau T of shape λ we get an element in $E^{\times\lambda}$ and these elements generate $E^{\times\lambda}$. Therefore their images e_T generate E^λ, i. e., the map $F \to E^\lambda$ is surjective. Properties (2) and (3) imply that the generators of Q map to zero, so $F/Q \twoheadrightarrow E^\lambda$ is surjective. We now routinely check that this is an isomorphism as follows. The vectors e_T for Young tableaux T give a basis of the tensor product $E^{\otimes\lambda}$. The vector space obtained by the relations (i) and (ii) is exactly the tensor product

$$\wedge^{\mu_1} E \otimes \cdots \otimes \wedge^{\mu_{\lambda_1}} E$$

(and the $\hat{e}_T$ with all columns strictly increasing forms a basis for this vector space). The relations (iii) then generate the vector space of relations $Q^\lambda(E)$, as follows from multilinearity and the fact that the e_i generate E. The lemma therefore follows from (17.1). □

The image of $\hat{e}_T$ under the map $F \to E^\lambda$ is denoted by e_T. For example,

$$e_{\begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 & 4 \\ \hline\end{array}} = e_{\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline\end{array}} + e_{\begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 & 4 \\ \hline\end{array}} = e_{\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 & 4 \\ \hline\end{array}} - e_{\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 4 \\ \hline\end{array}}$$

Since this notation is clumsy, we usually write T instead of e_T. For example, using this notation we have

$$\begin{array}{|c|c|c|c|}\hline 1 & 2 & 3 & 4 \\ \hline\end{array} = \begin{array}{|c|c|c|c|}\hline 3 & 1 & 4 & 2 \\ \hline\end{array} = \begin{array}{|c|c|c|c|}\hline 2 & 4 & 1 & 3 \\ \hline\end{array}$$

and

$$\begin{array}{|c|}\hline 1 \\ \hline 2 \\ \hline 3 \\ \hline 4 \\ \hline\end{array} = -\begin{array}{|c|}\hline 2 \\ \hline 1 \\ \hline 3 \\ \hline 4 \\ \hline\end{array} = \begin{array}{|c|}\hline 1 \\ \hline 3 \\ \hline 4 \\ \hline 2 \\ \hline\end{array}$$

17.4 Sylvester's lemma

A multilinear function $f : V^{\times d} \to \mathbb{C}$ is called *alternating* if

$$f(v_1, \ldots, v_i, v_{i+1}, \ldots, v_d) = -f(v_1, \ldots v_{i-1}, v_{i+1}, v_i, v_{i+2}, \ldots, v_d).$$

Lemma 17.2. *A multilinear function* $f : V^{\times d} \to \mathbb{C}$ *is alternating iff* $f(v_1, \ldots, v_d) = 0$ *whenever* $v_i = v_{i+1}$*.*

Proof. Clearly, if f is alternating, then

$$f(v_1, \ldots, v_i, v_i, v_{i+2}, \ldots, v_d) = -f(v_1, \ldots, v_i, v_i, v_{i+2}, \ldots, v_d)$$

and thus $f(v_1, \ldots, v_i, v_i, v_{i+2}, \ldots, v_d) = 0$ (because $\mathrm{char}(\mathbb{C}) \neq 2$).

For the other direction,

$$\begin{aligned}
&f(v_1,\ldots,v_i,v_{i+1},v_{i+2},\ldots,v_d)\\
&\quad= f(v_1,\ldots,v_i,v_i,v_{i+2},\ldots,v_d)+f(v_1,\ldots,v_i,v_{i+1}-v_i,v_{i+2},\ldots,v_d)\\
&\quad= f(v_1,\ldots,v_i,v_{i+1}-v_i,v_{i+2},\ldots,v_d)\\
&\quad= f(v_1,\ldots,v_i-v_{i+1},v_{i+1}-v_i,v_{i+2},\ldots,v_d)+f(v_1,\ldots,v_{i+1},v_{i+1}-v_i,v_{i+2},\ldots,v_d)\\
&\quad= -\big(f(v_1,\ldots,v_{i+1}-v_i,v_{i+1}-v_i,v_{i+2},\ldots,v_d)+f(v_1,\ldots,v_{i+1},v_i-v_{i+1},v_{i+2},\ldots,v_d)\big)\\
&\quad= -\big(f(v_1,\ldots,v_{i+1},v_i-v_{i+1},v_{i+2},\ldots,v_d)\big)\\
&\quad= -\big(f(v_1,\ldots,v_{i+1},v_i-v_{i+1},v_{i+2},\ldots,v_d)+f(v_1,\ldots,v_{i+1},v_{i+1},v_{i+2},\ldots,v_d)\big)\\
&\quad= -f(v_1,\ldots,v_{i+1},v_i,v_{i+2},\ldots,v_d).
\end{aligned}$$

□

Corollary 17.3. *For $V=\mathbb{C}^p$ the only alternating multilinear function $V^{p+1}\to\mathbb{C}$ is the zero function.*

Proof. If f is alternating, to calculate $f(v)$ we express $v\in V^{p+1}$ over the standard basis. By the pigeonhole principle at least one standard vector appears at least twice. By Lemma 17.2 $f(v)=0$. □

For the explicit construction of the irreducibles we need the following lemma, proved by Sylvester in 1851.

Lemma 17.4. *For any $p\times p$ matrices M and N, and $1\le k\le p$,*

$$\det(M)\cdot\det(N)=\sum\det(M')\cdot\det(N'),$$

where the sum is over all pairs (M',N') of matrices obtained from M and N by interchanging a fixed set of k columns of N with any k columns of M, preserving the ordering of the columns.

Proof. By the alternating property of determinants, w.l.o.g. the fixed set of columns of N are the first k columns. For vectors $v_1,\ldots,v_p\in\mathbb{C}^p$ we write $\det(v_1\cdots v_p)$ for the determinant of the matrix with these column vectors. We have to prove

$$\det(v_1\cdots v_p)\det(w_1\cdots w_p)=\sum_{i_1<\cdots<i_k}\det(v_1\cdots w_1\cdots w_k\cdots v_p)\det(v_{i_1}\cdots v_{i_k}w_{k+1}\cdots w_p),$$

where in the sum the vectors $w_1,\ldots,w_k$ are interchanged with the vectors $v_{i_1},\ldots,v_{i_k}$. It suffices to show that the difference of the two sides is an alternating function in the $p+1$ vectors $v_1,\ldots,v_p,w_1$, since any such function must vanish (see Cor. 17.3). For this it suffices to show (see Lemma 17.2) that the two sides are equal when two successive vectors v_i and v_{i+1} are equal (which is easy to see: The left hand side is zero and we can pair the nonzero summands on the right hand side such that each pair cancels out) and when $v_p=w_1$. In the latter case, fixing $v_p=w_1$, it suffices to show that the difference of the two sides is an alternating function of $v_1,\ldots,v_p,w_2$. Again, the case when $v_i=v_{i+1}$ is immediate. This time $v_p=w_2$ means $w_1=w_2$ and thus both sides vanish. □

Let $Z_{i,j}$ be variables, $1 \le i \le n$, $1 \le j \le m$. We write $\mathbb{C}[Z] := \mathbb{C}[Z_{1,1}, \dots, Z_{n,m}]$.

For each p-tuple $(i_1, \dots, i_p)$ of integers from $\{1, \dots, m\}$, with $p \le n$, we define the symbolic determinant

$$D_{i_1,i_2,\dots,i_p} := \det\begin{pmatrix} Z_{1,i_1} & \cdots & Z_{1,i_p} \\ \vdots & \ddots & \vdots \\ Z_{p,i_1} & \cdots & Z_{p,i_p} \end{pmatrix}.$$

For a Young tableau T we take the product of the column determinants:

$$D_T := \prod_{j=1}^{\lambda_1} D_{T(1,j),T(2,j),\dots,T(\mu_j,j)},$$

where $T(i,j)$ is the entry of T in the i-th row and j-th column of T.

Lemma 17.5. *There is a well-defined (canonical) homomorphism from E^λ to $\mathbb{C}[Z]$ that maps e_T to D_T for all Young tableaux T.*

Proof. Using Lemma 17.1, for well-definedness it suffices to show that the elements D_T satisfy the corresponding properties (i)-(iii) of Lemma 17.1. Properties (i) and (ii) follows from the alternating property of determinants. Property (iii) follows from Sylvester's lemma 17.4, applied to appropriate matrices. For this, suppose the two columns of T in which the exchange takes place have entries $i_1, \dots, i_p$ in the first and $j_1, \dots, j_q$ in the second. Set

$$M := \begin{pmatrix} Z_{1,i_1} & \cdots & Z_{1,i_p} \\ \vdots & \ddots & \vdots \\ Z_{p,i_1} & \cdots & Z_{p,i_p} \end{pmatrix} \qquad N := \begin{pmatrix} Z_{1,j_1} & \cdots & Z_{1,j_q} & 0 \\ \vdots & \ddots & \vdots & \mathrm{Id}_{p-q} \\ Z_{p,j_1} & \cdots & Z_{p,j_q} & \end{pmatrix}$$

Here the matrix N has a lower right identity matrix of size $p-q$, and an upper right $q \times (p-q)$ block of zeros. Note that $\det(N) = \det\begin{pmatrix} Z_{1,j_1} & \cdots & Z_{1,j_q} \\ \vdots & \ddots & \vdots \\ Z_{q,j_1} & \cdots & Z_{q,j_q} \end{pmatrix}$. Sylvester's lemma, applied to M and N and the fixed subset of columns in N being specified by the subset of the right column of T used in the exchange, translates precisely to the required equation. ◻

17.5 An explicit basis of the Schur module

Definition 17.6. A Young tableau is called *semistandard* if each row read from left to right is nondecreasing and each column read from top to bottom is strictly increasing.

Example of a semistandard Young tableau:

1	1	1	1	2	2	2	3
2	2	2	3	4			
3	3	4	4				

Theorem 17.7. *If $e_1, \dots, e_m$ is a basis of E, then e_T is a basis of E^λ, where T runs over all semistandard tableaux of λ with entries from $\{1, \dots, m\}$.*

Proof. On the set of Young tableaux we define an ordering:

- $T' > T$ if in the right-most column which is different, the lowest box where they differ has a larger entry in T'.

We first prove that the e_T generate E^λ. We also use $E^\lambda = F/Q$ from Lemma 17.1. We must show that, given any T that is not semistandard, we can write e_T as a linear combination of elements e_S with $S > T$ and elements in Q, because then we can use this process recursively to express every e_T as a linear combination of e_S with S semistandard and elements in Q.

We may assume that the entries in each column of T are strictly increasing by using relations (i) and (ii); Note that by making the the columns strictly increasing in T replaces T by T' with $T' > T$.

If the columns are strictly increasing, but T is not semistandard, then suppose the k-th entry of the j-th column is strictly larger than the k-th entry of the $(j+1)$-st column. Then we have a relation $e_T \equiv \sum_S e_S$, the sum over all S obtained from T by exchanging the top k entries of the $(j+1)$-st column with k entries in the j-th column (preserving their order). Since each such S has $S > T$, we proved that the e_T for semistandard T generate E^λ.

To prove that the e_T are linearly independent, we use Lemma 17.5, so it suffices to prove that the D_T are linearly independent as T varies over all semistandard tableaux T. For this we order the variables $Z_{i,j}$ in the order: $Z_{i,j} < Z_{i',j'}$ if $i < i'$ or both $i = i'$ and $j < j'$. We order the monomials in these variables lexicographically: $M_1 < M_2$ if the smallest $Z_{i,j}$ that occurs to a different power occurs to a smaller power in M_1 than in M_2. Note that if $M_1 < M_2$ and $N_1 < N_2$, then $M_1 N_1 < M_2 N_2$. It follows immediately from this definition that the smallest monomial that appears in a determinant $D_{i_1,\dots,i_p}$ if $i_1 < \cdots < i_p$ is the diagonal term $Z_{1,i_1} \cdots Z_{p,i_p}$. Therefore the smallest monomial occurring in D_T, if T has increasing columns, is $\prod (Z_{i,j})^{m_T(i,j)}$, where $m_T(i,j)$ is the number of times j occurs in the i-th row of T. This monomial occurs with coefficient 1.

Now order the semistandard tableaux by saying that $T < T'$ if the first row where they differ, and the first entry where they differ in that row, is smaller in T than in T'. Equivalently, $T < T'$ if for the smallest i for which there is a j with $m_T(i,j) \neq m_{T'}(i,j)$ and for the smallest such j we have $m_T(i,j) < m_{T'}(i,j)$. It follows that if $T < T'$, then the smallest monomial occurring in D_T is smaller than the smallest monomial occurring in $D_{T'}$ and thus smaller than *any* monomial occurring in $D_{T'}$. From this the linear independence follows: If $\sum \alpha_T D_T = 0$, take T minimal such that $\alpha_T \neq 0$, then the coefficient of $\prod (Z_{i,j})^{m_T(i,j)}$ in $\sum \alpha_T D_T$ is α_T. ◻

Remark 17.8. The proof of Theorem 17.7 provides an algorithm to express any e_T for a tableau T over the basis $(e_{T'})$ with T' semistandard. This algorithm is called the *straightening algorithm*. Together with the upcoming Lemma 17.9 this gives an algorithmic way of computing the action of GL_n in E^λ.

17.6 Highest weight vectors

Lemma 17.9. *Fix an arbitrary ordering on the set of boxes of a Young diagram λ. Let T be a Young tableau and let $1 \le j_1, \dots, j_n \le m$ be the entries in its boxes. We write $T = \lambda(j_1, \dots, j_n)$. If $g \in \mathsf{GL}_m$, then*

$$ge_T = \sum_{(i_1,\dots,i_n)\in\{1,\dots,m\}^n} g_{i_1,j_1} \cdots g_{i_n,j_n} e_{\lambda(i_1,\dots,i_n)}.$$

Proof. This is not only true for E^λ, but it is already true for $E^{\otimes|\lambda|}$. More precisely, for a Young tableau T consider the vector in $E^{\times\lambda}$ obtained from replacing each i in T by e_i. Let $\tilde{e}_T$ denote its image under the map $E^{\times\lambda} \to E^{\otimes|\lambda|}$. The action on $E^{\otimes|\lambda|}$ is given by

$$g(v_1 \otimes \cdots \otimes v_{|\lambda|}) = (gv_1) \otimes \cdots \otimes (gv_{|\lambda|}).$$

Thus by multilinearity of the tensor product we have

$$g\tilde{e}_T = \sum_{(i_1,\dots,i_n)\in[m]^n} g_{i_1,j_1} \cdots g_{i_n,j_n} \tilde{e}_{T(i_1,\dots,i_n)},$$

The claim follows by (17.1). □

Using part (a') and (a) of Theorem 11.15, the following lemma finishes the explicit construction of the polynomial irreducible representations of the general linear group and proves part (c) of Theorem 11.15, where the E_λ are the required irreducible representations.

Lemma 17.10. *Up to multiplication with a nonzero scalar, the only highest weight vector in E^λ is the vector e_T, where T is the semistandard tableau of shape λ whose i-th row contains only the integer i as entries.*

Proof. Let T be the semistandard tableau of shape λ whose i-th row contains only the integer i as entries. Let $g \in U_n$ be an upper triangular matrix with 1s on the main diagonal, i. e., $g_{i,j} = 0$ if $i > j$. From Lemma 17.9 it follows that the only nonzero $e_{T'}$ that can occur in ge_T is e_T itself and therefore e_T is a HWV.

Similary suppose that the p-th row of T is the first row that contains an element larger than p. Let $q > p$ be the smallest such misplaced element in row p. Define g to be the elementary matrix that has 1s on the main diagonal and an entry α in row p and column q. We see that $ge_T = \sum \alpha^{c_{T'}} e_{T'}$, where the sum is over all tableaux T' by exchanging some set (possibly empty) of the qs appearing in T to ps, and $c_{T'}$ is the number of such exchanges. Considering this as a univariate polynomial in α and looking at the linear coefficient, this is a sum over T' in which a single q is switched to a p. Some of these T' could have p appear in a column twice, but the other T' are pairwise distinct semistandard tableaux and there is at least one of them. Thus $ge_T \neq e_T$ and hence e_T is not a HWV. □

The explicit construction of the irreducible representations of the general linear group

This chapter finalizes our classification of the polynomial irreducible representations of GL_n: The irreducibles are indexed by partitions λ that have at most n parts.
We will use the explicit quotient space from Lemma 17.1 for several constructions in Chapters 19 and 20, in particular we will use it to construct the irreducibles of the symmetric group (the so-called Specht modules).

18 The algebraic Peter–Weyl theorem

In this chapter we prove the algebraic Peter–Weyl theorem. It will be used in Chapter 19 to prove the Schur–Weyl duality that is used to understand the highest weight vectors in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]$. Moreover, we will see that it can directly be used to find upper bounds on the multiplicities for example in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]$ by using so-called Kronecker coefficients. We will see how these coefficients can then be used in the multiplicity based approach outlined in Section 12.4.

18.1 Regular functions

Let $G = \mathsf{GL}_N$ and $v \in \mathbb{A}$, where $\mathbb{A}$ has a polynomial G-action. The coordinate ring $\mathbb{C}[\overline{Gv}]$ is a subring of the algebra $\mathbb{C}[Gv]$ that we define next. The ring $\mathbb{C}[Gv]$ will have a particulary nice representation theory.

Definition 18.1. A *regular function* f on a locally closed set $X \subseteq \mathbb{A}$ is a function $f\colon X \to \mathbb{C}$ defined on the whole of X as follows: There exist finitely many fractions of polynomials $\frac{f_i}{g_i}$ with $f_i, g_i \in \mathbb{C}[\mathbb{A}]$ such that

$$\forall x \in X\ \exists i : g_i(x) \neq 0$$

and

$$\forall x \in X\ \forall i : \text{either } g_i(x) = 0 \text{ or } \frac{f_i(x)}{g_i(x)} = f(x).$$

For a locally closed set X define $\mathbb{C}[X]$ to be the algebra of regular functions on X.

The following example is called the *glued double cusp* and was provided by Prof. Dr. Eike Lau.

Example 18.2. Let the closed set $Y \subseteq \mathbb{C}^5$ be cut out by the polynomials $T_1^3 - T_2^2$, $T_3^3 - T_4^2$, $T_1T_3 - T_5^2$, and $T_2T_4 - T_5^3$. Since $\{0\} \subseteq Y$ is closed, $X := Y \setminus \{0\}$ is a locally closed set. One can check that X is parameterized by two variables as follows:

$$X = \{(t_1, t_2, t_3, t_4, t_5) \mid \alpha, \beta \in \mathbb{C}, \alpha \neq 0 \text{ or } \beta \neq 0,$$
$$t_1 = \alpha^2, t_2 = \alpha^3, t_3 = \beta^2, t_4 = \beta^3, t_5 = \alpha\beta\}.$$

Now consider the following regular function defined by two fractions of polynomials:

$$f = \frac{t_5 t_1}{t_2} = \frac{t_4}{t_3},$$

whose value is just β in the above syntax. Although f is defined on the whole X, the two fractions of polynomials are not, because their denominators both have zeros in X. In fact, one can show that f cannot be written as a single fraction of polynomials.

If Gv is a cone, then $\mathbb{C}[Gv]$ is graded with the same argument as at the end of Chapter 4.

Remark 18.3. If X in Definition 18.1 is Zariski-closed, then $\mathbb{C}[X]$ coincides with our earlier definition of the coordinate ring, i. e., $\mathbb{C}[X] = \mathbb{C}[\mathbb{A}]/I(X)$.

Our main interest in $\mathbb{C}[X]$ stems from the map

$$\iota : \mathbb{C}[\overline{Gv}]_\delta \hookrightarrow \mathbb{C}[Gv]_\delta$$

that is the restriction of functions. Clearly ι is a linear map. We show that ι is injective: If $\iota(f) = 0$, then $f(w) = 0$ for all $w \in Gv$. Since f is continuous, $f(w) = 0$ for all $w \in \overline{Gv}$. □

With Corollary 12.5 we obtain

$$\mathrm{mult}_\lambda(\mathbb{C}[\overline{Gv}]_\delta) \leq \mathrm{mult}_\lambda(\mathbb{C}[Gv]_\delta). \tag{18.1}$$

18.2 Invariants under the stabilizer

Definition 18.4. Let a group G act on a set S. For $v \in S$ define the *stabilizer of v under the action of G* as

$$\mathrm{stab}_G(v) := \{g \in G \mid gv = v\}.$$

Definition 18.5. Let H be a group. For an H-representation V define the *set of H-invariants*

$$V^H := \{v \in V \mid \forall g \in H : gv = v\}.$$

For a group G the *group algebra* $\mathbb{C}[G]$ is defined as the set of finite formal sums of group elements from G. If G is finite, then $\mathbb{C}[G] \simeq \mathbb{C}^{|G|}$ as a vector space. The group $G \times G$ acts on $\mathbb{C}[G]$ via

$$((g_1, g_2)f)(g) = f(g_1^{-1} g g_2).$$

Theorem 18.6. *Let G be a finite group, let V be a G-representation, and let $v \in V$. Let $\mathbb{C}[G]^{\mathrm{stab}_G(v)}$ denote the set of right $\mathrm{stab}_G(v)$-invariants, i. e., the elements fixed under the action of $\{1\} \times \mathrm{stab}_G(v)$. Note that $\mathbb{C}[G]^{\mathrm{stab}_G(v)}$ is a representation of $G \times \{1\} \simeq G$. Then the map*

$$\varphi : \mathbb{C}[Gv] \to \mathbb{C}[G]^{\mathrm{stab}_G(v)},\ f \mapsto \big(g \mapsto f(gv)\big)$$

is an isomorphism of G-representations.

Proof. The finiteness of G implies that $\mathbb{C}[Gv]$ is the vector space of all functions on Gv and $\mathbb{C}[G]$ is the vector space of all functions on G, without any additional constraints on the functions.

First of all, we verify that φ is well-defined, i. e., that $\kappa : g \mapsto f(gv)$ is invariant under $\mathrm{stab}_G(v)$. Let $g_2 \in \mathrm{stab}_G(v)$. Then $(g_2\kappa)(g) = \kappa(gg_2) = f(gg_2v) = f(gv) = \kappa(g)$, thus κ is fixed under g_2.

The map φ is clearly linear.

The inverse map φ^{-1} maps the $\mathrm{stab}_G(v)$-invariant ψ to $\big(gv \mapsto \psi(g)\big) \in \mathbb{C}[Gv]$. But $\psi(g)$ depends on g and not just gv, so we have to verify that this is well-defined: If $g'v = gv$, then we have to show that $\psi(g)$ coincides with $\psi(g')$. We have $v = g'^{-1}gv$, thus $g'^{-1}g \in \mathrm{stab}_G(v)$. Since ψ is $\mathrm{stab}_G(v)$-invariant, $\psi = (\mathrm{id}, g'^{-1}g)\psi$ and therefore $(\mathrm{id}, g')\psi = (\mathrm{id}, g)\psi$. In particular, if we evaluate both sides at the identity we obtain $\psi(g') = \psi(g)$.

It is easy to verify that both maps are inverses of each other.

For the G-equivariance we have to show that $g(\varphi(f)) = \varphi(gf)$.

We have $\varphi(f) = (g' \mapsto f(g'v))$ and hence $g(\varphi(f)) = (g' \mapsto f((g^{-1}g')v))$. Therefore $\varphi(gf) = (g' \mapsto (gf)(g'v)) = (g' \mapsto f(g^{-1}g'v)) = g(\varphi(f))$. ☐

Without giving the proof we state that Theorem 18.6 also holds for large classes of groups, including GL_N. This is particularly interesting because the representation theoretic structure of $\mathbb{C}[G]_\delta^{\mathrm{stab}_G(v)}$ can be obtained in Theorem 18.9 below.

18.3 Algebraic Peter–Weyl theorem

Proposition 18.7. *Let G and H be groups and let V be a G-representation and W an H-representation. Then $V \otimes W$ is an irreducible $G \times H$-representation iff both V and W are irreducible. Moreover, every irreducible $G \times H$-representation is isomorphic to some $V \otimes W$ for V, W irreducible.*

We omit the proof, but refer the reader to [50].

Definition 18.8. If V is a G-representation, then the dual space V^* is also a G-representation via $(g\ell)(v) := \ell(g^{-1}v)$ for all $\ell \in V^*$, $v \in V$. This representation is called the *dual representation* of V.

Clearly $\dim V = \dim V^*$. Moreover, V is irreducible iff V^* is irreducible.

If V is a polynomial representation, then V^* is not necessarily also a polynomial representation. Indeed, for GL_N we have that both V and V^* are polynomial representations iff the type of V is $(0, 0, \dots, 0)$.

Theorem 18.9 (Algebraic Peter–Weyl theorem for finite groups)**.** *Let G be a finite group. On $\mathbb{C}[G]$ we have an action of $G \times G$ via $((g, g')f)(\tilde{g}) := f(g^{-1}\tilde{g}g')$ and we have*

$$\mathbb{C}[G] = \bigoplus_\lambda \{\lambda\}^* \otimes \{\lambda\},$$

where λ runs over all isomorphism types of irreducible G-representations and $\{\lambda\}$ denotes an irreducible G-representation of type λ.

Proof. We follow [54, Thm. 8.6.4.3].

Let V be a G-representation. Define the $G \times G$-equivariant linear map

$$i_V : V^* \otimes V \longrightarrow \mathbb{C}[G], \;\; i_V(\ell \otimes v)(g) = \ell(gv),$$

defined via linear continuation on all tensors.

We first show that if V is irreducible, then i_V is injective: $\ker(i_V)$ is a subrepresentation of $V^* \otimes V$. Since $i_V \neq 0$, $\ker(i_V) \neq V^* \otimes V$. $V^* \otimes V$ is an irreducible $G \times G$-representation (Proposition 18.7). Thus $\ker(i_V) = 0$. Hence i_V is injective.

This already proves that the right-hand side is contained in the left-hand side.

To finish the proof we show that $i_V(V^* \otimes V)$ equals the isotypic component of type V^* w.r.t. the action of $G \times \{1\}$.

Let W^* be an irreducible $G \times \{1\}$-representation that is isomorphic to V^*. Let $j : W^* \to \mathbb{C}[G]$ be a $G \times \{1\}$-morphism. We need to show that $j(W^*) \subseteq i_V(V^* \otimes V)$.

We identify W^* and V^*. Let $\ell \in V^*$ be arbitrary. Define $v \in V$ via $\ell(v) := j(\ell)(1_G)$. Note that $(g^{-1}\ell)(v) = (g^{-1}(j(\ell)))(1_G) = j(g^{-1}\ell)(1_G)$. Then $j(\ell) = i_V(\ell \otimes v)$ because:

$$(j(\ell))(g) = (j(\ell))(g \cdot 1_G) = (g^{-1} \cdot j(\ell))(1_G) = (j(g^{-1} \cdot \ell))(1_G) = (g^{-1}\ell)(v) = \ell(gv) = (i_V(\ell \otimes v))(g).$$

□

The group algebra $\mathbb{C}[G]$ for finite groups G is isomorphic (as a $G \times G$-representation) to the coordinate ring of G (where G is interpreted as finite subset of matrices in the set of all invertible matrices of some fixed size). The key fact is that with the coordinate ring interpretation Theorem 18.9 generalizes to all reductive algebraic groups, in particular for $G = \mathsf{GL}_N$. There the sum is over all types λ of *rational* representations, which includes all polynomial representations. In particular, using Theorem 18.6, the following corollary holds.

Corollary 18.10.

$$\mathrm{mult}_\lambda(\mathbb{C}[Gv]) = \dim(\{\lambda^*\}^{\mathrm{stab}_G(v)}).$$

Proof.

$$\begin{aligned}\mathrm{mult}_\lambda(\mathbb{C}[Gv]) &\overset{\text{Thm. 18.6}}{=} \mathrm{mult}_\lambda(\mathbb{C}[G]^{\mathrm{stab}_G(v)}) \overset{\text{Thm. 18.9}}{=} \mathrm{mult}_\lambda((\bigoplus_\mu \{\mu\} \otimes \{\mu\}^*)^{\mathrm{stab}_G(v)}) \\ &= \dim(\{\lambda^*\}^{\mathrm{stab}_G(v)}),\end{aligned}$$

where the last equality holds because $\mathrm{mult}_\lambda(\{\mu\}) = 1$ iff $\lambda = \mu$ and 0 otherwise. □

Several variants of Theorem 18.9 are also true, with minor changes in the proof: On $\mathbb{C}^{N \times N}$ we have the action of $\mathsf{GL}_N \times \mathsf{GL}_N$ via $(g_1, g_2)M = g_1 M g_2^t$. The algebraic Peter–Weyl theorem implies the decomposition of the coordinate ring of the matrix space

$$\mathbb{C}[\mathbb{C}^{N \times N}]_d = \bigoplus_{\lambda \vdash_N d} \{\lambda\} \otimes \{\lambda\}. \tag{18.2}$$

This generalizes to a $\mathsf{GL}_a \times \mathsf{GL}_b$ action on $\mathbb{C}^{a\times b}$:

$$\mathbb{C}[\mathbb{C}^{a\times b}]_d = \bigoplus_{\lambda \vdash_{\min(a,b)} d} \{\lambda\}_a \otimes \{\lambda\}_b, \tag{18.3}$$

where $\{\lambda\}_a$ denotes the irreducible GL_a-representation of type λ.

18.4 The determinant and rectangular Kronecker coefficients

Combining (18.1) and Corollary 18.10 we see that

$$\mathrm{mult}_\lambda(\mathbb{C}[\overline{Gv}]) \overset{(18.1)}{\leq} \mathrm{mult}_\lambda(\mathbb{C}[Gv]) \overset{\text{Cor. 18.10}}{=} \dim(\{\lambda^*\}^{\mathrm{stab}_G(v)}), \tag{18.4}$$

which could potentially be used to find multiplicity or even occurrence obstructions. If $G = \mathsf{GL}_{n^2}$ and $v = \det_n$, then $\mathrm{stab}_G(v)$ was determined by Frobenius in 1897 [35]:

$$H := \mathrm{stab}_{\mathsf{GL}_{n^2}}(\det_n) = \big((\mathsf{GL}_n \times \mathsf{GL}_n)/\mathbb{C}^*\big) \rtimes \mathbb{Z}_2.$$

The multiplicities $\dim\{\lambda\}^H$ are known as *rectangular symmetric Kronecker coefficients* $sk(\lambda, n\times d)$. We have seen in (18.4) that if $a_\lambda(\delta, d) > 0$ and $sk(\lambda, n \times d) = 0$, then the type λ occurs in the vanishing ideal $I(\overline{\mathsf{GL}_{n^2}\det_n})$. We will see a working application of this approach in Chapter 22.

An upper bound for $sk(\lambda, n \times d)$ is given by a similar coefficient: Given a partition λ of m with at most n^2 rows, we interpret it as a $\mathsf{GL}_n \times \mathsf{GL}_n$-representation via the map $(g, g') \mapsto g \otimes g'$, where the matrix $g \otimes g'$ is the Kronecker product of matrices. Then $\{\lambda\}$ decomposes into irreducible $\mathsf{GL}_n \times \mathsf{GL}_n$-representations and the Kronecker coefficient $k(\lambda, \mu, \nu)$ is defined as the multiplicity of the irreducible $\mathsf{GL}_n \times \mathsf{GL}_n$-representation $\{\mu\} \otimes \{\nu\}$ in $\{\lambda\}$. Mulmuley and Sohoni conjectured that the vanishing of the Kronecker coefficients should give enough elements in the vanishing ideal to separate $\mathrm{VP}_s \subsetneq \mathrm{VNP}$, but that was recently disproved [47] (and strengthened in [23]).

Coordinate rings of orbits

The coordinate rings of orbit closures are not well understood.
The coordinate rings of orbits are much better understood: Their multiplicities are dimensions of stabilizer-invariant subspaces.
Moreover, we will use the Algebraic Peter–Weyl theorem to prove the Schur–Weyl duality in Chapter 19.

19 Explicit HWV constructions via Schur–Weyl duality

In this chapter we give an explicit interpretation of $\mathbb{C}[\mathbb{A}]_d$, $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_m]_n$, in terms of tensors. We understand the action of GL_m using the famous Schur–Weyl duality. We derive useful results on plethysm coefficients and potential candidates for partitions λ to separate the determinant from the padded permanent, and related orbit closure questions. In particular we will prove a result by Kadish and Landsberg [48] that the first row of λ must be large and we will see that the degree d must be superpolynomially large.

As in Proposition 12.7, $\mathsf{HWV}_\lambda(W)$ denotes the vector space of highest weight vectors of weight λ in the GL_N-representation W.

19.1 Specht modules

In this section we describe without proof the irreducible representations of the symmetric group. The irreducible representations of $\mathfrak{S}_n$ are called *Specht modules*. The Specht modules of $\mathfrak{S}_n$ are indexed by partitions of n, i. e., partitions λ with $|\lambda| = n$. Note the difference to GL_n, where the irreducible polynomial representations are indexed by partitions λ with $\ell(\lambda) \le n$, but where $|\lambda|$ is arbitrary.

The construction of the Specht modules works as follows. For $n \in \mathbb{N}$ let $\{\lambda\}$ denote the irreducible GL_n-representation of type λ. We assume $|\lambda| = n$. Let $\{\lambda\}^0$ denote the weight space of $\{\lambda\}$ of weight $(1, 1, \dots, 1) \in \mathbb{N}^n$. We embed $\mathfrak{S}_n \subseteq \mathsf{GL}_n$ via permutation matrices as in Lemma 11.10. Lemma 11.10 says that $\mathfrak{S}_n$ acts on $\{\lambda\}^0$. It turns out (without proof) that

- $\{\lambda\}^0$ is irreducible as an $\mathfrak{S}_n$-representation,
- if $\lambda \ne \mu$, then $\{\lambda\}^0$ and $\{\mu\}^0$ are non-isomorphic $\mathfrak{S}_n$-representations,
- all irreducible representations of $\mathfrak{S}_n$ are obtained as $\{\lambda\}^0$ for some λ with $|\lambda| = n$.

We denote by $[\lambda] := \{\lambda\}^0$ the Specht module of type λ.

Example 19.1. 1. X_1^n is a HWV in $V = \mathbb{C}[X_1, \dots, X_M]_n$ of weight $\lambda = (n, 0, \dots, 0)$. For $n = M$ we have a unique line of weight $(1, 1, \dots, 1) = (1^n)$, which is $\mathbb{C} \cdot X_1 X_2 \cdots X_n$. Thus $\{(n)\}^0 = [(n)]$ is 1-dimensional. The action of $\mathfrak{S}_n$ permutes the positions of the variables in the monomial $X_1 X_2 \cdots X_n$, so the action of $\mathfrak{S}_n$ is trivial: $[(n)]$ is the trivial representation.

2. Take $V = \mathbb{C}$. GL_n acts on V via $gv := \det(g)v$. The weight is $(1, 1, \dots, 1) = (1^n)$. Thus $[(1^n)]$ is 1-dimensional. Moreover, if $\varrho(\pi) \in \mathsf{GL}_n$ denotes the permutation matrix of the permutation $\pi \in \mathfrak{S}_n$, then $\det(\varrho(\pi)) = \mathrm{sign}(\pi)$, so $[(1^n)]$ is the *alternating representation:* $\pi v = \mathrm{sign}(\pi)v$.

A semistandard Young tableau with n boxes in which each number $1, 2, \dots, n$ appears exactly once is called a *standard Young tableau*. A basis of $[\lambda]$ is given by e_T, where T goes over the standard Young tableaux of shape λ, see Section 17.3. For example, $[(2, 1)]$ is 2-dimensional and we have (when writing T instead of e_T)

$$(2\ 3)\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} = \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}$$

and

$$(1\ 2)\begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} = \begin{array}{|c|c|}\hline 2 & 1 \\ \hline 3 \\ \cline{1-1}\end{array} = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} + \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 1 \\ \cline{1-1}\end{array} = \begin{array}{|c|c|}\hline 1 & 2 \\ \hline 3 \\ \cline{1-1}\end{array} - \begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 \\ \cline{1-1}\end{array}.$$

19.2 Explicit Schur–Weyl duality

We write $\lambda \vdash_N d$ to denote that λ is a partition of d into at most N parts.

The group $\mathfrak{S}_d \times \mathsf{GL}_N$ acts on $\otimes^d \mathbb{C}^N$ via

$$(\pi, g)(v_1 \otimes \cdots \otimes v_N) := (gv_{\pi^{-1}(1)}) \otimes \cdots \otimes (gv_{\pi^{-1}(N)}).$$

Its decomposition into irreducibles is known as the Schur–Weyl duality.

Theorem 19.2 (Schur–Weyl duality)**.**

$$\bigotimes^d \mathbb{C}^N \cong \bigoplus_{\lambda \vdash_N d} \{\lambda\} \otimes [\lambda].$$

Proof sketch. We follow [49].

In (18.3), let $d = a$ and take the GL_d-weight $(1, 1, \dots, 1)$ space:

$$\mathbb{C}[\mathbb{C}^{d\times b}]^0_d = \bigoplus_{\lambda \vdash_{\min(d,b)} d} [\lambda] \otimes \{\lambda\}_b.$$

A partition with d boxes cannot have more than d rows, so $\ell(\lambda) \le d$ is a void restriction. We obtain

$$\mathbb{C}[\mathbb{C}^{d\times b}]^0_d = \bigoplus_{\lambda \vdash_b d} [\lambda] \otimes \{\lambda\}_b.$$

Degree d polynomials on $\mathbb{C}^{d\times b}$ that have GL_d-weight $(1, 1, \dots, 1)$ are linear combinations of the monomials $x_{1,j_1} x_{2,j_2} \cdots x_{d,j_d}$, $1 \le j_i \le b$. These are in bijection to the rank 1 tensors $x_{j_1} \otimes x_{j_2} \otimes \cdots \otimes x_{j_d}$, $1 \le j_i \le b$. This gives a canonical isomorphism

$$\mathbb{C}[\mathbb{C}^{d\times b}]^0_d \cong \bigotimes^d \mathbb{C}^b$$

and the result follows. $\square$

The highest weight vectors in Theorem 19.2 can be described explicitly as follows.

We denote by $X_1, \dots, X_N$ the standard basis vectors of $\mathbb{C}^N$. Let $\lambda \vdash D$ and μ denote the transpose of λ, so μ_i denotes the number of boxes in the i-th column of λ. For $j \le N$ we note that $v_{j\times 1} := X_1 \wedge X_2 \wedge \cdots \wedge X_j$ is a highest weight vector of weight $j \times 1$: If we use the definition of $x_{ii'}(\alpha)$ from the proof of Lemma 11.13, then for $i < i'$ we have

$$x_{ii'}(1) X_1 \wedge X_2 \wedge \cdots \wedge X_j = X_1 \wedge X_2 \wedge \cdots \wedge X_j + \underbrace{X_1 \wedge \cdots \wedge X_{i'-1} \wedge X_i \wedge X_{i'+1} \wedge \cdots \wedge X_j}_{=0}$$

We define now:

$$v_\lambda := v_{\mu_1 \times 1} \otimes \ldots \otimes v_{\mu_{\lambda_1} \times 1} \in \bigotimes^D V. \tag{19.1}$$

It is easy to check that v_λ is a nonzero highest weight vector of weight λ.

Proposition 19.3. *Let $\lambda \vdash_{\dim V} D$. Then the vector space $\mathsf{HWV}_\lambda(\bigotimes^D V)$ is spanned by the $\mathfrak{S}_D$-orbit of v_λ.*

Proof. Schur–Weyl duality provides a $\mathsf{GL}(V) \times \mathfrak{S}_D$-isomorphism

$$\bigotimes^D V \simeq \bigoplus_{\lambda \vdash_{\dim V} D} \{\lambda\} \otimes [\lambda].$$

Recalling that $\mathsf{HWV}_\lambda(\{\lambda\})$ is one-dimensional, we see that $\mathsf{HWV}_\lambda(\bigotimes^D V)$ is isomorphic to $[\lambda]$ as an $\mathfrak{S}_D$-module. From the irreducibility of $[\lambda]$ it follows that $\mathsf{HWV}_\lambda(\bigotimes^D V)$ is spanned by the $\mathfrak{S}_D$-orbit of any of its nonzero elements. $\square$

Note that Prop. 19.3 is even more explicit:

$$\text{The isomorphism } \bigotimes^D V \simeq \bigoplus_{\lambda \vdash_{\dim V} D} \{\lambda\} \otimes [\lambda] \text{ maps } v_\lambda \text{ to } T \otimes S, \tag{19.2}$$

where T is the semistandard tableau with only letters i in row i and S is the so-called *superstandard* tableau: S contains the entries $1, \ldots, |\lambda|$ ordered columnwise from left to right, top to bottom. For example, $v_{(3,2,1)}$ corresponds to

$$\begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 2 & 2 \\ \cline{1-2} 3 \\ \cline{1-1}\end{array} \otimes \begin{array}{|c|c|c|}\hline 1 & 4 & 6 \\ \hline 2 & 5 \\ \cline{1-2} 3 \\ \cline{1-1}\end{array}.$$

This provides a new basis for $\bigotimes^D V$ given by pairs (T, S) of semistandard (T) and standard (S) tableaux. This is a special case of the so-called *Robinson-Schensted-Knuth-correspondence*.

19.3 Polynomials as symmetric tensors: Plethysms

We follow [23].

In this section we establish the fundamental connection between tensors and the coordinate rings $\mathbb{C}[\mathbb{A}]_d$, $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_m]_n$. This leads to several results concerning the possible partitions λ that can be used to separate orbit closures.

19.3.1 Wreath products and symmetric powers of symmetric powers

We have seen in Section 16.2 that the dth symmetric power $Sym^d W$ of a vector space W can be defined as the $\mathfrak{S}_d$-invariant subspace of $\bigotimes^d W$. This construction is easily seen to work for arbitrary GL_N-representations W. In fact, if $V = \mathbb{C}^N$ we can choose $W = Sym^n V$ and define $Sym^d Sym^n V := Sym^d(Sym^n V)$ as the space of $\mathfrak{S}_d$-invariants in $\bigotimes^d(Sym^n V)$. This is a subrepresentation of $\bigotimes^d(\bigotimes^n V)$ in a natural way that we want to understand now.

We partition the position set $[dn] := \{1, \ldots, dn\}$ into the *blocks* $B_1, \ldots, B_d$, where $B_u := \{(u-1)n + v \mid 1 \le v \le n\}$. The subgroup of $\mathfrak{S}_{dn}$ of permutations that preserve the partition into blocks is called the *wreath product* $\mathfrak{S}_n \wr \mathfrak{S}_d$. It is generated by the permutations leaving the blocks invariant, and the permutations of the form $(u-1)n+v \mapsto (\tau(u)-1)n+v$ with $\tau \in \mathfrak{S}_d$, which simultaneously permute the blocks. Structurally, the wreath product is a semidirect product $\mathfrak{S}_n \wr \mathfrak{S}_d \simeq (\mathfrak{S}_n)^d \rtimes \mathfrak{S}_d$. Note that its order equals $d!\, n!^d$. Symmetrizing over $\mathfrak{S}_n \wr \mathfrak{S}_d$, we obtain the projection

$$\Sigma_{d,n}(w) := \frac{1}{d!\, n!^d} \sum_{\sigma \in \mathfrak{S}_n \wr \mathfrak{S}_d} \sigma(w) \tag{19.3}$$

onto the $\mathfrak{S}_n \wr \mathfrak{S}_d$-invariant subspace $(\bigotimes^{dn} V)^{\mathfrak{S}_n \wr \mathfrak{S}_d} \subseteq \bigotimes^{dn} V$.

It is crucial and readily verified that

$$(\bigotimes^{dn} V)^{\mathfrak{S}_n \wr \mathfrak{S}_d} = Sym^d Sym^n V. \tag{19.4}$$

Example 19.4. We give an example that will naturally lead to the connection with polynomials. First, it is easy to verify that in $[(2,2)]$ (recall Section 17.3 and Section 19.1 for the relations on tableaux)

$$\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array} + \begin{array}{|c|c|}\hline 2&1\\\hline 3&4\\\hline\end{array}$$

is invariant under $\mathfrak{S}_2 \wr \mathfrak{S}_2$. We write $a \wedge b := \frac{1}{2}(a \otimes b - b \otimes a)$ and $a \odot b := \frac{1}{2}(a \otimes b + b \otimes a)$. In particular $a \odot a = a \otimes a$. We write $x := e_1$ and $y := e_2$ and omit the tensor symbol between them (caveat: the omitted symbol is *not* the symmetric product, but the tensor product). We convert tableaux to tensors using (19.2), so in particular $\begin{array}{|c|}\hline 1\\\hline 2\\\hline\end{array} = x \wedge y = \frac{1}{2}(xy - yx)$ and $\begin{array}{|c|c|}\hline 1&3\\\hline 2&4\\\hline\end{array} = (x \wedge y) \otimes (x \wedge y) = \frac{1}{4}(xyxy - xyyx - yxxy + yxyx)$.

$$\begin{aligned}
\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array} + \begin{array}{|c|c|}\hline 2&1\\\hline 3&4\\\hline\end{array} &= \begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array} + \begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array} + \begin{array}{|c|c|}\hline 2&3\\\hline 1&4\\\hline\end{array} = 2\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array} - \begin{array}{|c|c|}\hline 1&3\\\hline 2&4\\\hline\end{array} \\
&= \tfrac{1}{2}(xxyy - yxxy - xyyx + yyxx) - \tfrac{1}{4}(xyxy - yxxy - xyyx + yxyx) \\
&= \tfrac{1}{2}xxyy - \tfrac{1}{4}yxxy - \tfrac{1}{4}xyyx + \tfrac{1}{2}yyxx - \tfrac{1}{4}xyxy - \tfrac{1}{4}yxyx \\
&= \tfrac{1}{2}(xxyy + yyxx) - \tfrac{1}{4}(xyxy + yxxy + xyyx + yxyx) \\
&= \tfrac{1}{2}(xxyy + yyxx) - \tfrac{1}{4}(xy + yx)^{\otimes 2} \\
&= \tfrac{1}{2}(2(x^{\odot 2}) \odot (y^{\odot 2})) - \tfrac{1}{4}(2x \odot y)^{\odot 2} \\
&= ((x^{\odot 2}) \odot (y^{\odot 2})) - (x \odot y)^{\odot 2} \\
&= ((x^{\odot 2}) \odot (y^{\odot 2})) - (x \odot y)^{\odot 2},
\end{aligned}$$

which reminds us of the discriminant function, but without the coefficients.

Why did we choose $\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array} + \begin{array}{|c|c|}\hline 2&1\\\hline 3&4\\\hline\end{array}$ in the previous example? We will see this in the Remark 19.11.

19.3.2 Polynomials as symmetric tensors

Let $W = \mathbb{C}^N$ with standard basis $X_1, \ldots, X_N$. For a list $I = (i_1, \ldots, i_d) \in \{1, \ldots, N\}^d$ we define

$$X_I := X_{i_1} \otimes X_{i_2} \otimes \cdots \otimes X_{i_d} \in \bigotimes^d W$$

Let $C_N(d)$ denote the set of all $\alpha \in \mathbb{N}^N$ with $|\alpha| = d$. These α are called *compositions* of d. For $I \in \{1, \ldots, N\}^d$ we define the *type* $\zeta(I) \in C_N(d)$, letting $\zeta(I)_i$ denote the number of appearances of i in I. For example $\zeta(1,2,5,3,4,2,1,3,2) = (2,3,2,1,1)$. We associate with $\alpha \in C_N(d)$ the *monomial*

$$X^\alpha := \frac{1}{\binom{d}{\alpha}} \sum_{\zeta(I)=\alpha} X_I, \tag{19.5}$$

where the sum is over all $\alpha \in C_N(d)$ such that $\zeta(I) = \alpha$ and $\binom{m}{\alpha}$ is the multinomial coefficient

$$\binom{d}{\alpha_1\ \alpha_2\ \cdots\ \alpha_N} := \frac{d!}{\alpha_1!\alpha_2!\cdots\alpha_N!}.$$

This agrees with our definition:

$$X^\alpha = X_{\alpha_1} \odot \cdots \odot X_{\alpha_d} = \frac{1}{d!} \sum_{\pi \in \mathfrak{S}_d} \pi(X_{\alpha_1} \otimes \cdots \otimes X_{\alpha_d})$$

Note that X^α is a symmetric tensor: $X^\alpha \in Sym^{|\alpha|}$.

Given a homogeneous degree d polynomial f we can interpret the evaluation at a point p via

$$f(p) = \langle f, p^{\otimes d} \rangle, \tag{19.6}$$

where $\langle , \rangle$ is the inner product that is inherited from the standard inner product on $\mathbb{C}^N$: Eq. (19.6) is easily checked for monomials f, and since evaluation is a linear function, eq. (19.6) holds for all homogeneous polynomials f. The inner product of two tensors is called a *tensor contraction*.

We formalize this correspondence as follows.

Proposition 19.5. *For a GL_N-representation W there is a GL_N-isomorphism $\mathbb{C}[W]_d \simeq Sym^d W$ given by mapping monomials as in (19.5), which implies that evaluating polynomials is given by tensor contraction as in (19.6).*

Iterating Proposition 19.5 twice we obtain the following corollary.

Corollary 19.6. *There is a natural GL_N-isomorphism*

$$\mathbb{C}[\mathbb{A}]_d \simeq Sym^d Sym^n V\,,$$

where $\mathbb{A} = \mathbb{C}[X_1, \ldots, X_{\dim V}]_n$.

Example 19.7. The tensor that corresponds (up to scale) to the discriminant via Corollary 19.6 is the tensor from Example 19.4:

$$2xxyy + 2yyxx - yxxy - xyyx - xyxy - yxyx.$$

We evaluate the discriminant at $aX^2 + bXY + cY^2$, which is a contraction with the tensor $(ax^2 + bx \odot y + cy^2)^{\otimes 2}$, according to (19.6). In order to do so, we expand first:

$$\begin{aligned}
(ax^2 + bx \odot y + cy^2)^{\otimes 2} &= (axx + \tfrac{b}{2}xy + \tfrac{b}{2}yx + cyy)^2 \\
&= a^2xxxx + \tfrac{ab}{2}xxxy + \tfrac{ab}{2}xxyx + acxxyy \\
&\quad + \tfrac{ab}{2}xyxx + \tfrac{b^2}{4}xyxy + \tfrac{b^2}{4}xyyx + \tfrac{bc}{2}xyyy \\
&\quad + \tfrac{ab}{2}yxxx + \tfrac{b^2}{4}yxxy + \tfrac{b^2}{4}yxyx + \tfrac{b}{2}cyxxx \\
&\quad + acyyxx + \tfrac{bc}{2}yyxy + \tfrac{bc}{2}yyyx + c^2yyyy.
\end{aligned}$$

The contraction is color-coded and yields $4ac - b^2$, which is the evaluation of (the negative of) the discriminant at $aX^2 + bXY + cY^2$.

We study (19.6) in more detail in Section 20.1 in the situation of Corollary 19.6.

19.3.3 Plethysm coefficients

In Section 12.4.1 we defined $a_\lambda(d,n)$ as the multiplicity of λ in $\mathbb{C}[\mathbb{A}]_d$ with $\mathbb{A} = \mathbb{C}[X_1,\dots,X_N]_n$. By Proposition 12.7 we know that

$$a_\lambda(d,n) = \dim \mathsf{HWV}_\lambda(\mathbb{C}[\mathbb{A}]_d).$$

With Corollary 19.6 we obtain

$$a_\lambda(d,n) = \dim \mathsf{HWV}_\lambda(Sym^d Sym^n V).$$

Proposition 19.8. *Formally $a_\lambda(d,n)$ depends on $m := \dim V$, i. e., we would need a symbol $a_{\lambda,m}(d,n)$. For $\ell(\lambda) > \dim V$ we have $a_{\lambda,\dim V}(d,n) := 0$, as there exists no $\mathsf{GL}_{\dim V}$-representation of type λ. If $\ell(\lambda) \le m_1$ and $\ell(\lambda) \le m_2$, then $a_{\lambda,m_1}(d,n) = a_{\lambda,m_2}(d,n)$.*

Proposition 19.8 justifies the notation $a_\lambda(d,n)$. This was claimed in Section 12.4.1.

Proof of Proposition 19.8. Note that by Schur–Weyl duality and (19.4) we have

$$\begin{aligned}
a_\lambda(d,n) &= \dim \mathsf{HWV}_\lambda(Sym^d Sym^n V) \\
&\overset{(19.4)}{=} \dim \mathsf{HWV}_\lambda((\otimes^{nd} V)^{\mathfrak{S}_n \wr \mathfrak{S}_d}) \\
&= \dim \mathsf{HWV}_\lambda((\bigoplus_{\mu \vdash_{\dim V} nd} \{\mu\} \otimes [\mu])^{\mathfrak{S}_n \wr \mathfrak{S}_d}) \\
&= \dim(\bigoplus_{\mu \vdash_{\dim V} nd} \mathsf{HWV}_\lambda(\{\mu\}) \otimes ([\mu])^{\mathfrak{S}_n \wr \mathfrak{S}_d}) \\
&= \dim [\lambda]^{\mathfrak{S}_n \wr \mathfrak{S}_d},
\end{aligned}$$

provided $\ell(\lambda) \le \dim V$. ◻

19.3.4 Semistandard tableaux again: Gay's theorem

We follow [46, Sec. 4.3(A)]. Since the HWVs of type λ in $Sym^d Sym^n V$ are explicitly described by $\mathfrak{S}_n \wr \mathfrak{S}_d$-invariants in $[\lambda]$ (see the proof of Proposition 19.8), we study those invariants now.

Recall that a standard tableau indexes a vector in $[\lambda] = \{\lambda\}^0$, but via Schur–Weyl duality it also indexes a vector in $\mathsf{HWV}_\lambda(\bigotimes^{dn} V) \cong [\lambda]$. This can be confusing at first, but this beautiful correspondence makes things a lot easier.

Lemma 19.9. *Let V be an $\mathfrak{S}_n \wr \mathfrak{S}_d$-representation. Then we have an action of $\mathfrak{S}_d$ on the invariant space $V^{\mathfrak{S}_n^d}$. Moreover, $V^{\mathfrak{S}_n \wr \mathfrak{S}_d} = (V^{\mathfrak{S}_n^d})^{\mathfrak{S}_d}$.*

Proof. Interchanging the block structure of invariant blocks keeps the blocks invariant, so we have an $\mathfrak{S}_d$-action. The second statement follows from our definition of $\mathfrak{S}_n \wr \mathfrak{S}_d$. □

It turns out that the $\mathfrak{S}_n^d$-invariants of $[\lambda]$ can be easily understood using semistandard tableaux. Let $\{\lambda\}^{d\times n}$ denote the $d \times n$ weight space in $\{\lambda\}$ (recall that a basis of $\{\lambda\}^{d\times n}$ is given by the e_T where T is semistandard of shape λ and content $d \times n$). Consider the map

$$\varphi : \{\lambda\}^0 \to \{\lambda\}^{d\times n} \tag{19.7}$$

that replaces each entry $1, 2, \ldots, n$ by 1, each entry $n+1, n+2, \ldots, 2n$ by 2, and so on. This is an application of a matrix in $\mathsf{End}(\mathbb{C}^{|\lambda|})$, for example:

$$\varphi\left(\begin{array}{|c|c|c|c|c|}\hline 2 & 1 & 3 & 5 & 8 \\ \hline 4 & 7 & 9 & 10 \\ \cline{1-4} 6 & 12 \\ \cline{1-2} 11 \\ \cline{1-1}\end{array}\right) = \begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 1 & 2 & 3 \\ \hline 2 & 3 & 3 & 4 \\ \cline{1-4} 2 & 4 \\ \cline{1-2} 4 \\ \cline{1-1}\end{array} \qquad \text{for } n = 3, d = 4$$

is obtained by applying the matrix

$$\begin{pmatrix}
1&1&1&0&0&0&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0&0&0&0\\
0&0&0&0&0&0&1&1&1&0&0&0\\
0&0&0&0&0&0&0&0&0&1&1&1\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0
\end{pmatrix} \cdot \begin{array}{|c|c|c|c|c|}\hline 2 & 1 & 3 & 5 & 8 \\ \hline 4 & 7 & 9 & 10 \\ \cline{1-4} 6 & 12 \\ \cline{1-2} 11 \\ \cline{1-1}\end{array} = \begin{array}{|c|c|c|c|c|}\hline 1 & 1 & 1 & 2 & 3 \\ \hline 2 & 3 & 3 & 4 \\ \cline{1-4} 2 & 4 \\ \cline{1-2} 4 \\ \cline{1-1}\end{array}.$$

For a standard tableau T there are two cases:

1. $\varphi(T)$ is semistandard or

2. $\varphi(T)$ has a column with a repeated entry.

Using the relations in the columns, we see that only the semistandard tableaux correspond to nonzero vectors. Moreover, distinct semistandard tableaux correspond to linearly independent vectors in $\{\lambda\}^{d\times n}$.

The map φ gives a way of representing the $\mathfrak{S}_n^d$-invariants:

Lemma 19.10 (Gay's theorem [38])**.** *Let $\mathfrak{S}_d$ act on $[\lambda]$ by permuting the d blocks, i. e., for example the transposition $(1\ 2)$ switches 1 with $n+1$, switches 2 with $n+2, \ldots,$ and switches n with $2n$. The map φ is $\mathfrak{S}_d$-equivariant. Moreover, if we restrict φ to the invariant space $[\lambda]^{\mathfrak{S}_n^d}$, then it becomes an isomorphism of $\mathfrak{S}_d$-representations $[\lambda]^{\mathfrak{S}_n^d}$ and $\{\lambda\}^{d\times n}$.*

Proof. Permuting blocks and then setting the whole block to a number has the same effect as setting the whole block to a number and then permuting the blocks. Thus φ is $\mathfrak{S}_d$-equivariant.

The inverse function φ^{-1} is given (up to scale) by summing over all ways of replacing each entry i in the semistandard tableau by all entries $i(n-1)+j$, $1 \le j \le n$. For example

$$\varphi^{-1}\Big(\begin{array}{|c|c|}\hline 1&1\\\hline 2&2\\\hline\end{array}\Big) \quad\mapsto\quad \begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array}+\begin{array}{|c|c|}\hline 2&1\\\hline 3&4\\\hline\end{array}+\begin{array}{|c|c|}\hline 1&2\\\hline 4&3\\\hline\end{array}+\begin{array}{|c|c|}\hline 2&1\\\hline 4&3\\\hline\end{array}$$

□

Remark 19.11. Note that

$$\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array}+\begin{array}{|c|c|}\hline 2&1\\\hline 3&4\\\hline\end{array}+\begin{array}{|c|c|}\hline 1&2\\\hline 4&3\\\hline\end{array}+\begin{array}{|c|c|}\hline 2&1\\\hline 4&3\\\hline\end{array} = 2\left(\begin{array}{|c|c|}\hline 1&2\\\hline 3&4\\\hline\end{array}+\begin{array}{|c|c|}\hline 2&1\\\hline 3&4\\\hline\end{array}\right),$$

which explains the choice at the end of Section 19.3.1.

Taking $\mathfrak{S}_d$-invariants in Lemma 19.10 we obtain:

Corollary 19.12. $a_\lambda(d,n) = \dim(\{\lambda\}^{d\times n})^{\mathfrak{S}_d}$.

The following example shows how we can calculate plethysm coefficients using tableaux.

Example 19.13. For $\lambda = (2,2)$ there is a unique semistandard tableau of shape λ with rectangular content 2×2:

$$\begin{array}{|c|c|}\hline 1&1\\\hline 2&2\\\hline\end{array}.$$

Thus $\dim[\lambda]^{\mathfrak{S}_2^2} = 1$. Let us consider the action of $\mathfrak{S}_2$:

$$\tfrac{1}{2}\left(\begin{array}{|c|c|}\hline 1&1\\\hline 2&2\\\hline\end{array}+(1\ 2)\begin{array}{|c|c|}\hline 1&1\\\hline 2&2\\\hline\end{array}\right) = \begin{array}{|c|c|}\hline 1&1\\\hline 2&2\\\hline\end{array}$$

and thus $\dim[\lambda]^{\mathfrak{S}_2\wr\mathfrak{S}_2} = 1 = a_{(2,2)}(2,2)$. This invariant tableau can therefore now be used with the above constructions to find the unique HWV of weight $(2,2)$: The discriminant, as it was done in Example 19.4.

Corollary 19.14. *Via Schur–Weyl duality we identify tableaux S of shape λ that have content $(1, 1 \dots, 1)$ with vectors $v_S \in \mathsf{HWV}_\lambda(\bigotimes^d \bigotimes^n V)$. The vector space $\mathsf{HWV}_\lambda(\bigotimes^d Sym^n V)$ has as a basis the highest weight vectors $\varphi^{-1}(T)$, where T ranges over all semistandard tableaux of shape λ with content $d \times n$. Moreover, the vector space $\mathsf{HWV}_\lambda(Sym^d Sym^n V)$ is spanned by the highest weight vectors $\sum_{\pi \in \mathfrak{S}_d} \varphi^{-1}(\pi T)$, where T ranges over all semistandard tableaux of shape λ with content $d \times n$.*

Example 19.15.

$$Sym^2 Sym^2 \mathbb{C}^m = \{(2, 2)\} \oplus \{(4)\} \text{ for all } m \geq 2.$$

Proof.

- There is a unique semistandard tableau of shape $(2, 2)$ and rectangular content 2×2:

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 2 & 2 \\ \hline \end{array}.$$

  It is invariant under $\mathfrak{S}_2$.

- There is a unique semistandard tableau of shape $(4, 0)$ and rectangular content 2×2:

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & 2 & 2 \\ \hline \end{array}.$$

  It is invariant under $\mathfrak{S}_2$.

- There is a unique semistandard tableau of shape $(3, 1)$ and rectangular content 2×2:

$$\begin{array}{|c|c|c|} \hline 1 & 1 & 2 \\ \hline 2 & \multicolumn{2}{c}{} \\ \cline{1-1} \end{array}. \tag{19.8}$$

  It vanishes under symmetrization over $\mathfrak{S}_2$.

- There are no semistandard tableaux of shape $(2, 1, 1)$ or $(1, 1, 1, 1)$ with rectangular content 2×2.

□

Some more small examples:

Theorem 19.16 (Howe's theorem [44, Thm. 4.3])**.** *Let $d > 1$. Let $d \times n := (n, n, \dots, n)$ denote the partition of nd whose Young diagram is rectangular with n columns and d rows.*

$$a_{d \times n}(d, n) = \begin{cases} 1 & \text{if } n \text{ is even}, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. There is exactly 1 semistandard tableau of shape $d \times n$ and rectangular content $d \times n$. For example:

$$\begin{array}{|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 2 & 2 & 2 & 2 & 2 & 2 \\ \hline 3 & 3 & 3 & 3 & 3 & 3 \\ \hline 4 & 4 & 4 & 4 & 4 & 4 \\ \hline \end{array}.$$

Applying a transposition $(i\ j) \in \mathfrak{S}_d$ gives a sign change for each column. So the tableau is invariant iff n is even. □

Proposition 19.17. *A partition λ is called a* nontrivial hook *if* $\lambda = (k, 1^m) := (k, \underbrace{1, 1, \ldots, 1}_{m \text{ times}})$ *with* $k \geq 1$ *and* $m \geq 1$. *If λ is a partition of nd and λ is a nontrivial hook, then the plethysm coefficient* $a_\lambda(d, n)$ *is zero.*

Proof. This generalizes the vanishing of eq. (19.8) under symmetrization.

Let $\lambda \vdash nd$ be a nontrivial hook. Pick any Young tableau T of shape λ with rectangular content $d \times n$. Then T will have at least two numbers a and b appearing in the first column, for example

$$\begin{array}{|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 2 & 2 & 3 & 3 & 4 & 4 \\ \hline 2 \\ \cline{1-1} 3 \\ \cline{1-1} 4 \\ \cline{1-1} \end{array}$$

Then pick two numbers a and b that appear in the first column. Then $T + (a\ b)T = 0$ and thus T vanishes when symmetrizing over a cardinality 2 subgroup of $\mathfrak{S}_d$, therefore T also vanishes when symmetrizing over the whole $\mathfrak{S}_d$. □

Explicit construction of HWVs

The explicit construction of highest weight vectors gives us first results about plethysm coefficients. In Chapter 20 we will lift these techniques to multiplicities in coordinate rings of orbit closures.

20 Tensor contraction

In this chapter we describe a combinatorial interpretation of the contraction in (19.6) and give several applications, including strong restrictions on the shape of partitions λ that can potentially serve as obstructions. We will see that those λ require a very long first part.

20.1 Contracting highest weight vectors in plethyms with rank one tensors

In this section we give a combinatorial interpretation of the contraction in (19.6). This will enable us to deduce restrictions on the possible λ that can serve as obstructions.

Let $V = \mathbb{C}^N$ and let $s : \{1, \ldots, dn\} \to \mathbb{C}^N$.

We view $\lambda \vdash dn$ as a Young diagram and, for convenience, denote by λ also the set of boxes of the diagram. Recall the map φ from (19.7).

Definition 20.1. Let T be a tableau of shape λ with content $d \times n$ and $\vartheta \colon \lambda \to [dn]$ be a bijection. This results in a tableau S_ϑ with content $(1, \ldots, 1)$. We say that ϑ *respects* T iff there exists a permutation $\tau \in \mathfrak{S}_d$ such that $\varphi(S_\vartheta) = T\tau(T)$.

Clearly for a given T there are $(n!)^d d!$ maps ϑ that respect T, which is the size of the wreath product group $\mathfrak{S}_n \wr \mathfrak{S}_d$.

Pictorially, the composition $s \circ \vartheta$ puts *vectors* in the tableau cells.

Let $j = (j_1, \dots, j_k)$ be a list of vectors in $\mathbb{C}^N$. We define

$$\det(j_1, \dots, j_k) \coloneqq \langle e_1 \wedge e_2 \wedge \dots \wedge e_k, j_1 \otimes j_2 \otimes \dots \otimes j_k \rangle, \tag{20.1}$$

which is the determinant of the top $k \times k$ submatrix of the $N \times k$ matrix j.

Suppose $\vartheta \colon \lambda \to [dn]$ respects the tableau T of shape λ with content $d \times n$, and take a map $s \colon \{1, \dots, dn\} \to \mathbb{C}^N$. We define the *value* $\mathrm{val}_\vartheta(s)$ of ϑ at $s \colon \{1, \dots, dn\} \to \mathbb{C}^N$ by

$$\mathrm{val}_\vartheta(s) \coloneqq \prod_{\text{column } c \text{ of } \lambda} \det(s(\vartheta(1, c)), \dots, s(\vartheta(\mu_c, c)), \tag{20.2}$$

where $\mu = \lambda^t$. This is natural in the following sense:

$$\mathrm{val}_\vartheta(s) \quad = \quad \langle e_1 \wedge \dots \wedge e_{\mu_1} \otimes \dots \otimes e_1 \wedge \dots \wedge e_{\mu_{\lambda_1}}, s(\vartheta(1,1)) \otimes s(\vartheta(2,1)) \otimes \dots \otimes s(\vartheta(\mu_{\lambda_1}, \lambda_1))\rangle$$

Pictorially, $s \circ \vartheta$ places vectors in the tableau and $\mathrm{val}_\vartheta(s)$ is the product over the column determinants.

Theorem 20.2. *Let T be a tableau of shape $\lambda \vdash dn$ with content $d \times n$ and let $\overline{v}_T$ be the corresponding HWV in $\mathsf{HWV}_\lambda(\bigotimes^d Sym^n V)$ (Cor. 19.14). Let $\tilde{v}_T \coloneqq \frac{1}{d!} \sum_{\pi \in \mathfrak{S}_d} \pi \overline{v}_T$ be the HWV in $\mathsf{HWV}_\lambda(Sym^d Sym^n V)$. Let $s \colon \{1, \dots, dn\} \to \mathbb{C}^N$ be a map. Then*

$$\langle \tilde{v}_T, s(1) \otimes \ldots \otimes s(dn) \rangle = \frac{1}{d!\, n!^d} \sum_\vartheta \mathrm{val}_\vartheta(s),$$

where the sum is over all bijections $\vartheta \colon \lambda \to \{1, \dots, dn\}$ respecting T.

Proof. Let S be a tableau with content $(1, 1, \dots, 1)$ such that $\varphi(S) = T$. Let $\pi \in \mathfrak{S}_{dn}$ any permutation such that $\pi S_0 = S$, where S_0 is the standard tableau that is ordered columnwise from left to right, top to bottom.

$$\begin{aligned}
\langle \tilde{v}_T, s(1) \otimes \ldots \otimes s(dn) \rangle &= \tfrac{1}{d!\, n!^d} \sum_{\sigma \in \mathfrak{S}_d \wr \mathfrak{S}_d} \langle \sigma v_S, s(1) \otimes \ldots \otimes s(dn) \rangle \\
&= \tfrac{1}{d!\, n!^d} \sum_{\sigma \in \mathfrak{S}_d \wr \mathfrak{S}_d} \langle v_S, \sigma(s(1) \otimes \ldots \otimes s(dn)) \rangle \\
&= \tfrac{1}{d!\, n!^d} \sum_{\sigma \in \mathfrak{S}_d \wr \mathfrak{S}_d} \langle \pi v_{S_0}, \sigma(s(1) \otimes \ldots \otimes s(dn)) \rangle \\
&= \tfrac{1}{d!\, n!^d} \sum_{\sigma \in \mathfrak{S}_d \wr \mathfrak{S}_d} \langle v_{S_0}, \pi^{-1}\sigma(s(1) \otimes \ldots \otimes s(dn)) \rangle \\
&= \tfrac{1}{d!\, n!^d} \sum_\vartheta \mathrm{val}_\vartheta(s),
\end{aligned}$$

where the last equality follows from $v_{S_0} = e_1 \wedge \dots \wedge e_{\mu_1} \otimes \dots \otimes e_1 \wedge \dots \wedge e_{\mu_{\lambda_1}}$. ☐

20.2 Applications: Waring rank and a proof of Weintraub's conjecture

20.2.1 The discriminant

We let

$$T := \begin{array}{|c|c|}\hline 1 & 1 \\ \hline 2 & 2 \\ \hline \end{array}$$

and aim to evaluate $\tilde{v}_T$ at a tensor of Waring rank 1: ℓ^2, where $\ell \in \mathbb{C}^M$.

We observe $(\ell^2)^{\otimes 2} = \ell \otimes \ell \otimes \ell \otimes \ell$ and set $s = (\ell, \ell, \ell, \ell)$.

$$\langle \tilde{v}_T, (\ell^2)^{\otimes 2} \rangle \;=\; \tfrac{1}{8} \sum_{\vartheta \text{ respecting } T} \mathrm{val}_\vartheta(s)$$

But every summand $\mathrm{val}_\vartheta(s)$ is zero, because it is a product of 2×2 determinants in which both columns are ℓ. Therefore the discriminant vanishes on Waring rank 1 polynomials.

20.2.2 Too many rows

The following observation generalizes the discriminant.

Observation 20.3. *Let T be a semistandard tableau of shape λ and content $d \times n$ with more than k rows. Then $\tilde{v}_T$ vanishes on all points of border Waring rank $\leq k$.*

Proof. Let $h = \ell_1^n + \ell_2^n + \cdots + \ell_k^n$ for linear forms ℓ_i.

$$\begin{aligned}
\langle \tilde{v}_T, h^{\otimes d} \rangle &= \sum_{1 \leq a_1, \ldots, a_d \leq k} \langle \tilde{v}_T, \ell_{a_1}^n \otimes \cdots \otimes \ell_{a_d}^n \rangle \\
&= \tfrac{1}{d!(n!)^d} \sum_{1 \leq a_1, \ldots, a_d \leq k} \sum_{\vartheta} \mathrm{val}_\vartheta(\ell_{a_1}, \ldots, \ell_{a_1}, \ell_{a_2}, \ldots, \ell_{a_2}, \ldots, \ldots, \ell_{a_d}, \ldots, , \ell_{a_d})
\end{aligned}$$

As for the discriminant, each summand vanishes independently. Indeed, for

$$s = (\ell_{a_1}, \ldots, \ell_{a_1}, \ell_{a_2}, \ldots, \ell_{a_2}, \ldots, \ldots, \ell_{a_d}, \ldots, , \ell_{a_d}),$$

the map $s \circ \vartheta$ places the ℓ_i in the tableau such that each position that has the same number in T gets the same ℓ_i. But the first column of T has more than k different numbers. So by the pigeonhole principle $s \circ \vartheta$ puts at least two coinciding ℓ_i in the first column. Thus $\mathrm{val}_\vartheta(s) = 0$ because a determinant with a repeating column is zero. $\square$

20.2.3 Aronhold's invariant

We can do a little bit better, i. e., use fewer rows, as Aronhold's invariant in $Sym^4 Sym^3 V$ shows (and this can be generalized):

$$T := \begin{array}{|c|c|c|c|}\hline 1 & 1 & 1 & 2 \\ \hline 2 & 2 & 3 & 3 \\ \hline 3 & 4 & 4 & 4 \\ \hline \end{array}.$$

Let $h \in Sym^3 V$ be of Waring rank 3, $h = \ell_1^3 + \ell_2^3 + \ell_3^3$. Then

$$h^{\otimes 4} = \ell_1^3 \otimes \ell_1^3 \otimes \ell_1^3 \otimes \ell_1^3 + \ell_1^3 \otimes \ell_1^3 \otimes \ell_1^3 \otimes \ell_2^3 + \cdots + \ell_3^3 \otimes \ell_3^3 \otimes \ell_3^3 \otimes \ell_3^3.$$

$$\begin{aligned} \langle \tilde{v}_T, h^{\otimes 4} \rangle &= \sum_{1 \le a,b,c,d \le 3} \langle \tilde{v}_T, \ell_a^3 \otimes \ell_b^3 \otimes \ell_c^3 \otimes \ell_d^3 \rangle \\ &= \tfrac{1}{4!(3!)^4} \sum_{1 \le a,b,c,d \le 3} \sum_{\vartheta} \mathrm{val}_\vartheta(\ell_a, \ell_a, \ell_a, \ell_b, \ell_b, \ell_b, \ell_c, \ell_c, \ell_c, \ell_d, \ell_d, \ell_d) \end{aligned}$$

Again each summand vanishes independently. Indeed, for

$$s = (\ell_a, \ell_a, \ell_a, \ell_b, \ell_b, \ell_b, \ell_c, \ell_c, \ell_c, \ell_d, \ell_d, \ell_d),$$

the map $s \circ \vartheta$ places the ℓ_i in the tableau such that each position that has the same number in T gets the same ℓ_i. For a nonzero summand it is required that ϑ puts different vectors ℓ_i on the numbers 1,2,3 because of the first column. But in T the number 4 shares columns with each 1,2,3 and there is no 4th different vector ℓ_i. Thus for every ϑ there is at least one column in which the determinant vanishes because of a repeated column.

20.2.4 Proof of Weintraub's conjecture

With one additional idea it is now straightforward to prove Weintraub's conjecture that allows us to create nonzero $\tilde{v}_T$:

Theorem 20.4 ([18])**.** *Let n be even. Given a partition λ of dn into at most d parts such that all λ_i are even. Then $a_\lambda(d, n) > 0$.*

Proof. If λ is even, $\ell(\lambda) \le d$, then we can fill it with content $d \times n$ such that each column appears twice (for example in a greedy fashion, taking the column pair with the most empty rows first). For example

$$T = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline 1 & 1 & 1 & 1 & 4 & 4 & 2 & 2 & 3 & 3 & 4 & 4 \\ \hline 2 & 2 & 2 & 2 & 1 & 1 \\ \cline{1-6} 3 & 3 & 3 & 3 \\ \cline{1-4} 4 & 4 \\ \cline{1-2} \end{array}\,.$$

Now choose a homogeneous polynomial h as the sum of d many homogeneous degree n linear forms with *real* coefficients:

$$h = \ell_1^n + \cdots + \ell_d^n,$$

where each $\ell_i \in \mathbb{R}^d$. Here we choose each ℓ_i in a "generic" way, i. e., its entries should be algebraically independent or at least all top $k \times k$ determinants of all subsets of k vectors ℓ_i should be nonzero. We will see that

$$\langle \tilde{v}_T, h^{\otimes d} \rangle > 0. \tag{20.3}$$

Indeed,

$$\langle \tilde{v}_T, h^{\otimes d}\rangle = \tfrac{1}{d!(n!)^d} \sum_{1\le a_1,a_2,\dots,a_d\le d} \sum_{\vartheta} \mathrm{val}_\vartheta(\ell_{a_1},\dots,\ell_{a_1},\ell_{a_2},\dots,\ell_{a_2},\dots\dots,\ell_{a_d},\dots,\ell_{a_d})$$

When expanding we get a sum of products of determinants, but in each product each determinant appears an even number of times. Therefore we sum over squares of real numbers! Since squaring nonzero real numbers results in positive real numbers, (20.3) will be positive. It remains to show that there is at least one positive summand, but that is given for example for the summand ϑ in which ℓ_i is placed on i. □

20.3 Application: Obstructions require long first rows

For a partition λ we define $\bar{\lambda}$ to be its *body*, i. e., $\bar{\lambda}$ is obtained from λ by removing its first row.

The following insight is due to Kadish and Landsberg [48]. It puts a strong restriction on the types λ that can be used to separate the determinant from the padded permanent.

Proposition 20.5 ([48]). *Let* $\Omega_{n,m} := \overline{\mathsf{GL}_{n^2}(X_{1,1}^{n-m}\mathrm{per}_m)}$. *If* $\lambda \vdash nd$ *occurs in* $\mathbb{C}[\Omega_{n,m}]_d$, *then* $\ell(\lambda) \le m^2+1$ *and* $|\bar{\lambda}| \le md$.

Proof. Suppose that $\lambda \vdash nd$ satisfies $\ell(\lambda) > m^2+1$ or $|\bar{\lambda}| > md$. We need to show that λ does not occur in $\mathbb{C}[\Omega_{n,m}]_d$. We will show that $\langle \tilde{v}_T, h^{\otimes d}\rangle = 0$ for every semistandard tableau T of shape λ with content $d\times n$ and every $h \in \Omega_{n,m}$. So fix a tableau T of shape λ with content $d \times n$ and fix a point $h \in \Omega_{n,m}$.

Assume first that $\ell := \ell(\lambda) > m^2+1$. We prove that $\langle \tilde{v}_T, h^{\otimes d}\rangle = 0$. We use an argument that is very similar to Observation 20.3. For this, it suffices to show that $\langle \tilde{v}_T, t\rangle = 0$ for all tensors $t = s(1)\otimes\cdots\otimes s(dn)$, where $s\colon \{1,\dots,dn\} \to \{e_1,\dots,e_{m^2+1}\}$ and e_i are the standard basis vectors of $\mathbb{C}^{m^2+1}$. Indeed, for every ϑ we have $\mathrm{val}_\vartheta(s) = 0$ because the determinant that corresponds to the first column is zero: It is a determinant of a $\ell\times\ell$ matrix, $\ell > m^2+1$, whose last $\ell-(m^2+1)$ rows are zero. Thus $\langle \tilde{v}_T, h^{\otimes d}\rangle = 0$.

Assume now $|\bar{\lambda}| > md$, so that $\lambda_1 < (n-m)d$. For $g \in \mathsf{GL}_{n^2}$ let $Z := gX_{1,1}$ so that $Z^{n-m}\cdot g\mathrm{per}_m \in \mathsf{GL}_{n^2}(X_{1,1}^{n-m}\mathrm{per}_m)$. Let $q := Z^{n-m}g\mathrm{per}_m$. We can express $q^{\otimes d}$ as a linear combination of tensors $t = s(1)\otimes\cdots\otimes s(dn)$, where $s\colon \{1,\dots,dn\} \to \mathbb{C}^M$ maps at least $(n-m)d$ elements to the vector Z. Fix such a tensor t. It suffices to show that $\langle \tilde{v}_T, t\rangle = 0$. Indeed, each summand of $\sum_\vartheta \mathrm{val}_\vartheta(s)$ vanishes independently. This can be seen as follows. Since $\lambda_1 < (n-m)d$, the partition λ has less than $(n-m)d$ columns. By the pigeonhole principle, there is a column c in which $s\circ\vartheta$ puts a vector Z in at least two boxes. Thus the determinant corresponding to this column vanishes. □

Explicit construction of HWVs

Evaluation of highest weight vectors can be defined via tensor contraction. In several cases this tensor contraction can be fully understood, so that nontrivial results about multiplicities in coordinate rings of orbit closures can be deduced, for example equations for Waring rank.

21 Good occurrence obstructions for determinant vs padded permanent do not exist

In this chapter, we will see that showing $(X_1)^{n-m}\mathrm{per}_m \notin \overline{\mathsf{GL}_{n^2}\mathrm{det}_n}$ for superpolynomially large n cannot be achieved with occurrence obstructions. The padding is crucial for the proof of this result. It is astonishing that such a seemingly small modification has such a large impact (on the proof, the complexity of the polynomial is of course not changed by much). Multiplicity obstructions might still work, as well as occurrence obstructions in models of computation where no padding is involved. One such example is homogeneous iterated matrix multiplication. The hope that multiplicity obstructions might work (at least in models without padding) comes from the fact that in many situations group orbits and orbit closures can be reconstructed without loss of information from the multiplicities in their coordinate rings [58, 83, 22].

We roughly follow [23]. The crucial tool is the semigroup property that is explained in Section 21.2.1. It is a well-known theorem of independent interest: If λ and μ occur in the coordinate ring of a GL_n-orbit closure, then $\lambda + \mu$ occurs as well. This means that we only have to prove the occurrence for a large enough set of building blocks so that we can obtain every partition as a sum of building blocks. We will see that this essentially works in the range of all “interesting” partitions. This splitting of partitions into building blocks is the main argument, see Section 21.2.2. The range of all “interesting” partitions is cut down significantly by the requirement to have a long first row and only very few rows, see Section 20.3. Moreover, we rule out that a polynomially bounded number of boxes in the Young diagram suffices, see Section 21.1. This means that we have only polynomially many rows, but superpolynomially many columns and a very long first row. We show directly that the rectangular partitions with even number of columns and an additional long first row occur (rectangular blocks with an even number of columns are particularly easy to understand, and the long first row is handled by a lifting theorem, see Section 21.1.3), so using these building blocks we can construct any partition that has even row lengths, few rows, and a long first row. If there were no odd row lengths, then we would be done at this point. We handle odd row lengths by proving that wide rectangular blocks with one additional other column and an additional long first row also occur. For this type of building block we can prove occurrence directly by evaluating on a padded power sum, see Section 21.2.3.

21.1 The degree lower bound

In this section we prove a lower bound on the degree d that (polynomial) obstructions must have: $d > \sqrt{\frac{n}{m}}$. In particular if we want to prove superpolynomial lower bounds on the border determinantal complexity of the permanent, we need superpolynomially high degree.

21.1.1 Padded low-rank embedding

We will apply the following theorem with $s = md$.

Theorem 21.1. *Let n, s, d be positive integers such that $n \geq sd$ and*

$$Z, v_{1,1}, \ldots, v_{1,s}, v_{2,1}, \ldots, \ldots, v_{d,s} \in V.$$

Then we have $Z^{n-s}(v_{1,1}v_{1,2}\cdots v_{1,s} + \cdots + v_{d,1}v_{d,2}\cdots v_{d,s}) \in \overline{\mathsf{GL}_{n^2}\mathrm{det}_n}$.

Proof. Let $X_{1,1}, \ldots, X_{d,s}, \ldots, X_n$ denote the standard basis of V (that last $n - ds$ variables are indexed by just one integer). Writing the polynomial $X_{1,1}X_{1,2}\cdots X_{1,s} + \cdots + X_{d,1}X_{d,2}\cdots X_{d,s}$ as a formula requires at most $(s-1)d + d - 1 = sd - 1$ many additions and multiplications. Valiant's construction [80] implies that $X_{1,1}X_{1,2}\cdots X_{1,s} + \cdots + X_{d,1}X_{d,2}\cdots X_{d,s}$ has the determinantal complexity at most $sd \leq n$, i. e., it can be written as the determinant of an $n \times n$-matrix with affine linear entries in $X_{1,1}, \ldots, X_{d,s}$. The determinantal complexity is invariant under invertible linear transformations. Hence the determinantal complexity of $v_{1,1}v_{1,2}\cdots v_{1,s} + \cdots + v_{d,1}v_{d,2}\cdots v_{d,s}$ is at most n, for any linearly independent system $v_{1,1}, \ldots, v_{d,s}$ of linear combinations of $X_1, \ldots, X_{n^2}$. By homogenizing with respect to a new variable Y and then substituting Y by $Z \in V$, we see that $Z^{n-s}(v_{1,1}v_{1,2}\cdots v_{1,s} + \cdots + v_{d,1}v_{d,2}\cdots v_{d,s}) \in \overline{\mathsf{GL}_{n^2}\mathrm{det}_n}$. Since $\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}$ is closed, we can drop the assumption that the $v_{i,j}$ are linearly independent, by taking limits. □

21.1.2 Low-rank evaluation

We present now a useful lemma on the evaluation of polynomials at "points of low rank".

Lemma 21.2. *Let $f \in Sym^d W$ be such that $f(\sum_{j=1}^r v_j) \neq 0$ for some $v_1, \ldots, v_r \in W$. Then there exists $S \subseteq \{1, \ldots, r\}$ with $|S| \leq d$ and $f(\sum_{j \in S} v_j) \neq 0$.*

Proof. (an alternative proof is given in [23, Lem. 3.1]) Let $[r] := \{1, \ldots, r\}$. For a subset $S \subset [r]$ we write $v_S := v_{S_1} + v_{S_2} + \cdots + v_{S_{|S|}}$. Let $v := v_{[r]}$. For a map $\sigma : [d] \to [r]$ we write $v_\sigma := v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(d)}$. In the following inclusion/exclusion calculation we can assume all sets S_i to have at most i elements. We will express $f(v)$ as a linear combination of evaluations

$f(v_S)$, where $|S| \le d$, $S \subseteq [r]$. A two-headed arrow "$\twoheadrightarrow$" indicates a surjective map.

$$
\begin{aligned}
f(v) &= \langle f, v^{\otimes d} \rangle = \sum_{\sigma:[d]\to[r]} \langle f, v_\sigma \rangle \\
&= \sum_{S_d \subset [r]} \sum_{\sigma:[d]\twoheadrightarrow S_d} \langle f, v_\sigma \rangle \\
&= \sum_{S_d \subset [r]} \left(f(v_{S_d}) - \sum_{\sigma:[d]\to S_d \text{ not surj}} \langle f, v_\sigma \rangle \right) \\
&= \sum_{S_d \subset [r]} \left(f(v_{S_d}) - \sum_{S_{d-1} \subsetneq S_d} \sum_{\sigma:[d]\twoheadrightarrow S_{d-1}} \langle f, v_\sigma \rangle \right) \\
&= \sum_{S_d \subset [r]} \left(f(v_{S_d}) - \sum_{S_{d-1} \subsetneq S_d} \left(f(v_{S_{d-1}}) - \sum_{S_{d-2} \subsetneq S_{d-1}} \sum_{\sigma:[d]\twoheadrightarrow S_{d-2}} \langle f, v_\sigma \rangle \right) \right) \\
&= \sum_{S_d \subset [r]} \left(f(v_{S_d}) - \sum_{S_{d-1} \subsetneq S_d} \left(f(v_{S_{d-1}}) - \sum_{S_{d-2} \subsetneq S_{d-1}} \left(\cdots \left(\sum_{S_1 \subsetneq S_2} \sum_{\sigma:[d]\twoheadrightarrow S_1} \underbrace{\langle f, v_\sigma \rangle}_{=f(v_{S_1})} \right) \cdots \right) \right) \right)
\end{aligned}
$$

Hence we see that $f(v)$ can be expressed as a linear combination of evaluations $f(v_S)$, where $|S| \le d$. Since $f(v) \neq 0$ there exists S such that $f(v_S) \neq 0$. $\square$

A direct corollary of Lemma 21.2 (using v_j to be the monomial basis) is the following.

Corollary 21.3. *Let V be a finite dimensional $\mathbb{C}$-vector space and $d, n \ge 1$. If $f \in Sym^d Sym^n V$ is nonzero, then there exists a polynomial $h \in Sym^n V$ that has at most d nonzero coefficients such that $f(h) \neq 0$.*

21.1.3 The inner degree tableau lifting

Given a tableau with shape λ and content $d \times n$ we define the *lifted tableau* to be the tableau of shape $\lambda + (d)$ where we append the entries $1, \ldots, d$ to the first row. For example:

$$
\begin{ytableau}
1 & 1 & 1 & 1 & 4 & 4 & 2 & 2 & 3 & 3 & 4 & 4 \\
2 & 2 & 2 & 2 & 1 & 1 \\
3 & 3 & 3 & 3 \\
4 & 4
\end{ytableau}
\quad \overset{\text{lifting}}{\longmapsto} \quad
\begin{ytableau}
1 & 1 & 1 & 1 & 4 & 4 & 2 & 2 & 3 & 3 & 4 & 4 & 1 & 2 & 3 & 4 \\
2 & 2 & 2 & 2 & 1 & 1 \\
3 & 3 & 3 & 3 \\
4 & 4
\end{ytableau}
$$

If $\tilde{v}_T$ is a HWV in $Sym^d Sym^n V$, then $\tilde{v}_{T'}$ is a HWV in $Sym^d Sym^{n+1} V$, where T' is the lifted tableau of T. The lifting is an injective map $\mathsf{HWV}_\lambda(Sym^d Sym^n V) \hookrightarrow \mathsf{HWV}_{\lambda+(n)}(Sym^d Sym^{n+1} V)$.

The proof can be readily obtained from the semigroup property of the partitions in the coordinate ring $\mathbb{C}[\mathsf{GL}_d(X_1X_2\cdots X_d)]$, but we omit the details, see [23].

We can apply the lifting several times and also call the result a lifted tableau.

Recall the monomial

$$X^\alpha = \frac{1}{d!}\sum_{\pi\in\mathfrak{S}_d}\pi(X_{\alpha_1}\otimes\cdots\otimes X_{\alpha_d})$$

We define a linear map $Sym^n V \to Sym^{n+1}V$ that is suited to this lifting (and which is basically a rescaled multiplication with X_1):

$$X_1 \boxdot (\sum_{\pi\in\mathfrak{S}_d}\pi(X_{\alpha_1}\otimes\cdots\otimes X_{\alpha_d})) := \frac{1}{k}\sum_{\pi\in\mathfrak{S}_{d+1}}\pi(X_{\alpha_1}\otimes\cdots\otimes X_{\alpha_d}\otimes X_1)$$

where $k = |\{j \mid \alpha_j = 1\}| + 1$ is the X_1-degree of the right-hand side. The crucial property is that the summands on the left-hand side are in bijection (taking into account the rescaling factor $\frac{1}{k}$) to the summands on the right-hand side which have X_1 as their last tensor factor. In terms of monomials, this is a rescaled multiplication with X_1. The rescaling factor is chosen so that is works well with the lifting map $\tilde{v}_T \mapsto \tilde{v}_{T'}$:

$$\langle \tilde{v}_T, (p)^{\otimes d}\rangle = \langle \tilde{v}_{T'}, (X_1 \boxdot p)^{\otimes d}\rangle \tag{21.1}$$

which can be seen immediately by comparing summands in the tensor contractions (this is also the reason why we state (21.1) in terms of tensors and not monomials, which would also be perfectly possible. In fact, [23, Sec. 5(b)] gives a more algebraic presentation of this construction). Applying $X_1 \boxdot h$ to a polynomial preserves the number of monomials that have a nonzero coefficient, but it changes the coefficients individually, depending on the number of occurrences of X_1 in each monomial. We write $(X_1)^{n-m} \boxdot p := X_1 \boxdot (X_1 \boxdot (\cdots \boxdot p))$.

Corollary 21.4. *Let n, s, d be positive integers such that $n \geq sd$ and*

$$Z, v_{1,1}, \ldots, v_{1,s}, v_{2,1}, \ldots, \ldots, v_{d,s} \in V.$$

Then we have $Z^{n-s} \boxdot (v_{1,1}v_{1,2}\cdots v_{1,s} + \cdots + v_{d,1}v_{d,2}\cdots v_{d,s}) \in \overline{\mathsf{GL}_{n^2}\mathsf{det}_n}$.

Proof. $Z^{n-s} \boxdot (v_{1,1}v_{1,2}\cdots v_{1,s} + \cdots + v_{d,1}v_{d,2}\cdots v_{d,s})$ has the required format for applying Theorem 21.1 directly: $Z^{n-s} \boxdot (v_{1,1}v_{1,2}\cdots v_{1,s} + \cdots + v_{d,1}v_{d,2}\cdots v_{d,s}) = Z^{n-s}(\tilde{v}_{1,1}\tilde{v}_{1,2}\cdots\tilde{v}_{1,s} + \cdots + \tilde{v}_{d,1}\tilde{v}_{d,2}\cdots\tilde{v}_{d,s})$ for some $\tilde{v}_{i,j}$. ☐

We write $\lambda^{\sharp D}$ for the partition of the "lifted shape" $\lambda + (D - |\lambda|)$, that arises from λ by extending the first row so that $\lambda^{\sharp D}$ has D boxes.

Corollary 21.5. *If $a_\lambda(d, n) > 0$, then $a_{\lambda^{\sharp dN}}(d, N) > 0$ for all $N \geq n$.*

Proof. If $a_\lambda(d, n) > 0$, then there is some nonzero $\tilde{v}_T \in \mathsf{HWV}_\lambda(Sym^d Sym^n V)$ and therefore some $h \in Sym^n V$ such that $\tilde{v}_T(h) \neq 0$. We obtain $\tilde{v}_{T'}$ that satisfies $\tilde{v}_{T'}(X_1^{N-n} \boxdot h) = \tilde{v}_T(h) \neq 0$ according to (21.1). As $\tilde{v}_{T'}$ is a nonzero HWV of weight $\lambda^{\sharp dN}$, we get that $a_{\lambda^{\sharp dN}}(d, N) > 0$. ☐

$$\begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&1&2&2&2&3&3&3&3&4&4&4\\ \hline 2&2&2&2&4&4\\ \cline{1-6} 3&3&3\\ \cline{1-3} 4&4\\ \cline{1-2}\end{array} \xmapsto{\text{shortening}} \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|}\hline 1&1&1&1&1&1&2&2&3&3&3&4&4\\ \hline 2&2&2&2&4&4\\ \cline{1-6} 3&3&3\\ \cline{1-3} 4&4\\ \cline{1-2}\end{array}$$

Figure 10: Shortening the first row. Here $n = 7$, $d = 4$, $M = 6$, $\lambda_2 + |\bar{\lambda}| = 18 \leq 24 = Md$ $\lambda = (16, 6, 4, 2)$. Then $\mu = (12, 6, 4, 2)$. The crucial property is that for each number there is a singleton box containing that number in the left tableau.

We will use the following proposition with $M = md$.

Proposition 21.6. *Suppose that $\lambda \vdash nd$ satisfies $\lambda_2 \leq M$ and $\lambda_2 + |\bar{\lambda}| \leq Md$. Then every HWV of weight λ in $Sym^d Sym^n V$ is obtained by lifting a HWV in $Sym^d Sym^M V$ of weight μ, where $\mu \vdash Md$ such that $\bar{\mu} = \bar{\lambda}$.*

Proof. Note that $\lambda_2 + |\bar{\lambda}| \leq Md$ is the number of boxes of λ that appear in columns that are not singleton columns. We can therefore shorten the given λ to a partition $\mu \vdash Md$ by removing singleton columns, see Figure 10 for an example. Indeed, if T is semistandard of shape λ with content $d \times n$, then each number can appear in non-singleton columns at most M times, because $\lambda_2 \leq M$. Therefore shortening T to shape $\mu \vdash Md$ can be done by removing $(n - M)$ of each number $1, \ldots, d$ from singleton columns, which gives the HWV we searched for. $\square$

21.1.4 The degree lower bound

If we want to separate with polynomials, then the following proposition gives a lower bound on the possible degree.

Proposition 21.7. *Let $\lambda \vdash nd$ be such that there exists a positive integer m satisfying $|\bar{\lambda}| \leq md$ and $d \leq \sqrt{\frac{n}{m}}$. Then every nonzero HWV of weight λ in $Sym^d(Sym^n V)$ does not vanish on $\overline{\mathsf{GL}_{n^2}\det_n}$.*

In particular, to show superpolynomial lower bounds on the border determinantal complexity of per_m *we need superpolynomially high degree d.*

Proof. The case $d = 1$ is trivial as (n) occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]_1$. So suppose $d \geq 2$.

Let $F \in \mathsf{HWV}_\lambda(Sym^d Sym^n V)$. We have $\lambda_2 \leq |\bar{\lambda}| \leq md$ and $\lambda_2 + |\bar{\lambda}| \leq 2|\bar{\lambda}| \leq 2md \leq md \cdot d$. Therefore, we are in the setting of Proposition 21.6 with respect to the lifting $Sym^d Sym^{md} V \to Sym^d Sym^n V$. We conclude that F arises by an inner degree lifting from a HWV $f \in Sym^d Sym^{md} V$ of weight $\lambda - (d(n - m))$.

By Corollary 21.3, there are $v_{1,1}, \ldots, v_{d,md} \in V$ such that f does not vanish on

$$p := v_{1,1} v_{1,2} \cdots v_{1,md} + \cdots + v_{d,1} v_{d,2} \cdots v_{d,md}.$$

Using (21.1) we see that F does not vanish on $q := X_1^{n-m} \boxdot p$. By Corollary 21.4 we have $q \in \overline{\mathsf{GL}_{n^2}\det_n}$ since $n \geq md \cdot d$ (i. e., $d \leq \sqrt{\frac{n}{m}}$). Therefore, F does not vanish on $\overline{\mathsf{GL}_{n^2}\det_n}$. $\square$

21.1.5 A first row that is too long

In a mostly analogous way one can prove the following proposition that takes care of cases with a huge λ_1. We omit the details.

Proposition 21.8. *Let $\lambda \vdash nd$ and assume there exist positive integers s, m such that $\ell(\lambda) \le m^2$, $\lambda_1 \ge nd - s$, $m^2 s^2 \le n$, and $m^2 s \le d$. Then every nonzero $h \in \mathsf{HWV}_\lambda(Sym^d Sym^n V)$ of weight λ does not vanish on $\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}$.*

21.2 No occurrence obstructions

In this section we prove that occurrence obstructions cannot prove superpolynomial lower bounds on $\mathrm{dc}(\mathrm{per}_m)$.

Theorem 21.9. *Let n, d, m be positive integers with $n \ge m^{25}$ and $\lambda \vdash nd$. If λ occurs in $\mathbb{C}[\overline{Z^{n-m}\mathrm{per}_m}]$, then λ also occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]$. In particular occurrence obstructions cannot show superpolynomial lower bounds on $dc(\mathrm{per}_m)$.*

Proof. We may assume that $m \ge 2$, as the case $m = 1$ is trivial. Suppose that $\lambda \vdash nd$ occurs in $\mathbb{C}[\overline{Z^{n-m}\mathrm{per}_m}]$ and $n \ge m^{25}$. Proposition 20.5 implies that $|\bar{\lambda}| \le md$ and $\ell(\lambda) \le m^2$.

In the case of "small degree", where $n \ge md^2$, Proposition 21.7 implies that λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]$.

So we may assume that $d > \sqrt{n/m}$. In this case we have $d \ge \sqrt{m^{25}/m} = m^{12}$. We conclude by two further case distinctions.

If $|\bar{\lambda}| < m^{10}$, we can apply Proposition 21.8 with $s := m^{10}$ since $\lambda_2 \le |\bar{\lambda}| \le s$ (note that $\lambda_1 \ge nd - s$ iff $\lambda_2 \le s$), $m^2 s^2 = m^{22} \le n$, and $m^2 s = m^{12} \le d$. Thus λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]_d$.

We handled the case where the d is small or $|\lambda|$ is small. Finally, we come to the most interesting case. If $|\bar{\lambda}| \ge m^{10}$, then an explicit construction (Corollary 21.22) tells us that λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]_d$. □

The explicit construction mentioned in the proof crucially uses the so-called *semigroup property* that we introduce in the next section.

21.2.1 The semigroup property

To have an explicit construction of HWVs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]$ we use the *semigroup property*, which allows us to construct HWVs as products of HWVs of smaller degrees.

Let $\mathbb{A} := \mathbb{C}^N$. A Zariski-closed subset $Z \subseteq \mathbb{A}$ is called an *irreducible subvariety* if Z is not a union of two distinct proper Zariski-closed subsets of Z. More generally, in a topological space a subset is *irreducible* if it is not the union of two distinct proper closed subsets.

Lemma 21.10. *The affine space $\mathbb{A} := \mathbb{C}^N$ is irreducible.*

Proof. Let $V(f_1,\dots,f_r)$ denote the simultaneous vanishing set of $f_1,\dots,f_r$, where $f_i \in \mathbb{C}[\mathbb{A}]$. If $\mathbb{A}$ is reducible, then $\mathbb{A} = V(f_1,\dots,f_r) \cup V(g_1,\dots,g_s)$, where $f_i, g_j \in \mathbb{C}[\mathbb{A}]$, $f_i, g_j \neq 0$. In particular $\mathbb{A} \subseteq V(f_1) \cup V(g_1) = V(f_1 g_1)$. Since $f_1 g_1 \neq 0$ there exists a point $x \in \mathbb{A}$ with $(f_1 g_1)(x) \neq 0$ (Lemma 1.9). But this means that $\mathbb{A} \nsubseteq V(f_1 g_1)$, which is a contradiction. □

Lemma 21.11. *The image $f(X)$ of an irreducible set X under a continuous map f is irreducible.*

Proof. If $f(X) = Y_1 \cup Y_2$ with nontrivial distinct Zariski-closed subsets Y_i, then $X = f^{-1}(Y_1) \cup f^{-1}(Y_2)$. Since f is continuous, $f^{-1}(Y_i)$ is closed. Moreover, if $f^{-1}(Y_1) = X$, then $f(X) = Y_1$, in contradiction to $Y_1 \subsetneq f(X)$. Analogously it holds $f^{-1}(Y_2) \neq X$. Thus X is a union of nontrivial closed subsets, in contradiction to X being irreducible. □

Lemma 21.12. *The closure $\overline{Y}$ of every irreducible set Y is irreducible.*

Proof. Assume that $\overline{Y}$ is not irreducible, i. e., $\overline{Y} = S \cup T$ with closed subsets $S \subseteq \overline{Y}$ and $T \subseteq \overline{Y}$, and $S, T \neq \overline{Y}$. We have $Y \subseteq \overline{Y} \subseteq S \cup T$. Thus $Y = (S \cup T) \cap Y = (S \cap Y) \cup (T \cap Y)$, where both $(S \cap Y)$ and $(T \cap Y)$ are closed in Y (subspace topology). The decomposition is nontrivial: if $S \cap Y = Y$, then $Y \subseteq S$, and therefore $\overline{Y} \subseteq \overline{S} = S$ and thus $\overline{Y} = S$, a contradiction. Analogously for T. Therefore Y is not irreducible. □

Lemma 21.13. *The orbit closure $\overline{\mathsf{GL}_{n^2}\det_n} \subseteq \mathbb{C}[X_1,\dots,X_{n^2}]_n$ is an irreducible subvariety.*

Proof. By Lemma 21.10 the affine matrix space $\mathbb{C}^{n^2 \times n^2} = \mathsf{End}_{n^2}$ is irreducible. Hence the orbit $\mathsf{End}_{n^2}\det_n$ is irreducible by Lemma 21.11. Its closure is irreducible by Lemma 21.12. □

Remark 21.14. Lemma 21.13 holds in high generality: The orbit closure of any connected algebraic group under a polynomial group action is irreducible.

Claim 21.15. *In a domain (i. e., a ring without zero divisors) we have:*

$$\textit{If } ax = ay \textit{ with } a \neq 0\textit{, then } x = y.$$

Proof.

$$ax = ay \Rightarrow a(x-y) = 0 \Rightarrow a = 0 \text{ or } x = y.$$

□

Lemma 21.16. *For an irreducible subvariety Z the coordinate ring $\mathbb{C}[Z]$ has no zero divisors.*

Proof. Let f, g with $fg = 0$ in $\mathbb{C}[Z]$, i. e., $fg(z) = 0$ for all $z \in Z$. Since fg vanishes on Z, $Z = V(fg) \cap Z = (V(f) \cap Z) \cup (V(g) \cap Z)$. Since Z is irreducible, $V(f) \cap Z$ and $V(g) \cap Z$ cannot be both proper subsets of Z. Therefore either $V(f) \cap Z = Z$ or $V(g) \cap Z = Z$. W.l.o.g. $V(f) \cap Z = Z$, thus $f = 0$ in $\mathbb{C}[Z]$. □

Proposition 21.17 (The semigroup property)**.** *Let $G = \mathsf{GL}_m$ act polynomially on $\mathbb{A}$. Let the cone $Z \subseteq \mathbb{A}$ be an irreducible subvariety that is closed under the action of G. Then the coordinate ring of Z in degree d is a G-representation and we have the following:*

If the type λ occurs with positive multiplicity m_1 in $\mathbb{C}[Z]_{d_1}$ and the type μ occurs with positive multiplicity m_2 in $\mathbb{C}[Z]_{d_2}$, then the type $\lambda + \mu$ occurs with multiplicity at least $\max(m_1, m_2)$ in $\mathbb{C}[Z]_{d_1+d_2}$.

Proof. W.l.o.g. let $m_1 \le m_2$. Let f be a HWV of weight λ in $\mathbb{C}[Z]_{d_1}$. Let $F_1, \dots, F_{m_2}$ be a basis of HWVs of weight μ in $\mathbb{C}[Z]_{d_2}$. Then $fF_1, \cdots fF_{m_2}$ are linearly independent HWVs of weight $\lambda + \mu$ in $\mathbb{C}[Z]_{d_1+d_2}$, as can be seen as follows.

Assume a nontrivial linear combination of zero:

$$\alpha_1 fF_1 + \ldots + \alpha_{m_2} fF_{m_2} = 0$$

We conclude

$$f(\alpha_1 F_1 + \ldots + \alpha_{m_2} F_{m_2}) = f0.$$

Since $\mathbb{C}[Z]$ has no zero divisors, using Claim 21.15 it follows $\alpha_1 F_1 + \ldots + \alpha_{m_2} F_{m_2} = 0$, a contradiction to the linear independence of the F_i. ☐

Corollary 21.18. *Let $G = \mathsf{GL}_{n^2}$ and $v = \det_n$ or $v = Z^{n-m}\mathrm{per}_m$. If $\mathrm{mult}_\lambda(\mathbb{C}[\overline{Gv}])_{d_1} > 0$ and $\mathrm{mult}_\mu(\mathbb{C}[\overline{Gv}])_{d_2} > 0$, then*

$$\mathrm{mult}_{\lambda+\mu}(\mathbb{C}[\overline{Gv}])_{d_1+d_2} > \max(\mathrm{mult}_\lambda(\mathbb{C}[\overline{Gv}])_{d_1}, \mathrm{mult}_\mu(\mathbb{C}[\overline{Gv}])_{d_2}).$$

Moreover, if $a_\lambda(d_1, n) > 0$ and $a_\mu(d_2, n) > 0$, then $a_{\lambda+\mu}(d_1 + d_2, n) > \max(a_\lambda(d_1, n), a_\mu(d_2, n))$.

Proof. Both facts are direct corollaries of Prop. 21.17. ☐

21.2.2 Building blocks and the splitting technique

We construct as "building blocks" certain partitions that occur in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]$ and combine them with the semigroup property Corollary 21.18.

A first building block is the following.

Proposition 21.19. *Let $n \ge k\ell$ and ℓ be even. Then $(k \times \ell)^{\sharp nk}$ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]_k$.*

Proof. Let T denote the tableau of shape $k \times \ell$ with content $k \times \ell$ from Theorem 19.16. Suppose $n \ge k\ell$ and let $F \in Sym^k Sym^n V$ denote the lifting of $v_T \in Sym^k Sym^\ell V$. Hence F is a highest weight vector of weight $(k \times \ell)^{\sharp nk}$. Choose $p \in Sym^\ell V$ with at most k nonzero coefficients and the property that $v_T(p) \neq 0$ (Corollary 21.3). Applying (21.1), we obtain with $q := X_1^{n-\ell} \boxdot p$ that $\langle F, q^{\otimes k}\rangle = \langle v_T, p^{\otimes k}\rangle \neq 0$. Even if we rescale its coefficients, the determinantal complexity of p is less than $k\ell \le n$. Therefore F does not vanish on $\overline{\mathsf{GL}_{n^2}\det_n}$ and the assertion follows. ☐

We postpone the proof of the following technical result to Section 21.2.3. (It is based on an explicit construction of a highest weight vector.)

Theorem 21.20. *Let $2 \le b, c \le m^2$ and let $n \ge 24m^6$. Then there exists an even $i \le 2m^4$, such that*

$$\lambda = b \times 1 + c \times i + 1 \times j$$

occurs in $\mathbb{C}[\overline{\mathsf{GL}_n^2 \det_n}]_{3m^4}$ for $j = 3m^4 n - b - ic$.

The splitting strategy in the following proof is a refinement of the one in [47]. The proof relies on Theorem 21.20 and on the semigroup property (Corollary 21.18).

Proposition 21.21. *Given a partition λ with $|\lambda| = nd$ such that there exists $m \ge 2$ with $\ell(\lambda) \le m^2$, $m^{10} \le |\bar{\lambda}| \le md$, $n \ge 24m^6$, and $d > 4m^6$. Then λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2} \det_n}]_d$.*

Proof. Let $L := \ell(\lambda)$ and c_k denote the number of columns of length k in λ for $1 \le k \le L$. Let K be the index $k \ge 2$, for which c_k is maximal, i. e., $c_K = \max(c_k; k = 2, \dots, L)$. By assumption, we have $2 \le K \le m^2$ and

$$m^{10} \le |\bar{\lambda}| = \sum_{k=2}^{L} (k-1) c_k \le c_K \sum_{k=2}^{L} (k-1) \le c_K \frac{L^2}{2} \le c_K \frac{m^4}{2},$$

hence $c_K \ge 2m^6$.

Let S denote the set of integers $k \in \{2, \dots, L\}$ for which c_k is odd. For $k \in S$ we define the partition

$$\omega_k := k \times 1 + K \times i_k,$$

where the even integer $i_k \le 2m^4$ is taken from Theorem 21.20, so that $\omega_k^{\sharp 3nm^4}$ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2} \det_n}]_{3m^4}$. (Here we have used the assumption $n \ge 24m^6$.)

Assume first that $K \notin S$, that is, c_K is even. Then we can split λ vertically in rectangles as follows:

$$\begin{aligned}
\lambda &= 1 \times c_1 + \sum_{\substack{k=2 \\ k \notin S \cup \{K\}}}^{L} k \times c_k + \sum_{\substack{k=2 \\ k \in S}}^{L} k \times c_k + K \times c_K \\
&= 1 \times c_1 + \sum_{\substack{k=2 \\ k \notin S \cup \{K\}}}^{L} k \times c_k + \sum_{\substack{k=2 \\ k \in S}}^{L} k \times (c_k - 1) + \sum_{k \in S} \omega_k + K \times \Big(c_K - \sum_{k \in S} i_k\Big).
\end{aligned}$$

If, for $k \le L$, we set $d_k := c_k$ if $k \notin S \cup \{K\}$ and $d_k := c_k - 1$ if $k \in S$, and define $d_K := c_K - \sum_{k \in S} i_k$, then the above can be briefly written as

$$\lambda = 1 \times c_1 + \sum_{k=2}^{L} k \times d_k + \sum_{k \in S} \omega_k. \tag{21.2}$$

By construction, all d_k are even. It is crucial to note that, using $i_k \le 2m^4$,

$$d_K = c_K - \sum_{k \in S} i_k \ge c_K - (m^2 - 1) \cdot 2m^4 \ge c_K - 2m^6 \ge 0.$$

The last inequality is due to our observation at the beginning of the proof.

In the case where $K \in S$, we achieve the same decomposition as in (21.2) with the modified definition $d_K := c_K - 1 - \sum_{k \in S} i_k$. Here, as well $d_K \geq 0$ and all d_k are even.

We need to round down rational numbers to the next even number, so for $a \in \mathbb{Q}$ we define $\llbracket a \rrbracket := 2\lfloor a/2 \rfloor$. Note that $\llbracket a \rrbracket \geq a - 2$ for all $a \in \mathbb{Q}$. Hence $\llbracket n/k \rrbracket \geq n/k - 2 \geq 2$ for all $2 \leq k \leq m^2$, since $n \geq 4m^2$.

Using division with remainder, let us write $d_k = q_k \llbracket \frac{n}{k} \rrbracket + r_k$ with $0 \leq r_k < \llbracket \frac{n}{k} \rrbracket$. Then we split $k \times d_k = q_k(k \times \llbracket \frac{n}{k} \rrbracket) + k \times r_k$. Since d_k is even and $\llbracket n/k \rrbracket$ is even, r_k is even as well. From (21.2) we obtain that the partition

$$\mu := \sum_{k=2}^{L} q_k((k \times \llbracket n/k \rrbracket)^{\sharp nk}) + \sum_{k=2}^{L} (k \times r_k)^{\sharp nk} + \sum_{k \in S} \omega_k^{\sharp 3nm^4} \tag{21.3}$$

coincides with λ in all but possibly the first row.

Since $\llbracket n/k \rrbracket \leq n/k$, $r_k \leq n/k$, and both $\llbracket n/k \rrbracket$ and r_k are even, Proposition 21.19 implies that $(k \times \llbracket n/k \rrbracket)^{\sharp nk}$ and $(k \times r_k)^{\sharp nk}$ occur as highest weights in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]_k$. Moreover, Theorem 21.20 tells us that $\omega_k^{\sharp 3nm^4}$ occurs as a highest weight in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]_{3m^4}$. The semigroup property implies that μ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]$.

Claim. $|\mu| \leq dn$.

Let us finish the proof assuming the claim. If $|\mu| \leq dn$, we can obtain λ from μ by adding boxes to the first row of μ. Note that $|\lambda| - |\mu|$ is a multiple of n. Since $(n) \in \mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]$, the semigroup property implies that λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]_d$.

It remains to verify the claim. From (21.3) we get

$$|\mu| \;\leq\; \sum_{k=2}^{L} (q_k nk + nk + 3nm^4).$$

We have, using $\llbracket a \rrbracket \geq a - 2$,

$$q_k \leq \frac{d_k}{\llbracket n/k \rrbracket} \leq \frac{kd_k}{n - 2k}.$$

This implies

$$|\mu| \;\leq\; n \sum_{k=2}^{L} \Big(\frac{k^2 d_k}{n - 2k} + k + 3m^4 \Big).$$

Using $d_k \leq c_k$ and $L \leq m^2$, we get

$$|\mu| \;\leq\; n \sum_{k=2}^{L} \frac{m^2}{n - 2m^2} k c_k + n \sum_{k=2}^{m^2} k + 3nm^4(m^2 - 1).$$

Noting that $\sum_{k=2}^{L} kc_k = |\bar{\lambda}| + \lambda_2 \leq 2|\bar{\lambda}|$, we continue with

$$\begin{aligned} |\mu| &\leq \frac{nm^2}{n-2m^2} \cdot 2|\bar{\lambda}| + n\Big(\frac{m^2(m^2+1)}{2} + 3m^4(m^2-1)\Big) \\ &\leq \frac{nm^2}{12m^6 - m^2} \cdot |\bar{\lambda}| + n\Big(3m^6 - \frac{5}{2}m^4 + \frac{1}{2}m^2\Big), \end{aligned}$$

where we have used $n > 24m^6$ for the second inequality. Plugging in the assumptions $|\bar{\lambda}| \leq dm$ and $d > 4m^6$, we obtain

$$|\mu| \leq \frac{dnm^3}{11m^6} + 3nm^6 \leq \frac{dn}{11} + 3nm^6 \leq \frac{dn}{11} + \frac{3dn}{4} < dn,$$

which shows the claim and completes the proof. □

Corollary 21.22. *Let $m \geq 2$, $n \geq m^{25}$, $\lambda \vdash nd$, $|\bar{\lambda}| \leq md$, $\ell(\lambda) \leq m^2$, $d > \sqrt{n/m}$, $|\bar{\lambda}| \geq m^{10}$. Then λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]_d$.*

Proof. To apply Proposition 21.21 we need to ensure that $n \geq 24m^6$ and $d > 4m^6$. Indeed, $n \geq m^{25} \geq 32m^{20} \geq 24m^6$ and $d > \sqrt{n/m} \geq \sqrt{m^{24}} = m^{12} \geq 4m^{10} \geq 4m^6$. □

At this point, to finish the proof of Theorem 21.9 it just remains to prove Theorem 21.20.

21.2.3 Explicit constructions of tableaux and positivity of plethysms

The goal of this section is to provide the proof of Theorem 21.20, which finishes the proof of Theorem 21.9. We achieve this by a direct construction of a HWV of type λ. The first Proposition 21.23 treats a simple case, while Proposition 21.24 covers the full generalization. Since the degree is low enough, we can then use the methods from Section 21.1 to show that λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\mathrm{det}_n}]$.

Proposition 21.23. *Let $t \geq r$, $i \geq 2t+3$ be positive integers and let $n \geq i$ and $d \geq 2t+i+1$. Let $\nu = (t+1) \times i + (r+1) \times 1 + (j)$, where $j = dn - (t+1)i - (r+1)$. Then $a_\nu(d,n) > 0$.*

Proof. We may assume that $n = i$ and $d = 2t+i+1$ (see Lemma 21.18 for $d > 2t+i+1$ and Corollary 21.5 if $n > i$).

Let T be a tableau of shape ν labeled with the integers $1, 2, 3, \ldots, d$, each appearing n times, as explained in Figure 11 for the case $t = 5$, $r = 3$ and $i = 13$. Formally, if $1 \leq k \leq r$, the row $k+1$ of T has $i+1$ boxes: $k+1$ boxes are labeled k, and the remaining $i-k$ boxes are labeled $2t+1-k$. If $r < k \leq t$, then the row $k+1$ of T has i boxes: $k+1$ boxes are labeled k and the remaining $i-k-1$ boxes labeled $2t+1-k$. The first row of T starts with the first $i+1$ boxes labeled with $2t+1, \ldots, d = 2t+i+1$, respectively, and all the remaining j labels are put in the singleton columns of T such that each integer in $1, \ldots, d$ appears exactly n times. Note that each integer $1, \ldots, d$ appears in at least one singleton column, since $n \geq i \geq 2t+3$.

11	12	13	14	15	16	17	18	19	20	21	22	23	24	1	1	1	...
1	1	10	10	10	10	10	10	10	10	10	10	10	10				
2	2	2	9	9	9	9	9	9	9	9	9	9	9				
3	3	3	3	8	8	8	8	8	8	8	8	8	8				
4	4	4	4	4	7	7	7	7	7	7	7	7					
5	5	5	5	5	5	6	6	6	6	6	6	6					

Figure 11: Prop. 21.24: $t = 5$, $r = 3$, $i = 13$, $d = 24$, $n = 13$, $D = 10$, $dn = 312$, $j = 230$.

Put $D \coloneqq 2t$. By construction, for any $1 \le u \le D$ in T, u appears in row 1 and in a unique row $k_u + 1$ for some $1 \le k_u \le t$. Let $\beta(u)$ denote the number of occurrences of u in row $k_u + 1$. Note that $2 \le \beta(1) < \beta(2) < \ldots < \beta(D)$ by construction.

We consider now the tensor

$$\Phi \coloneqq \bigotimes_{u=1}^{D} \left(e_{k_u+1}^{\otimes\beta(u)} \otimes e_1^{\otimes(n-\beta(u))}\right) \otimes \bigotimes_{u=D+1}^{d} e_1^n,$$

which, more precisely, is defined by the map, $s\colon [dn] \to \mathbb{C}^N$,

$$s_{(u-1)n+v} = \begin{cases} e_{k_u+1} & \text{if } 1 \le u \le D \text{ and } 1 \le v \le \beta(u) \\ e_1 & \text{otherwise.} \end{cases}$$

Since Φ is of rank 1, the tensor contraction $\langle \tilde{v}_T, \Phi\rangle$ from Section 20.1 simplifies: Since the $\beta(u)$ are all distinct, the only nonzero summands in the expansion of $\langle \tilde{v}_T, \Phi\rangle$ satisfy $(s \circ \vartheta)(\square) = e_{\mathrm{row}(\square)}$. These summands have $\mathrm{val}_\vartheta(s) = 1$, which makes the overall contraction nonzero. $\square$

By generalizing this construction in the proof, we can show the following.

Proposition 21.24. *Let t, r be positive integers, $i \in [\frac{(r+2t)^2}{2t}, \frac{(r+2t)^2}{2t} + r + t + 1]$, and let $n > 6t + 2r$ and $d > r + 2t + i$. Let $\nu = (t+1) \times i + (r+1) \times 1 + (j)$, where $j = dn - (r+1) - (t+1)i$. Then $a_\nu(d[n]) > 0$.*

Proof. If $r < t$ then we can directly apply Proposition 21.23, noticing that

$$2t + 2 < \frac{(1+2t)^2}{2t} \le \frac{(r+2t)^2}{2t} \le i \le \frac{(t+2t)^2}{2t} + r + t + 1 \le \frac{11}{2}t + r + 1 \le 6t + r \le n.$$

Let now $r \ge t$. The proof is similar to the proof of Proposition 21.23, so we describe a more general construction which applies in the case $r < t$ as well. Define $e \coloneqq 2(\lfloor (r-1)/(2t)\rfloor + 1)$, so that $r \le te \le r + 2t - 1$ and e is even. Put

$$i' \coloneqq (te+1)\frac{e}{2} \;\le\; (r+2t)\frac{e}{2} \;\le\; (r+2t)(\lfloor \frac{r-1}{2t} \rfloor + 1) \;\le\; (r+2t)\frac{(r+2t-1)}{2t} \;\le\; i.$$

We will prove the statement for $i = i'$. When $i > i'$, the tableau construction below can be modified by increasing the number of appearances of the t largest labels by $i - i' \le r + t$ in the subtableau T' as defined below. By assumption, $n > 6t + 2r \ge te + 2$ and $d > r + 2t + i \ge te + i + 1$. Indeed, we will prove the statement for the more general case in which we do not require $n > 6t + 2r$ and $d > r + 2t + i$, but only $n \ge te + 2$ and $d \ge te + i + 1$. It suffices to prove the statement with $n = te + 2$ and $d = te + i + 1$.

Let T be a tableau of shape ν filled with the labels $1, 2, 3, \dots, d = te + i + 1$, each number appearing $n = te + 2$ times, as in Figure 12 for the case $t = 2, r = 8, e = 4, i = 18, n = 10, d = 27$.

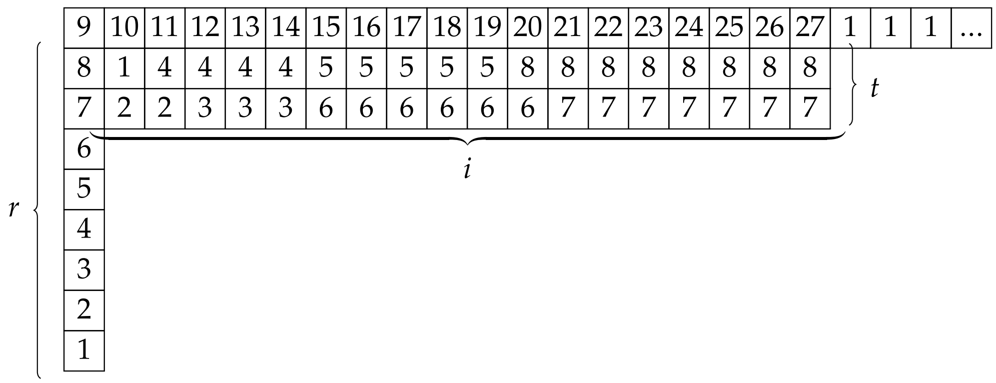

Figure 12: Prop. 21.24: $t = 2, r = 8, e = 4, i = 18, n = 10, d = 27$.

In the first row and in the first $i + 1$ colums we have the labels $te + 1, \dots, te + i + 1$. In the first column and in the rows 2 to $r + 1$ we have the labels $te, te - 1 \dots, te - r + 1$. The remaining rectangular $t \times i$ subshape of T, denoted T', consisting of the columns 2 to $i + 1$ and the rows 2 to $t + 1$, is filled with the remaining labels $1, \dots, te$, so that each label appears a different number of times. More precisely, for each $1 \le s \le te$, let the label s appear in T' exactly s times and only in row $\min(\ell, 2t - \ell + 1)$, where $s \equiv \ell \pmod{2t}$, $1 \le \ell \le 2t$. (Note that the first row in T', which we are referring to, is actually the second row in T.) So the row k of T' contains the e different labels $k, 2t + 1 - k, 2t + k, 4t + 1 - k, \dots, t(e - 2) + k, te + 1 - k$, each appearing that many times, adding up to the row length of

$$\sum_{\alpha=1}^{e/2} \Big((2(\alpha - 1)t + k) + (2\alpha t + 1 - k) \Big) = (te + 1)\frac{e}{2} = i.$$

The remaining labels of each kind are then put in the singleton boxes of T.

As in Proposition 21.23, we show that the corresponding highest-weight vector $\tilde{v}_T$ in $\mathsf{HWV}_\nu(Sym^d Sym^n V)$ is nonzero by contracting it with a particular monomial tensor Φ. For each label u, $1 \le u \le d$, let the associated monomial be

$$m_u = \bigotimes_{\square \in T,\ \mathrm{label}(\square) = u} X_{\mathrm{row}(\square)},$$

where the product goes over all boxes of T labeled u and for each such box we take the variable X whose index is the row the box is in. Again, let $\Phi := \otimes_{u=1}^{d} m_u$ be the tensor. The nonvanishing of the contraction $\langle v_T, \Phi\rangle$ can be seen analogously as in Proposition 21.23. □

Finally we can complete the proof of the promised technical result.

Proof of Theorem 21.20. We apply Proposition 21.24 with $r = b-1 \le m^2-1$ and $t = c-1 \le m^2-1$. We have

$$\frac{(r+2t)^2}{2t} = \frac{(b+2c-3)^2}{2(c-1)} \le \max\left(\frac{(b+1)^2}{2}, \frac{(b+2m^2-3)^2}{2(m^2-1)}\right) \le m^4,$$

where we use the fact that $(b+2c-3)^2/(2(c-1))$ is a convex function of c and so attains its maximum at the end points of the interval $[2, m^2]$. We can then find an even integer i in the interval $[\frac{(r+2t)^2}{2t}, \frac{(r+2t)^2}{2t} + r + t + 1] \subseteq [1, m^4 + 2m^2]$. By Proposition 21.24, there exists a highest weight vector f of weight $\nu = b \times 1 + c \times i + 1 \times j'$ in $Sym^d Sym^N V$ for

$$d := 3m^4 > 3m^2 + 2m^2 + m^4 \ge r + 2t + i, \quad N := 8m^2 > 6t + 2r.$$

By Proposition 21.3 we have $\langle f, h^d\rangle \ne 0$ for a generic polynomial $h \in Sym^N V$ that has at most d nonzero coefficients. Moreover, by Corollary 21.4, $q := X_1^{n-N} \boxdot p$ is contained in $\overline{\mathsf{GL}_{n^2}\det_n}$ for all $n \ge dN$, in particular for $n \ge 24m^6$. Consider the lifting $F \in Sym^d Sym^n V$ of f; it has the weight $\lambda = \nu^{\sharp dn}$ with $dn = 3m^4 n$. By (21.1) we see that $F(X_1^{n-m} \boxdot h) = f(h) \ne 0$. Therefore, λ occurs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]_{3m^4}$. □

> **No occurrence obstructions**
>
> By an explicit construction of HWVs in $\mathbb{C}[\overline{\mathsf{GL}_{n^2}\det_n}]$ we showed that occurrence obstructions are too weak to separate $Z^{n-m}\mathrm{per}_m$ from $\overline{\mathsf{GL}_{n^2}\det_n}$ for n being superpolynomially large in m.
> The proof relied heavily on the padding of the permanent (i. e., the partition having a very long first part). Other algebraic computational models do not have that property. Multiplicity obstructions could also still be an option to prove Valiant's conjecture.

22 Occurrence obstructions for matrix multiplication

In this last chapter we show that obstructions can indeed show complexity lower bounds. In most parts we follow [20], [21], and [46].

We will construct occurrence obstructions that show the border rank lower bound

$$\underline{R}(\langle m, m, m\rangle) \ge \tfrac{3}{2}m^2 - 2.$$

Recall from the end of Chapter 14 that

$$\underline{R}(\langle m, m, m\rangle) \le n \quad \text{iff} \quad \langle m, m, m\rangle \in \overline{\mathsf{GL}_n^3\langle n\rangle},$$

which is equivalent to $\overline{\mathsf{GL}_n^3\langle m,m,m\rangle} \subseteq \overline{\mathsf{GL}_n^3\langle n\rangle}$ (see Lemma 3.16). By Proposition 18.7 the irreducible polynomial representations of GL_n^3 are given by triples $\lambda = (\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)})$ of partitions into at most n parts each. For odd m and $n \geq \frac{3}{2}m^2 - 1$ let $\kappa := \frac{m^2-1}{2}$ and set $\lambda^{(1)} = \lambda^{(2)} = \lambda^{(3)} = ((2\kappa+1)\times 1) + (1 \times \kappa)$ and $d = |\lambda| = 3\kappa + 1$. We will show that

$$\begin{aligned} \mathrm{mult}_\lambda(\mathbb{C}[\overline{\mathsf{GL}_n^3\langle n\rangle}]_d) &= 0 & (22.1)\\ &< 1 = \mathrm{mult}_\lambda(\mathbb{C}[\overline{\mathsf{GL}_n^3\langle m,m,m\rangle}]_d) & (22.2)\end{aligned}$$

The irreducible representation in the coordinate ring of the ambient space is unique:

$$\mathrm{mult}_\lambda(\mathbb{C}[\textstyle\bigotimes^3\mathbb{C}^n]) = 1\,,$$

which is proved in [69, Thm. 2.1], see also [70, Thm. 3(4.)]. We construct the HWV of weight λ and evaluate it at a point in $\overline{\mathsf{GL}_n^3\langle m,m,m\rangle}$ where the evaluation is nonzero, which proves (22.2), see Section 22.3. For (22.1) we present two proofs. In the first one we examine the HWV that we constructed (see Section 22.2), while in the second one (see Section 22.4) we use the approach presented in (18.4), i. e., we study

$$\mathrm{mult}_{\lambda,\lambda',\lambda''}(\mathbb{C}[\overline{\mathsf{GL}_n^3\langle n\rangle}]) \leq \mathrm{mult}_{\lambda,\lambda',\lambda''}(\mathbb{C}[\mathsf{GL}_n^3\langle n\rangle]) = \dim(\{\lambda\}\otimes\{\lambda'\}\otimes\{\lambda''\})^{\mathsf{stab}\langle n\rangle}. \quad (22.3)$$

After developing the general theory in Section 22.1 and the theory tailored to equations for low border rank tensors in Section 22.2, the triple hook partitions λ that we use here can be found in small cases by hand. Proving that these specific HWVs vanish on $\mathsf{GL}_n^3\langle n\rangle$ then follows from the general theory with a quick argument, see Section 22.2. One main problem with this approach is to prove that the HWV that we construct is not the zero function. We do this be evaluating at a point in the orbit of the matrix multiplication tensor. This point is carefully chosen so that we can completely understand the evaluation of the HWV at this point and verify that it is nonzero, see Section 22.3. There are many points that one could use here. Our analysis of the nonvanishing of the evaluation is quite long, but we are not aware of a point for which this part would be significantly shorter (and the evaluation still nonzero).

Remark 22.1. [57] proved $\underline{R}(\langle m,m,m\rangle) \geq 2m^2 - m$. [56] improved this to $\underline{R}(\langle m,m,m\rangle) \geq 2m^2 - \log_2 m - 1$. Barriers to these proof methods were discovered in [32, 37]. There are no known barriers for using occurrence obstructions or multiplicity obstructions for the border rank of matrix multiplication. [39] use representation theory in a different way to prove $\underline{R}(\langle m,m,m\rangle) \geq \frac{5}{4}m^2$.

22.1 Highest weight vectors in the tensor setting

In this section we describe the HWVs in $\bigotimes^d(\bigotimes^3\mathbb{C}^n)$, because each HWV in $Sym^d(\bigotimes^3\mathbb{C}^n)$ arises from one in $\bigotimes^d(\bigotimes^3\mathbb{C}^n)$ via symmetrization over $\mathfrak{S}_d$. This is analogous to studying semistandard tableaux with rectangular content in Section 19.3. The natural equivalent to semistandard tableaux with rectangular content are triples of standard tableaux, but we want to use a more graphical description via hypergraphs.

22.1.1 Highest Weight Vectors

For a partition $\lambda \vdash d$ recall the definition of v_λ from (19.1). We want to use the convenient bra-ket notation

$$\langle\widehat{\lambda}| := v_\lambda. \tag{22.4}$$

We write $\lambda \overset{*}{\vdash} d$ to denote that $\lambda = (\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)})$ with $\lambda^{(1)} \vdash d$, $\lambda^{(2)} \vdash d$, and $\lambda^{(3)} \vdash d$. For a partition triple $\lambda \overset{*}{\vdash}_n d$ we define

$$\langle\widehat{\lambda}| := \mathsf{reorder}_{3,n}\left(\langle\widehat{\lambda^{(1)}}| \otimes \langle\widehat{\lambda^{(2)}}| \otimes \langle\widehat{\lambda^{(3)}}|\right), \tag{22.5}$$

where for $a, b \in \mathbb{N}$ the linear isomorphism $\mathsf{reorder}_{a,b} \colon \bigotimes^a \bigotimes^b \mathbb{C}^n \to \bigotimes^b \bigotimes^a \mathbb{C}^n$ is defined on rank 1 tensors as follows:

$$\bigotimes_{i=1}^{a}\left(\bigotimes_{j=1}^{b} v_{ij}\right) \mapsto \bigotimes_{j=1}^{b}\left(\bigotimes_{i=1}^{a} v_{ij}\right), \quad v_{ij} \in \mathbb{C}^n. \tag{22.6}$$

Since $\langle\widehat{\lambda^{(i)}}|$ is a HWV of weight $\lambda^{(i)}$ in $\bigotimes^d \mathbb{C}^n$, $\langle\widehat{\lambda}|$ is a highest weight vector of weight λ in $\bigotimes^d \bigotimes^3 \mathbb{C}^n$. Moreover, since the $\langle\widehat{\lambda^{(i)}}|\pi$, $\pi \in \mathfrak{S}_d$, generate $\mathsf{HWV}_{\lambda^{(i)}}(\bigotimes^d \mathbb{C}^n)$ as a vector space (cf. Proposition 19.3), we also see following:

Claim 22.2. *The highest weight vector space* $\mathsf{HWV}_\lambda(\bigotimes^d \bigotimes^3 \mathbb{C}^n)$ *is generated by* $\langle\widehat{\lambda}|\pi$ *with* $\pi \in \mathfrak{S}_d^3$.

Embed $\mathfrak{S}_d \hookrightarrow \mathfrak{S}_d^3$, $\pi \mapsto (\pi, \pi, \pi)$. Let $\mathcal{P}_d : \bigotimes^d \bigotimes^3 \mathbb{C}^n \to Sym^d \bigotimes^3 \mathbb{C}^n$ denote the symmetrization over $\mathfrak{S}_d$. Since the actions of GL_n^3 and $\mathfrak{S}_d^3$ commute, we draw the following important conclusion.

Claim 22.3. *The tensors* $\langle\widehat{\lambda}|\pi\mathcal{P}_d$ *with* $\pi \in \mathfrak{S}_d^3$ *span* $\mathsf{HWV}_\lambda(Sym^d \bigotimes^3 \mathbb{C}^n)$.

These tensors span the HWV space, but they are rarely a basis, because the symmetrization operator $\mathcal{P}_d$ has a nonzero kernel.

22.1.2 Set Partitions

In this subsection we start deriving a more graphical interpretation of the highest weight vectors. Let $\wp(S)$ denote the powerset of a finite set S, i. e., the set of all subsets of S. Given a set S, we call a subset $\Lambda \subseteq \wp(S)$ of the powerset $\wp(S)$ a *set partition of* S, if for all $s \in S$ there exists exactly one set $e_s \in \Lambda$ with $s \in e_s$. We call e_s the *hyperedge corresponding to* s. If $|e_s| = 1$, then e_s is called a *singleton hyperedge*. The *type* of a set partition Λ is defined as the partition $\mu \vdash |S|$ obtained from sorting the multiset $\{|e| : e \in \Lambda\}$ and transposing the partition afterwards. Let $V(\Lambda) := \bigcup_{e \in \Lambda} e = S$ denote the ground set.

For a given partition $\lambda \vdash d$, we can define a canonical set partition Λ of $\{1, \dots, d\}$ as follows, where μ is the transpose of λ: Let $\omega_i := \sum_{j=1}^{i} |\mu_j|$ be the number of boxes in the first i columns of λ. We define the disjoint hyperedges

$$e_i := \{\omega_{i-1} + 1, \omega_{i-1} + 2, \dots, \omega_i\}$$

and set $\Lambda := \{e_i \mid 1 \leq i \leq \lambda_1\}$. For example, if $\lambda = (4,3,1)$, then $\mu = (3,2,2,1)$ and $e_1 = \{1,2,3\}$, $e_2 = \{4,5\}$, $e_3 = \{6,7\}$, $e_4 = \{8\}$, corresponding to the superstandard tableau $\mathcal{T}_\lambda$ (see Section 19.2) of shape λ, see Figure 13.

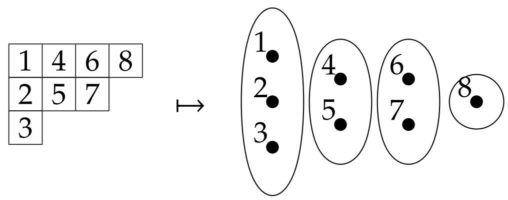

Figure 13: The superstandard tableau $\mathcal{T}_\lambda$ of shape $\lambda = (4,3,1)$ and its set partition.

Analogously, from any standard Young tableau T of λ we obtain a set partition Λ_T of $\{1,\ldots,d\}$ with type λ by grouping together each column in a hyperedge.

The map $T \mapsto \Lambda_T$ is not injective in general. Our aim is to classify the fibers. Two Young tableaux T and T' are mapped to the same $\Lambda_T = \Lambda_{T'}$, iff T' can be obtained from T by permuting entries inside of columns and by permuting whole columns of the same length. This observation gives rise to the following definition (see the right hand side of Figure 14 for an example).

Definition 22.4. An *ordered set partition* Λ of a vertex set $V(\Lambda) := \{1,\ldots,d\}$ of type λ is a set partition of $V(\Lambda)$ of type λ endowed with (1) linear orderings on each hyperedge $e \in \Lambda$ and (2) for each length $1 \leq \ell \leq \ell(\lambda)$ a linear ordering on the set $\{e \in \Lambda : |e| = \ell\}$ of hyperedges with the same cardinality ℓ.

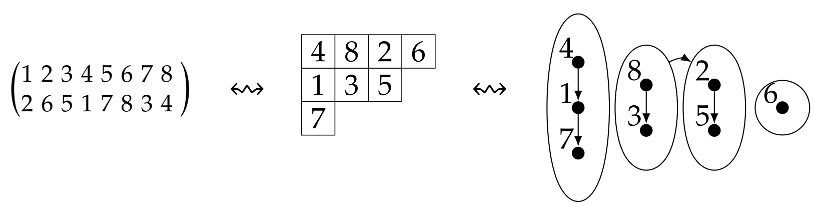

Figure 14: The bijections between permutations, Young tableaux, and ordered set partitions of type $\lambda = (4,3,1)$. The orderings are shown with arrows pointing from the smaller element to the bigger.

The above discussion gives an explicit bijection between the set of Young tableaux T of λ and the set of ordered set partitions of type λ by grouping together each column of T in one hyperedge, ordered from top to bottom, and ordering hyperedges of equal length by their appearance in T from left to right, see Figure 14 for an example. The columns of T are ordered from left to right and this induces an additional linear ordering on the set of hyperedges of an ordered set partition, which is consistent with the single linear orderings of hyperedges of the same length. In particular we can speak of the ith hyperedge of Λ. Additionally, since the hyperedges are linearly ordered, we have a linear order on the set of vertices $V(\Lambda)$ and hence we can write $V(\Lambda)_i$ for the ith element of $V(\Lambda)$.

The permutations $\pi \in \mathfrak{S}_d$ are in bijection to the Young tableaux of type λ via replacing the entry i in $\mathcal{T}_\lambda$ with the integer $\pi^{-1}(i)$. Hence we get an *explicit bijection* between $\mathfrak{S}_d$ and the set of ordered set partitions of type λ. For a given ordered set partition Λ of type λ we denote by π_Λ the corresponding permutation. We can state a first observation:

$$\pi_\Lambda(V(\Lambda)_i) = i. \tag{22.7}$$

The crucial property of our bijection is shown in the upcoming Claim 22.5, for whose statement we introduce some notation.

The group $\mathfrak{S}_d$ acts naturally on $(\mathbb{C}^n)^d$ by permuting the positions as follows:

$$\pi(\zeta_1, \zeta_2, \dots, \zeta_d) := (\zeta_{\pi^{-1}(1)}, \zeta_{\pi^{-1}(2)}, \dots, \zeta_{\pi^{-1}(d)}).$$

Given a linearly ordered subset $(e, \prec) \subseteq \{1, \dots, d\}$ with ℓ elements, $\ell \le d$, where the order $\prec$ is not necessarily consistent with the natural order on $\{1, \dots, d\}$, we define the list elements $e^1, \dots, e^\ell$ via $e = \{e^1, \dots, e^\ell\}$ satisfying $e^1 \prec \dots \prec e^\ell$. For example, for the leftmost hyperedge e in Figure 14 we have $e^1 = 4$, $e^2 = 1$, and $e^3 = 7$. Given a vector $\zeta \in (\mathbb{C}^n)^d$, we define the *restriction* $\zeta_{|_e}$ of ζ to e as

$$(\zeta_1, \zeta_2, \dots, \zeta_d)_{|_e} := (\zeta_{e^1}, \dots, \zeta_{e^\ell}).$$

Claim 22.5. *Fix $\lambda \vdash_n d$ and let e_i denote the ith column of $\mathcal{T}_\lambda$. Let Λ be an ordered set partition. Then, for the ith hyperedge e of Λ, we have*

$$(\zeta_1, \zeta_2, \dots, \zeta_d)_{|_e} = \big(\pi_\Lambda(\zeta_1, \zeta_2, \dots, \zeta_d)\big)_{|_{e_i}}$$

for all $\zeta \in (\mathbb{C}^n)^d$.

Proof. Note that $\pi_\Lambda(e^j) = (e_i)^j$ according to (22.7). Now the proof is straightforward as follows:

$$\begin{aligned}(\zeta_1, \dots, \zeta_d)_{|_e} &= (\zeta_{e^1}, \dots, \zeta_{e^d}) = (\zeta_{\pi_\Lambda^{-1}((e_i)^1)}, \dots, \zeta_{\pi_\Lambda^{-1}((e_i)^{|e_i|})}) \\ &= (\pi_\Lambda(\zeta_1, \dots, \zeta_d))_{|_{e_i}} \end{aligned}$$ □

From set partition triples to highest weight vectors Our main motivation for looking at set partitions is the construction of highest weight vectors. For each ordered set partition Λ of type $\lambda \vdash_n d$ we have $\pi_\Lambda \in \mathfrak{S}_d$ and hence obtain a nonzero highest weight vector

$$f_\Lambda := \pi_\Lambda^t(\langle\widehat{\lambda}|)$$

of weight λ, provided $n \ge \ell(\lambda)$. We conveniently write $f_\Lambda = \langle\widehat{\lambda}|\pi_\Lambda$. This roughly corresponds to v_T in Chapter 20.

We next want to determine the *projective stabilizer* $Y_\lambda \subseteq \mathfrak{S}_d$ of $\langle\widehat{\lambda}|$, which is defined as

$$Y_\lambda := \{\tau \in \mathfrak{S}_d : \langle\widehat{\lambda}|\tau = \pm\langle\widehat{\lambda}|\}. \tag{22.8}$$

Consider the Young subgroup $Y_\lambda^{\text{inner}} := \mathfrak{S}(e_1) \times \cdots \times \mathfrak{S}(e_{\lambda_1})$, where we recall that e_i denotes the ith column of $\mathcal{T}_\lambda$. For $\tau \in Y_\lambda^{\text{inner}}$ we have $\langle\widehat{\lambda}|\tau = \text{sgn}(\tau)\langle\widehat{\lambda}|$, hence $Y_\lambda^{\text{inner}} \subseteq Y_\lambda$. Let Y^{outer} denote the group that interchanges columns of the same length in $\mathcal{T}_\lambda$ while preserving the order in each column. For $\tau \in Y_\lambda^{\text{outer}}$ we have $\langle\widehat{\lambda}|\tau = \langle\widehat{\lambda}|$, hence $Y_\lambda^{\text{outer}} \subseteq Y_\lambda$. One can prove that the projective stabilizer Y_λ is the group generated by Y_λ^{inner} and Y_λ^{outer}.

We are interested in classifying the left cosets of $Y_\lambda \subseteq \mathfrak{S}_d$. The ordered set partition corresponding to π and the ordered set partition corresponding to $\tau\pi$ for $\tau \in Y_\lambda^{\text{inner}}$ are the same up to reordering the elements in each hyperedge. For $\tau \in Y_\lambda^{\text{outer}}$ the ordered set partitions corresponding to π and $\tau\pi$ are the same up to reordering the hyperedges. All reorderings can be obtained by applying elements of Y_λ. If we forget about the orderings of ordered set partitions, we obtain the following claim.

Claim 22.6. *For a fixed partition $\lambda \vdash d$ there is a bijection between the left cosets of $Y_\lambda \subseteq \mathfrak{S}_d$ and the set of set partitions of type λ.*

Hence a set partition Λ of type $\lambda \vdash_n d$ uniquely determines a highest weight vector of weight λ

$$f_\Lambda := \pm\langle\widehat{\lambda}|\pi_\Lambda \in \bigotimes^d \mathbb{C}^n$$

up to a sign, where π_Λ is the permutation corresponding to some ordering of Λ.

Contraction We proceed as in Chapter 20. A finite sequence $\zeta = (\zeta_1, \dots, \zeta_d)$ of vectors ζ_i in $\mathbb{C}^n$ is called a *list*. A map whose domain is a vertex set is sometimes called a *labeling* of the vertex set. If a vertex set e is linearly ordered, then we can identify lists and labelings with codomain $\mathbb{C}^n$. Given a list $\zeta = (\zeta_1, \dots, \zeta_d)$, we write $|\zeta\rangle := |\zeta_1\rangle \otimes \cdots \otimes |\zeta_d\rangle$. We want to analyze how the scalar product $\langle\widehat{\lambda}|\pi|\zeta\rangle$, for $\pi \in \mathfrak{S}_d$ and $|\zeta\rangle \in \bigotimes^d \mathbb{C}^n$, can be interpreted combinatorially using set partitions.

For a fixed $\pi \in \mathfrak{S}_d$ and a list ζ we define $\tilde{\zeta} := \pi\zeta$ to obtain

$$\langle\widehat{\lambda}|\pi|\zeta\rangle = \langle\widehat{\lambda}|\tilde{\zeta}\rangle = \prod_{i=1}^{\lambda_1} \langle\widehat{{}^t\lambda_i}|\tilde{\zeta}_{|e_i}\rangle.$$

Note that for $\ell \leq n$ and a list $\tilde{\zeta} = (\tilde{\zeta}_1, \dots, \tilde{\zeta}_\ell)$ with $\tilde{\zeta}_i \in \mathbb{C}^n$ $\ell \times \ell$-matrix $\big(\langle i|\tilde{\zeta}_j\rangle\big)_{i,j}$, see also (20.1).

Now fix an ordered set partition Λ. Given a hyperedge $e \in \Lambda$ and a hyperedge labeling $\zeta^e : e \to \mathbb{C}^n$, we can interpret ζ^e as a list (since e is linearly ordered) and write $|\zeta^e\rangle$. We define the *evaluation*

$$\text{val}_e(\zeta^e) := \langle\widehat{\ell}|\zeta^e\rangle \in \mathbb{C}.$$

Note that the evaluation $\mathrm{val}_e(\zeta^e)$ is, up to sign, invariant under changing the linear order of e. For a labeling $\zeta\colon V(\Lambda) \to \mathbb{C}^n$ we define the *evaluation of the ordered set partition Λ at the labeling ζ* by

$$\mathrm{val}_\Lambda(\zeta) := \prod_{e\in\Lambda} \mathrm{val}_e(\zeta_{|_e}).$$

Proposition 22.7. *Let Λ be an ordered set partition. Let $\zeta\colon V(\Lambda) \to \mathbb{C}^n$ be a labeling. We have*

$$\mathrm{val}_\Lambda(\zeta) = \langle\widehat{\lambda}|\pi_\Lambda|\zeta\rangle.$$

Proof. According to Claim 22.5, for the ith hyperedge e of Λ we have

$$\zeta_{|_e} = (\pi_\Lambda\zeta)_{|_{e_i}}.$$

Therefore, if e has size ℓ, then

$$\mathrm{val}_e(\zeta_{|_e}) = \langle\widehat{\ell}|\zeta_{|_e}\rangle = \langle\widehat{\ell}|(\pi_\Lambda\zeta)_{|_{e_i}}\rangle.$$

The claim follows by definition of $\langle\widehat{\lambda}|$ in (22.4). ☐

22.1.3 Obstruction Designs

We want to describe the highest weight vectors of $\bigotimes^d\bigotimes^3\mathbb{C}^n$ with set partitions as we did for $\bigotimes^d\mathbb{C}^n$. For this we make the following definition, analogously to Definition 22.4.

Definition 22.8. An *ordered set partition triple* $\mathcal{H}$ consists of a vertex set $V(\mathcal{H}) = \{1,\dots,d\}$ and three ordered set partitions $E^{(k)} := E^{(k)}(\mathcal{H})$, $k \in \{1,2,3\}$, of $V(\mathcal{H})$. The elements of each $E^{(k)}$ are called *hyperedges*.

The ordered set partition triple $\mathcal{H}$ is said to have *type* λ, where $\lambda \vdash^{\!*} d$ is a partition triple, if the set partition $E^{(k)}$ has type $\lambda^{(k)}$ for all $1 \le k \le 3$.

Via our explicit bijections between $\mathfrak{S}_d$ and the set of set partitions of a fixed type, we get an explicit bijection between $\mathfrak{S}_d^3$ and the set of ordered set partition triples of type λ, see Figure 15. The permutation triple corresponding to an ordered set partition $\mathcal{H}$ is denoted by $\pi_{\mathcal{H}}$.

Analogously to (22.8), for partition triples $\lambda \vdash^{\!*} d$ we define the *projective stabilizer* Y_λ of $\langle\widehat{\lambda}| = \mathsf{reorder}_{3,n}\big(\langle\widehat{\lambda^{(1)}}| \otimes \langle\widehat{\lambda^{(2)}}| \otimes \langle\widehat{\lambda^{(3)}}|\big)$, see (22.5), as

$$Y_\lambda := \{\tau \in \mathfrak{S}_d^3 : \langle\widehat{\lambda}|\tau = \pm\langle\widehat{\lambda}|\}.$$

One can show that $Y_\lambda = Y_{\lambda^{(1)}} \times Y_{\lambda^{(2)}} \times Y_{\lambda^{(3)}}$, where $Y_{\lambda^{(k)}}$ is the projective stabilizer defined in (22.8). Therefore we can again forget about the orderings and arrive at the following definition.

Definition 22.9. A *set partition triple* $\mathcal{H}$ consists of a vertex set $V(\mathcal{H}) = \{1,\dots,d\}$ and three set partitions $E^{(k)}$, $k \in \{1,2,3\}$ of $V(\mathcal{H})$.

The above discussion implies the following claim, analogously to Claim 22.6.

$$\begin{matrix} \sigma^{(1)} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \\ \sigma^{(2)} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix} \\ \sigma^{(3)} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix} \end{matrix} \quad \leftrightsquigarrow$$

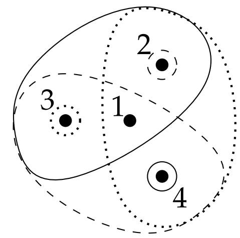

Figure 15: A set partition triple $\mathcal{H}$ of type $(\lambda, \lambda, \lambda)$, where $\lambda = (2, 1, 1)$ (and thus $\mu = (3, 1)$), and its corresponding permutation triple $\pi_{\mathcal{H}}$ with inverse σ. The solid lines represent the first hyperedge set partition, the dashed lines represent the second one, and the dotted lines represent the third one. To simplify the picture, the hyperedge orderings respect the natural ordering on the natural numbers and are not depicted.

Claim 22.10. *For a fixed partition triple $\lambda \vdash^{*} d$ there is a bijection between the left cosets of $Y_\lambda \subseteq \mathfrak{S}_d$ and the set of set partition triples of type λ.*

So each set partition triple $\mathcal{H}$ defines (up to sign) the highest weight vector

$$f_{\mathcal{H}} := \pm \langle \widehat{\lambda} | \pi_{\mathcal{H}} \in \bigotimes^d \bigotimes^3 \mathbb{C}^n$$

of weight λ^*, where $\lambda \vdash^{*}_{n} d$ denotes the type of $\mathcal{H}$.

Triple Contraction A finite sequence of vectors in $(\mathbb{C}^n)^3$ shall be called a *triple list*. Given a triple list ζ containing d triples, we write

$$\zeta = \begin{pmatrix} \zeta_1^{(1)}, \dots, \zeta_d^{(1)} \\ \zeta_1^{(2)}, \dots, \zeta_d^{(2)} \\ \zeta_1^{(3)}, \dots, \zeta_d^{(3)} \end{pmatrix}.$$

Moreover, we write $\zeta^{(k)} := (\zeta_1^{(k)}, \dots, \zeta_d^{(k)})$ and $\zeta_i := (\zeta_i^{(1)}, \zeta_i^{(2)}, \zeta_i^{(3)})$, and we write $|\zeta\rangle := \mathsf{reorder}_{3,d}(|\zeta^{(1)}\rangle \otimes |\zeta^{(2)}\rangle \otimes |\zeta^{(3)}\rangle) \in \bigotimes^d \bigotimes^3 \mathbb{C}^n$, where $\mathsf{reorder}_{3,d}$ is the linear map defined in (22.6). We want to analyze how the tensor contraction $\langle \widehat{\lambda} | \pi | \zeta \rangle$ can be interpreted combinatorially using set partitions. For an ordered subset $e \subseteq V(\mathcal{H})$ of vertices we identiy triple lists $(\zeta_1, \dots, \zeta_{|e|})$ with labelings on e whose codomain is $(\mathbb{C}^n)^3$.

We define the evaluation function for ordered set partition triples as follows: Given a labeling $\zeta \colon V(\mathcal{H}) \to (\mathbb{C}^n)^3$, we set

$$\mathrm{val}_{\mathcal{H}}(\zeta) := \mathrm{val}_{E^{(1)}(\mathcal{H})}(\zeta^{(1)}) \cdot \mathrm{val}_{E^{(2)}(\mathcal{H})}(\zeta^{(2)}) \cdot \mathrm{val}_{E^{(3)}(\mathcal{H})}(\zeta^{(3)}).$$

Proposition 22.11. *Let $\mathcal{H}$ be an ordered set partition triple of type λ. Let $\zeta \colon V(\mathcal{H}) \to (\mathbb{C}^n)^3$ be a labeling. We have*

$$\mathrm{val}_{\mathcal{H}}(\zeta) = \langle \widehat{\lambda} | \pi_{\mathcal{H}} | \zeta \rangle.$$

Proof. By (22.5) we have $\langle\widehat{\lambda}| = \mathsf{reorder}_{3,d}\big(\langle\widehat{\lambda}^{(1)}| \otimes \langle\widehat{\lambda}^{(2)}| \otimes \langle\widehat{\lambda}^{(3)}|\big)$. The claim follows with Proposition 22.7. □

Symmetrization Let $\mathcal{P}_d \colon \bigotimes^d \mathbb{C}^n \twoheadrightarrow Sym^d\mathbb{C}^n$ denote the symmetrization over $\mathfrak{S}_d$. Since $\tau\mathcal{P}_d = \mathcal{P}_d$ for all $\tau \in \mathfrak{S}_d$, we get $\langle\widehat{\lambda}|\pi\mathcal{P}_d = \langle\widehat{\lambda}|\pi\tau\mathcal{P}_d$ for all $\tau \in \mathfrak{S}_d$. Hence the polynomial described by a set partition triple is independent of the numbering of its vertices. This explains the following definition.

Definition 22.12. An *obstruction predesign* is defined to be an equivalence class of set partition triples under renumbering of the vertices. When depicting obstruction predesigns, we omit the vertex numbering of the corresponding set partition triple.

So each obstruction predesign describes some polynomial $\langle\widehat{\lambda}|\pi\mathcal{P}_d \in Sym^d \bigotimes^3 \mathbb{C}^n$ of degree d up to sign. Since we do not care about the sign, we abuse notation in the following way: For every obstruction predesign $\mathcal{H}$ we implicity fix an ordered set partition triple $\mathcal{H}'$ in a way such that $\mathcal{H}$ is obtained from $\mathcal{H}'$ by forgetting about orderings and vertex numbers. Then we define $\mathrm{val}_{\mathcal{H}}(\zeta) := \mathrm{val}_{\mathcal{H}'}(\zeta)$.

Corollary 22.13. *Let $\mathcal{H}$ be an ordered set partition triple with d vertices. Let $\xi \colon V(\mathcal{H}) \to (\mathbb{C}^n)^3$ be a labeling. We have*

$$\frac{1}{d!}\sum_{\zeta \in \mathfrak{S}_d \xi} \mathrm{val}_{\mathcal{H}}(\zeta) = \langle\widehat{\lambda}|\pi_{\mathcal{H}}\mathcal{P}_d|\xi\rangle.$$

Proof. Follows from Proposition 22.11 and the definition of the symmetrization $\mathcal{P}_d$. □

It is straightforward to verify (see, e. g., [46, Lemma 7.2.7]) that the polynomials described by obstructions predesigns are the zero function if they are not *obstruction designs*:

Definition 22.14. An *obstruction design* $\mathcal{H}$ is an obstruction predesign $\mathcal{H}$ which satisfies

$$|e_1 \cap e_2 \cap e_3| \le 1 \text{ for all hyperedge triples } (e_1, e_2, e_3) \in E^{(1)} \times E^{(2)} \times E^{(3)}.$$

Proposition 22.15. *For a partition triple λ we have*

$$\mathsf{HWV}_\lambda(Sym^d \textstyle\bigotimes^3 \mathbb{C}^n) = \mathit{span}\{f_{\mathcal{H}} : \mathcal{H} \text{ is an obstruction design of type } \lambda\}.$$

In particular

$$k(\lambda) = \dim \mathit{span}\{f_{\mathcal{H}} : \mathcal{H} \text{ is an obstruction design of type } \lambda\}.$$

Proof. According to Claim 22.3, for $\lambda \vdash^*_n d$, we have that

$$\mathsf{HWV}_\lambda(Sym^d \textstyle\bigotimes^3 \mathbb{C}^n) = \mathrm{span}\{\langle\widehat{\lambda}|\pi_{\mathcal{H}}\mathcal{P}_d : \pi \in \mathfrak{S}_d^3\}.$$

But since obstruction designs $\mathcal{H}$ determine $f_{\mathcal{H}} = \langle\widehat{\lambda}|\pi_{\mathcal{H}}\mathcal{P}_d$ up to a sign, the first assertion follows. The rest of the proposition follows from the fact that $k(\lambda) = \dim \mathsf{HWV}_\lambda(Sym^d \bigotimes^3 \mathbb{C}^n)$. □

Polynomial Evaluation We now describe how to evaluate the polynomial $f_{\mathcal{H}}$ corresponding to an obstruction design $\mathcal{H}$ at a point $|w\rangle = \sum_{i=1}^{r} |w_i^{(1)}\rangle \otimes |w_i^{(2)}\rangle \otimes |w_i^{(3)}\rangle \in \bigotimes^3 \mathbb{C}^n$. We calculate

$$f_{\mathcal{H}}(w) = \langle\widehat{\lambda}|\pi\mathcal{P}_d|w^{\otimes d}\rangle = \langle\widehat{\lambda}|\pi|w^{\otimes d}\rangle = \sum_{J\in\{1,\dots,r\}^d} \langle\widehat{\lambda}|\pi|w_{J_1}w_{J_2}\cdots w_{J_d}\rangle$$
$$\overset{\text{Prop. 22.11}}{=} \sum_{J\in\{1,\dots,r\}^d} \mathrm{val}_{\mathcal{H}}(w_{J_1}, w_{J_2}, \dots, w_{J_d}). \tag{22.9}$$

We can interpret $\zeta := (w_{J_1}, w_{J_2}, \dots, w_{J_d})$ as a vertex labeling $\zeta\colon V(\mathcal{H}) \to (\mathbb{C}^n)^3$ and see that the sum in (22.11) is over all vertex labelings $\zeta\colon V(\mathcal{H}) \to (\mathbb{C}^n)^3$ with $\zeta(y) \in \{w_i \mid 1 \le i \le r\}$ for all $y \in V(\mathcal{H})$.

22.1.4 The obstruction design for the hook triple

The obstruction design $\mathcal{H} := \mathcal{H}_\kappa$ that we construct consists of d vertices divided into disjoint sets $V^{(1)} \dot\cup V^{(2)} \dot\cup V^{(3)} \dot\cup \{y^0\}$, where $|V^{(k)}| = \kappa$ for all $1 \le k \le 3$. There are only three hyperedges of size larger than 1, called $e^{(k)}$ for $1 \le k \le 3$. We set $e^{(k)} := V^{(k+1)} \cup V^{(k+2)} \cup \{y^0\}$, where $V^{(k)} := V^{(k-3)}$ for $k > 3$. The obstruction design $\mathcal{H}$ is depicted in Figure 16.

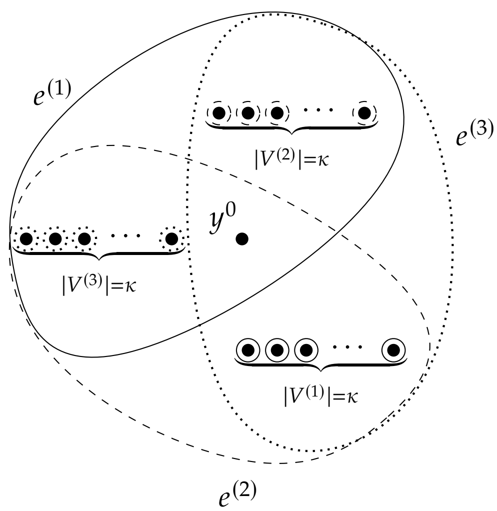

Figure 16: The family of obstruction designs corresponding to the hook partition triple.

22.2 Vanishing at low-rank points

In this subsection we prove (22.1), relying on our precise knowledge of the obstruction design $\mathcal{H}$ defined in Section 22.1.4. A second proof that only uses the invariance properties of the unit tensor is presented in Section 22.4.

Let $A \in (\mathbb{C}^{3\kappa\times 3\kappa})^3$ be arbitrary. We define the triple list

$$w := \big((A^{(1)}|1\rangle, A^{(2)}|1\rangle, A^{(3)}|1\rangle), \dots, (A^{(1)}|3\kappa\rangle, A^{(2)}|3\kappa\rangle, A^{(3)}|3\kappa\rangle)\big).$$

According to (22.9) we have

$$f_{\mathcal{H}}(\langle 3\kappa\rangle) = \sum_{J\in\{1,\dots,3\kappa\}^{3\kappa+1}} \mathrm{val}_{\mathcal{H}}(Aw_{J_1}, \dots, Aw_{J_{3\kappa+1}}). \tag{$*$}$$

The crucial property of $\mathcal{H}$ is that for each pair of vertices $\{y_1, y_2\}$ there exists a hyperedge e of $\mathcal{H}$ containing both y_1 and y_2. By the pigeon-hole principle, for each labeling $J\colon V(\mathcal{H}) \to \{1, \dots, 3\kappa\}$ there exists a pair of vertices $\{y_1, y_2\}$ such that $J(y_1) = J(y_2)$. The crucial property of $\mathcal{H}$ implies that y_1 and y_2 lie in a common hyperedge e. Hence $\mathrm{val}_e((Aw_{J_1}, \dots, Aw_{J_{3\kappa+1}})_{|_e}) = 0$, because it is the determinant of a matrix with two equal columns. Therefore, each summand in ($*$) vanishes, which proves (22.1).

This argument is analogous to Section 20.2.3.

22.3 Nonvanishing at the matrix multiplication tensor

In this rather technical section we prove (22.2) by an explicit construction of a matrix triple $A = (A^{(1)}, A^{(2)}, A^{(3)})$ consisting of maps $A^{(k)}\colon \mathbb{C}^{m\times m} \to \mathbb{C}^{m^2}$. We make use of the fact that m is odd.

For notational convenience, we define the triples

$$t_{ijl} := \big(|(ij)\rangle, |(jl)\rangle, |(li)\rangle\big) \in (\mathbb{C}^{m\times m})^3 \tag{22.10}$$

and the triple list w of length m^3 obtained by concatenating all t_{ijl} for $1 \le i, j, l \le m$ in any order. Recall that

$$\langle m, m, m\rangle = \sum_{i,j,l} t^{(1)}_{ijl} \otimes t^{(2)}_{ijl} \otimes t^{(3)}_{ijl}.$$

We put

$$\mathcal{T} := \{t_{ijl} \mid 1 \le i, j, l \le m\}.$$

$$f_{\mathcal{H}}(A\langle m, m, m\rangle) = \sum_{J\in\{1,\dots,m^3\}^d} \mathrm{val}_{\mathcal{H}}(Aw_{J_1}, \dots, Aw_{J_d}). \tag{$*$}$$

Consider the polynomial ring $\Gamma = \mathbb{C}[X_1, \dots, X_N]$, where X_i are indeterminates. According to Lemma 1.9, if a function $f \in \Gamma$ is nonzero, then there exist values $\alpha_i \in \mathbb{C}$, $1 \le i \le N$, such that $f(\alpha_1, \dots, \alpha_N) \neq 0$. We will define the $m^2 \times m^2$ matrix triple A with matrix entries being affine linear in the indeterminates X_i. Hence we write the sum $f_{\mathcal{H}}(A\langle m, m, m\rangle)$ as an element of Γ. We will provide a monomial of $f_{\mathcal{H}}(A\langle m, m, m\rangle)$ in the X_i with nonzero coefficient in (22.11).

Invariance in each $V^{(k)}$ We use the short notation $\mathrm{val}_e(\zeta) := \mathrm{val}_e(\zeta^{(k)}_{|e})$ for a hyperedge $e \in E^{(k)}$ and a triple labeling ζ. We start out with the following easy claim.

Claim 22.16. *Let $\sigma\colon V(\mathcal{H}) \to V(\mathcal{H})$ be a bijection satisfying $\sigma(V^{(k)}) = V^{(k)}$ for all $1 \le k \le 3$. For every triple labeling $\zeta\colon V(\mathcal{H}) \to (\mathbb{C}^{m^2})^3$ we have*

$$\mathrm{val}_{\mathcal{H}}(\zeta) = \mathrm{val}_{\mathcal{H}}(\zeta \circ \sigma).$$

Proof. It suffices to show the claim for a transposition $\tau = \sigma$ exchanging two elements of $V^{(1)}$, because the situation for $V^{(2)}$ and $V^{(3)}$ is completely symmetric. We have $\prod_{e\in E^{(1)}} \mathrm{val}_e(\zeta) = \prod_{e\in E^{(1)}} \mathrm{val}_e(\zeta \circ \tau)$, because up to reordering both products have the same factors. For $2 \le k \le 3$ we have $\mathrm{val}_e(\zeta) = \mathrm{val}_e(\zeta \circ \tau)$ for every singleton hyperedge $e \in E^{(k)}$ and $\mathrm{val}_{e^{(k)}}(\zeta) = -\mathrm{val}_{e^{(k)}}(\zeta \circ \tau)$. Therefore $\prod_{e\in E^{(k)}} \mathrm{val}_e(\zeta) = -\prod_{e\in E^{(k)}} \mathrm{val}_e(\zeta\circ\tau)$. As a result we get $\mathrm{val}_{\mathcal{H}}(\zeta) = (-1)^2\mathrm{val}_{\mathcal{H}}(\zeta\circ\tau)$. □

Special structure of the matrix triple Recall that m is odd and $\kappa = \frac{m^2-1}{2}$. Let $a := \frac{m+1}{2}$. Define the set $O_m := \{1,\dots,m\} \times \{1,\dots,m\} \setminus \{(a,a)\}$ consisting of $m^2 - 1$ pairs. Fix an arbitrary bijection

$$\varphi\colon O_m \to \{2,\dots,m^2\}.$$

Let $\bar{i} := m + 1 - i$ for $1 \le i \le m$. (We may think of the map $i \mapsto \bar{i}$ as a reflection at a; note $\bar{a} = a$.) Let

$$\Gamma := \mathbb{C}[\{X_i^{(k)} : 1 \le k \le 3,\ 1 \le i \le m\}]$$

denote the polynomial ring in $3m$ variables. For each $1 \le k \le 3$ we define the linear map $A^{(k)}\colon \mathbb{C}^{m\times m} \to \mathbb{C}^{m^2}$ by

$$A^{(k)}|(ij)\rangle := \begin{cases} X_a^{(k)}|1\rangle & \text{if } i = j = a \\ |\varphi(i,\bar{i})\rangle + X_i^{(k)}|1\rangle & \text{if } i \ne j \text{ and } j = \bar{i}\,. \\ |\varphi(i,j)\rangle & \text{if } j \ne \bar{i} \end{cases}$$

Hence $A^{(k)}$ looks as follows:

$$\left(\begin{array}{ccccccccc|c}
X_a^{(k)} & X_1^{(k)} & X_2^{(k)} & \cdots & X_{a-1}^{(k)} & X_{a+1}^{(k)} & \cdots & X_{m-1}^{(k)} & X_m^{(k)} & \\
 & 1 & & & & & & & & \\
 & & 1 & & & & & & & \\
 & & & \ddots & & & & & & \mathbf{0} \\
 & & & & 1 & & & & & \\
 & & & & & 1 & & & & \\
 & & & & & & \ddots & & & \\
 & & & & & & & 1 & & \\
 & & & & & & & & 1 & \\
\hline
\multicolumn{9}{c|}{\mathbf{0}} & id_{m^2-m}
\end{array}\right), \qquad (**)$$

where we arranged the rows and columns as follows: The left m columns correspond to the vectors $|(i\bar{i})\rangle$, where the leftmost one corresponds to $|(a,a)\rangle$. The top row corresponds to the vector $|1\rangle$ and the following $m-1$ rows correspond to the vectors $|\varphi(i,\bar{i})\rangle$. Recall that $f_{\mathcal{H}}(A\langle m,m,m\rangle)$ is a sum of products of determinants of submatrices of $A^{(k)}$.

Note that this is just a identity matrix with some additional entries in the top row (and the top left entry changed from a 1 to a variable). These top row entries are put there intentionally, so that the inner product with a column vector and the first standard basis vector is nonzero. The specific arrangement of the variables in the top row is not so crucial, but it helps in analyzing the evaluation $f_{\mathcal{H}}(A\langle m,m,m\rangle)$ later. Many other choices of matrices work here. This specific choice of A strikes a balance: It has a lot of zeros, so it is possible to calculate the evaluation $f_{\mathcal{H}}(A\langle m,m,m\rangle)$, but it has enough nonzero entries at the correct positions to actually ensure $f_{\mathcal{H}}(A\langle m,m,m\rangle) \neq 0$.

The sum $f_{\mathcal{H}}(A\langle m,m,m\rangle)$ is an element of Γ and we are interested in its coefficient of the monomial $\mathcal{X}$, where

$$\mathcal{X} := \prod_{k=1}^{3} X_a^{(k)} \prod_{i=1}^{m} \left(X_i^{(k)}\right)^{|i-\bar{i}|}. \tag{22.11}$$

We remark that the degree of $\mathcal{X}$ is $3(1+\sum_{i=1}^{m}|i-\bar{i}|)$. It is readily checked that $\sum_{i=1}^{m}|i-\bar{i}| = \kappa$.

Fix any numbering of the vertices of $\mathcal{H}$. For $J \in \{1,\ldots,m^3\}^d$ we abuse notation and define the map $J\colon V(\mathcal{H}) \to \mathscr{T}$ via $J(y) := w_{J_y}$. With this notation, (∗) becomes

$$\sum_J \mathrm{val}_{\mathcal{H}}\big(AJ(1),\ldots,AJ(d)\big),$$

or $\sum_J \mathrm{val}_{\mathcal{H}}(AJ)$ in short notation. We call a triple labeling $J\colon V(\mathcal{H}) \to \mathscr{T}$ *nonzero*, if the coefficient of $\mathcal{X}$ in the polynomial $\mathrm{val}_{\mathcal{H}}(AJ)$ is nonzero. Note that the sum of the evaluations of all nonzero triple labelings is the coefficient of $\mathcal{X}$ in the polynomial $f_{\mathcal{H}}(A\langle m,m,m\rangle)$. We will count and classify all nonzero triple labelings and show that they evaluate to the same nonzero value. This implies that the coefficient of $\mathcal{X}$ in $f_{\mathcal{H}}(A\langle m,m,m\rangle)$ is a sum without cancellations and is hence nonzero.

Separate Analysis of the Three Layers Given a triple labeling $J\colon V(\mathcal{H}) \to \mathscr{T}$, we define $J^{(k)}\colon V(\mathcal{H}) \to \{|(ij)\rangle \mid 1 \le i,j \le m\}$ by composing J with the projection to the kth component.

Claim 22.17. *Fix a nonzero triple labeling J and fix $1 \le k \le 3$. For all $y \in V^{(k)}$ we have $J^{(k)}(y) = |(i\bar{i})\rangle$ for some $1 \le i \le m$.*

Proof. Let $y \in V^{(k)}$. Since $\{y\} \in E^{(k)}$ we have $\langle 1|A^{(k)}|J^{(k)}(y)\rangle \neq 0$. From the definition of A it follows that $J^{(k)}(y) = |(ij)\rangle$ and the third case $j \neq \bar{i}$ is excluded. Hence $j = \bar{i}$. □

Claim 22.18. *For every nonzero triple labeling J we have $J(y^0) = (|(aa)\rangle, |(aa)\rangle, |(aa)\rangle)$.*

Proof. Let J be a nonzero triple labeling. Hence the coefficient of $\mathcal{X}$ in $\mathrm{val}_{\mathcal{H}}(AJ(1), \dots, AJ(d))$ is nonzero. For the following argument it is important to keep the structure of the matrix $A^{(k)}$ in mind, cf. (**). Recall that $f_{\mathcal{H}}(A\langle m, m, m\rangle)$ is a sum of products of certain subdeterminants of $A^{(k)}$ that are determined by the hyperedges in $E^{(k)}(\mathcal{H})$. Since the degree of $X_a^{(k)}$ in $\mathcal{X}$ is 1, we have that for all $1 \le k \le 3$ there is exactly one vertex $y_k \in V(\mathcal{H})$ with $J^{(k)}(y_k) = |(aa)\rangle$. Recall that the hyperedge $e^{(k)}$ has size $2\kappa + 1 = m^2$. Since J is a nonzero triple labeling, $J^{(k)}$ is injective on hyperedges and hence $|\{J^{(k)}(y) : y \in e^{(k)}\}| = m^2$. But since the image $J^{(k)}(V(\mathcal{H}))$ has cardinality at most m^2, $J^{(k)}$ is actually bijective on $e^{(k)}$. Since there is only one vertex y satisfying $J^{(k)}(y) = |(aa)\rangle$, namely the vertex $y = y_k$, it follows $y_k \in e^{(k)}$. Since $e^{(1)} \cap e^{(2)} \cap e^{(3)} = \{y^0\}$, it remains to show that $y_1 = y_2 = y_3$.

The structure of the matrix multiplication tensor implies that $J(y_1) = (|(aa)\rangle, |(ai)\rangle, |(ia)\rangle)$ for some $1 \le i \le m$. If $a = i$, then, by definition of y_2 and y_3 and uniqueness, we have $y_1 = y_2 = y_3$ and we are done.

Now assume $a \ne i$ and $y_1 \ne y^0$. W.l.o.g. $y_1 \in V^{(3)}$. Using Claim 22.17 we conclude that $J^{(3)}(y_1) = |i\bar{i}\rangle$ for some $1 \le i \le m$. Hence $\bar{i} = a$ contradicting $i \ne a$. Thus we have shown that $y_1 = y^0$. Similarly, we show that $y_2 = y_3 = y^0$. ☐

Claim 22.19. *For each nonzero triple labeling J we have $J^{(k)}(V^{(k)}) = \{|(i\bar{i})\rangle \mid 1 \le i \le m\} \setminus \{|(aa)\rangle\}$, where the preimage of each $|(i\bar{i})\rangle$ under $J^{(k)}$ has size $|i - \bar{i}|$.*

Proof. According to Claim 22.18 we have $J(y^0) = |(aa)(aa)(aa)\rangle$. For the following look again at the structure of $A^{(k)}$, cf. (**). Since $A^{(k)}|(aa)\rangle$ is a multiple of $|1\rangle$, we have that $\mathrm{val}_{e^{(k)}}(J)$ is a multiple of $X_a^{(k)}$. Moreover, for $i \ne a$, the variable $X_i^{(k)}$ does not appear in the expansion of $\mathrm{val}_{e^{(k)}}(J^{(k)})$. Since for a fixed $1 \le k \le 3$ there are $\kappa = \sum_{i=1}^{m} |i - \bar{i}|$ many contributions of a factor $X_i^{(k)}$ in the monomial $\mathcal{X}$, these factors must be contributed at vertices in $V^{(k)}$. Since $|V^{(k)}| = \kappa$, the only possibility is that all $y \in V^{(k)}$ satisfy $J^{(k)}(y) = |i\bar{i}\rangle$ for some $1 \le i \le m$, $i \ne a$. The specific requirement for the number of factors $X_i^{(k)}$ which are encoded in $\mathcal{X}$ in (22.11) finishes the proof. ☐

Coupling the Analysis of the Three Layers Define the bijective map

$$\tau\colon O_m \to O_m, \quad \tau(ij) = (j\bar{i}),$$

which corresponds to the rotation by 90°. Clearly, $\tau^4 = id$. The map τ induces a map $\wp(O_m) \to \wp(O_m)$ on the powerset, which we also call τ. Define the involution (taking the complement)

$$\iota\colon \wp(O_m) \to \wp(O_m), \quad S \mapsto O_m \setminus S.$$

Clearly, we have $\tau \circ \iota = \iota \circ \tau$. We will only be interested in subsets $S \subseteq O_m$ with exactly $|O_m|/2 = \kappa$ many elements and their images under τ and ι. The subsets $S \subseteq O_m$ that satisfy $\iota(S) = \tau(S)$ will be of special interest. Geometrically, these are the sets that get inverted when rotating by 90°, see Figure 17 for examples.

In the following we identify the sets $J^{(k)}(V^{(k')})$, for $1 \le k, k' \le 3$, with their corresponding subsets of O_m.

In Claim 22.19 we analyzed the labels $J^{(k)}(V^k)$. In the next claim we turn to $J^{(k)}(V^{k'})$, where $k \neq k'$.

Claim 22.20. *Every nonzero triple labeling J is completely determined by the image $J^{(1)}(V^{(3)})$ (up to permutations in the $V^{(k)}$, see Claim 22.16) as follows.*

- $J^{(2)}(V^{(3)}) = \tau(J^{(1)}(V^{(3)}))$,
- $J^{(2)}(V^{(1)}) = \iota(J^{(2)}(V^{(3)}))$,
- $J^{(3)}(V^{(1)}) = \tau(J^{(2)}(V^{(1)}))$,
- $J^{(3)}(V^{(2)}) = \iota(J^{(3)}(V^{(1)}))$,
- $J^{(1)}(V^{(2)}) = \tau(J^{(3)}(V^{(2)}))$.

Moreover, $\tau(J^{(1)}(V^{(3)})) = \iota(J^{(1)}(V^{(3)}))$.

Proof. According to Claim 22.19 we have that each vertex $y \in V^{(3)}$ satisfies

$$J(y) = \big(|(ij)\rangle, |(\tau(ij))\rangle, |(\bar{i}i)\rangle\big)$$

for some $1 \le i, j \le m$, $i \neq a$. In particular, using that τ is injective, we have

$$\tau(J^{(1)}(V^{(3)})) = J^{(2)}(V^{(3)}).$$

Since J is nonzero, $J^{(2)}$ is injective on $e^{(2)}$. We even have that $J^{(2)}$ is bijective on $e^{(2)}$, because $|e^{(2)}| = m^2$. Using that $e^{(2)} = V^{(1)} \dot\cup V^{(3)} \dot\cup \{y^0\}$ we see that

$$J^{(2)}(V^{(1)}) = O_m \setminus J^{(2)}(V^{(3)}) = \iota(J^{(2)}(V^{(3)})).$$

For the same reason, we can deduce $J^{(3)}(V^{(1)}) = \tau(J^{(2)}(V^{(1)}))$ and $J^{(3)}(V^{(2)}) = \iota(J^{(3)}(V^{(1)}))$. And applying these arguments one more time we get $J^{(1)}(V^{(2)}) = \tau(J^{(3)}(V^{(2)}))$ and $J^{(1)}(V^{(3)}) = \tau(J^{(1)}(V^{(2)}))$. Summarizing (recall $\tau \circ \iota = \iota \circ \tau$) we have

$$J^{(1)}(V^{(3)}) = \tau^3 \iota^3 (J^{(1)}(V^{(3)})) = \tau^{-1}\iota(J^{(1)}(V^{(3)})),$$

which is equivalent to $\tau(J^{(1)}(V^{(3)})) = \iota(J^{(1)}(V^{(3)}))$. □

Additionally to the constraint $\tau(J^{(1)}(V^{(3)})) = \iota(J^{(1)}(V^{(3)}))$ given in Claim 22.20, Claim 22.19 implies that in $J^{(1)}(V^{(3)})$ there are $|i - \bar{i}|$ many elements of the form $|(i\bar{i})\rangle$ for each $1 \le i \le m$.

This motivates the following definition.

Definition 22.21. A subset $S \subseteq O_m$ is called *valid*, if

(1) $|S| = \frac{m^2-1}{2} = \kappa$,

(2) $\tau(S) = \iota(S)$,

(3) $|p^{-1}(i)| = |i - \bar{i}|$ for all $1 \le i \le m$

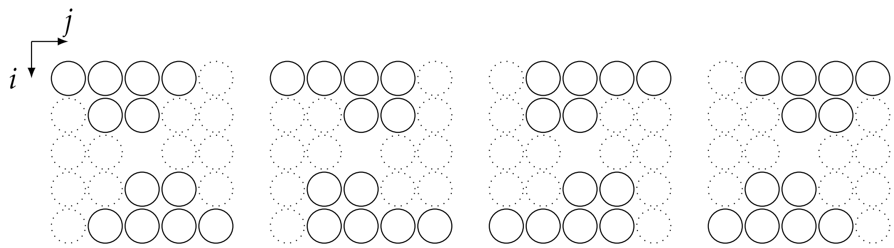

Figure 17: In each of the four pictures the vertices with solid border form a valid set for $m = 5$. The vertex in row i and column j represents the tuple (ij). The dotted vertices do *not* belong to the valid sets. Note that each vertex that does not lie on one of the two diagonals either lies in all valid sets or in no one. According to Lemma 22.23, there are no other valid sets for $m = 5$.

where $p\colon S \to \{1, \ldots, m\}$ is the projection to the first component, see Figure 17 for an example.

Proposition 22.22. *For all nonzero triple labelings J we have that $J^{(1)}(V^{(3)})$ is a valid set. On the other hand, for every valid set S there exists exactly one nonzero triple labeling J with $J^{(1)}(V^{(3)}) = S$ up to permutations in the $V^{(k)}$.*

Proof. For the first statement, property (2) of Definition 22.21 follows from Claim 22.20 and property (3) of Definition 22.21 follows from Claim 22.19. The second statement can be readily checked with Claim 22.18 and Claim 22.20. ◻

The next claim classifies all valid sets.

Lemma 22.23. *A set $S \subseteq O_m$ is valid iff the following conditions are all satisfied (see Figure 18 for an illustration):*

(1) $\Big\{(ij) \mid (i < j \text{ and } i < \bar{j}) \text{ or } (i > j \text{ and } i > \bar{j})\Big\} \subseteq S$*, represented by solid vertices in Figure 18.*

(2) $\Big\{(ij) \mid (i > j \text{ and } i < \bar{j}) \text{ or } (i < j \text{ and } i > \bar{j})\Big\} \cap S = \emptyset$*, represented by dotted vertices in Figure 18.*

(3) For all $1 \le i \le \frac{m-1}{2}$ there are two mutually exclusive cases, (a) and (b), represented by the two vertices x_i and the two vertices $\overline{x_i}$, respectively, in Figure 18.

(a) $\{(ii), (\bar{i}\,\bar{i})\} \subseteq S$ *and* $\{(i\bar{i}), (\bar{i}i)\} \cap S = \emptyset$,

(b) $\{(i\bar{i}), (\bar{i}i)\} \subseteq S$ *and* $\{(ii), (\bar{i}\,\bar{i})\} \cap S = \emptyset$.

These choices results in $2^{\frac{m-1}{2}}$ valid sets.

Proof. As indicated in Figure 18, for each tuple (ij) we call i the *row* of (ij). For S to be valid, according to Definition 22.21(3), S must contain $|i - \bar{i}|$ elements in row i and according to Definition 22.21(2), $\tau(s) \notin S$ for all $s \in S$.

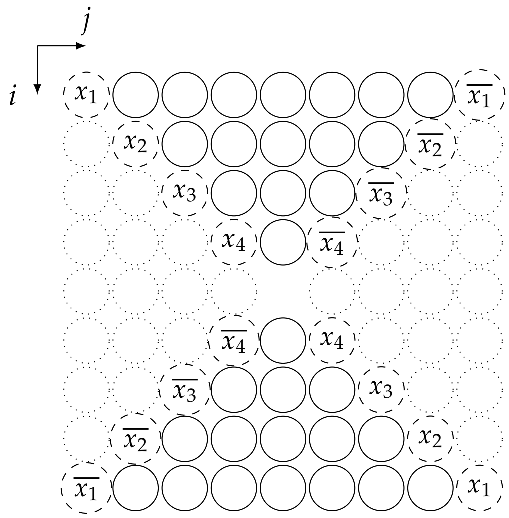

Figure 18: The case $m = 9$. Vertices that appear in all valid subsets are drawn with a solid border. Vertices that appear in no valid subset are drawn with a dotted border. Vertices that appear in half of all valid subsets are drawn with a dashed border. These contain a vertex label x_i or $\overline{x_i}$. Each valid set corresponds to a choice vector $x \in \{\text{true}, \text{false}\}^4$ determining whether the x_i or the $\overline{x_i}$ are contained in S. This results in $2^4 = 16$ valid sets $S \subseteq O_m$.

In particular, S must contain $m-1$ elements in row 1. If $(11) \in S$, then $(1m) \notin S$, because $\tau(11) = (1m)$. Hence there are only two possibilities: **(a):** $\{(1j) \mid 1 \le j < m\} \subseteq S$ or **(b):** $\{(1j) \mid 1 < j \le m\} \subseteq S$. By symmetry, for row m we get **(a′):** $\{(mj) \mid 1 \le j < m\} \subseteq S$ or **(b′):** $\{(mj) \mid 1 < j \le m\} \subseteq S$. But since $\tau(1m) = (mm)$ and $\tau(m1) = (11)$, the fact $\tau(S) = \iota(S)$ implies that **(a)** iff **(b′)** and that **(a′)** iff **(b)**. We are left with the two possibilities $\big($**(a)** and **(b′)**$\big)$ or $\big($**(a′)** and **(b)**$\big)$.

Now consider row 2. We have $\tau(21) = (1, m-1) \in S$ and hence $(21) \notin S$. In the same manner we see $(2m) \notin S$. We are left to choose $m-3$ elements from the $m-2$ remaining elements in row 2. The same argument as for row 1 gives two possibilities: **(a):** $\{(2j) \mid 2 \le j < m-1\} \subseteq S$ or **(b′):** $\{(2j) \mid 2 < j \le m-1\} \subseteq S$. Analogously for row $m-1$ we have **(a):** $\{((m-1), j) \mid 2 \le j < m-1\} \subseteq S$ or **(b′):** $\{((m-1), j) \mid 2 < j \le m-1\} \subseteq S$. With the same reasoning as for the rows 1 and m we get **(a)** iff **(b′)** and that **(a′)** iff **(b)**. Again we are left with the two possibilities $\big($**(a)** and **(b′)**$\big)$ or $\big($**(a′)** and **(b)**$\big)$.

Continuing these arguments we end up with $2^{\frac{m-1}{2}}$ possibilities. It is easy to see that each of these possibilities gives a valid set. $\square$

The following claim finishes the proof of (22.2).

Claim 22.24. *All nonzero triple labelings J have the same coefficient of $\mathcal{X}$ in the polynomial* $\mathrm{val}_{\mathcal{H}}(AJ)$*.*

Proof. Take two nonzero triple labelings J and J'. According to Proposition 22.22, both sets

$J^{(1)}(V^{(3)})$ and $J'^{(1)}(V^{(3)})$ are valid sets. Because of Lemma 22.23, it suffices to consider only the case where $J^{(1)}(V^{(3)})$ and $J'^{(1)}(V^{(3)})$ differ by a single involution $\sigma\colon O_m \to O_m$, where for some fixed $1 \le i \le \frac{m-1}{2}$ we have $\sigma(ii) = (i\bar{i})$ and $\sigma(\bar{i}\,\bar{i}) = (\bar{i}i)$, and σ is constant on all other pairs. We remark that σ restricted to the four pairs $\{(ii), (\bar{i}i), (i\bar{i}), (\bar{i}\,\bar{i})\}$ corresponds to a reflection in the second component.

We analyze the labels that are affected by this reflection. We only perform the analysis for one of the two symmetric cases, namely for $\{|(ii)\rangle, |(\bar{i}\,\bar{i})\rangle\} \subseteq J^{(1)}(V^{(3)})$. Note that this implies

$$\big\{\big(|(ii)\rangle, |(i\bar{i})\rangle, |(\bar{i}i)\rangle\big), \big(|(\bar{i}\,\bar{i})\rangle, |(\bar{i}i)\rangle, |(i\bar{i})\rangle\big)\big\} \subseteq J(V^{(3)}), \tag{†}$$

according to Claim 22.19. We adapt the notation from (22.10) to our special situation and write $t_{000} := t_{\bar{i}\bar{i}\bar{i}}$, $t_{001} := t_{\bar{i}\bar{i}i}$, $\ldots$, $t_{111} := t_{iii}$. Using this notation, (†) reads as follows: $\{t_{110}, t_{001}\} \subseteq J(V^{(3)})$. Using Claim 22.20 we get

$$\{t_{110}, t_{001}\} \subseteq J(V^{(3)}),\ \ \{t_{101}, t_{010}\} \subseteq J(V^{(2)}),\ \ \{t_{011}, t_{100}\} \subseteq J(V^{(1)}).$$

Applying the involution σ to $J^{(1)}(V^{(3)})$, we can use Claim 22.19 again to get

$$\big\{\big(|(i\bar{i})\rangle, |(\bar{i}\,\bar{i})\rangle, |(\bar{i}i)\rangle\big), \big(|(\bar{i}i)\rangle, |(ii)\rangle, |(i\bar{i})\rangle\big)\big\} \subseteq J'(V^{(3)}).$$

Applying Claim 22.20 and using our short syntax, we get:

$$\{t_{100}, t_{011}\} \subseteq J'(V^{(3)}),\ \ \{t_{001}, t_{110}\} \subseteq J'(V^{(2)}),\ \ \{t_{010}, t_{101}\} \subseteq J'(V^{(1)}).$$

We see that exactly the same triples occur in $J(V(\mathcal{H}))$ as in $J'(V(\mathcal{H}))$. We focus now on $J^{(1)}$ and $J'^{(1)}$ and see the following:

$$\{(ii), (\bar{i}\,\bar{i})\} \subseteq J^{(1)}(V^{(3)}) \text{ and } \{(i\bar{i}), (\bar{i}i)\} \subseteq J^{(1)}(V^{(2)})$$

and

$$\{(i\bar{i}), (\bar{i}i)\} \subseteq J'^{(1)}(V^{(3)}) \text{ and } \{(\bar{i}\,\bar{i}), (ii)\} \subseteq J'^{(1)}(V^{(2)}).$$

This gives exactly two switches of positions in $e^{(1)} = V^{(2)} \,\dot{\cup}\, V^{(3)} \,\dot{\cup}\, \{y^0\}$, hence

$$\mathrm{val}_{e^{(1)}}(AJ) = (-1)^2 \mathrm{val}_{e^{(1)}}(AJ') = \mathrm{val}_{e^{(1)}}(AJ').$$

Analogously we can prove that $\mathrm{val}_{e^{(k)}}(AJ) = \mathrm{val}_{e^{(k)}}(AJ')$ for all $2 \le k \le 3$ and therefore $\mathrm{val}_{\mathcal{H}}(AJ) = \mathrm{val}_{\mathcal{H}}(AJ')$. □

(22.2) is completely proved.

22.4 The coordinate ring of the unit tensor orbit

In this section we determine the multiplicities in the coordinate ring of the unit tensor. This proves (22.1) via (22.3). As a warm-up we study the closely related case of the power sum polynomial.

22.4.1 Warm-up: The coordinate ring of the power sum orbit

We use the notation from [20]: m is now the number of variables and D is the degree. The formulas in this subsection are unpublished calculations by Ikenmeyer and Panova. Parts also appear in the unpublished preprint [66].

The power sum is the polynomial $x_1^D + \cdots + x_m^D$. Let $H := \mathbb{Z}_D^m \rtimes \mathfrak{S}_m$ denote its stabilizer (see, e. g. [26, Ch. 2]). Let $\lambda \vdash Dd$.

If $\varrho \vdash_{\overline{m}} d$ is a partition, then the frequency notation $\kappa \in \mathbb{N}^m$ is defined via

$$\kappa_i = |\{j \mid \varrho_j = i\}|.$$

E. g., the frequency notation of $\varrho = (3,3,2,0)$ is $(0,1,2,0)$. We observe that $|\varrho| = \sum_i i\kappa_i$.

The group $\mathfrak{S}_m$ acts on $\mathbb{N}^m$ by permuting the positions. Note that under this action we have $\mathrm{stab}\varrho = \mathfrak{S}_{\kappa_1} \times \mathfrak{S}_{\kappa_2} \times \cdots \times \mathfrak{S}_{\kappa_m}$.

Theorem 22.25. $\dim\{\lambda\}^H = \sum_{\varrho \vdash_{\overline{m}} d} \sum_{\substack{\mu^1,\mu^2,\ldots,\mu^d \\ \mu^i \vdash \kappa_i Di}} c^{\lambda}_{\mu^1,\mu^2,\ldots,\mu^d} \prod_{i=1}^d a_{\mu^i}(\kappa_i, iD)$, *where κ is the frequency notation of ϱ, and $c^{\lambda}_{\mu^1,\mu^2,\ldots,\mu^d}$ is the* multi-Littlewood-Richardson coefficient *that denotes the multiplicity of $\{\lambda\}$ in the tensor product $\{\mu^1\} \otimes \ldots \otimes \{\mu^d\}$.*

Proof.

$$\{\lambda\}^H = (\{\lambda\}^{\mathbb{Z}_D^m})^{\mathfrak{S}_m} = \left(\bigoplus_{\substack{\gamma \in \mathbb{N}^m \\ |\gamma| = d}} [\lambda]^{G_\gamma} \right)^{\mathfrak{S}_m}$$

where for $\gamma \in \mathbb{N}^m, |\gamma| = d$, $G_\gamma \subseteq \mathfrak{S}_{dD}$ is defined as the Young subgroup $\mathfrak{S}_{\gamma_1 D} \times \cdots \times \mathfrak{S}_{\gamma_m D}$. The last equality can be seen using the tableau bases on both sides: While a basis of $\{\lambda\}$ is given by semistandard tableaux, by symmetrizing the basis over $\mathbb{Z}_D^m$ we see that a basis of $\{\lambda\}^{\mathbb{Z}_D^m}$ is given by semistandard tableaux in which each entry $1, \ldots, m$ occurs a multiple of D times, i. e., the content is $D\gamma$ for $\gamma \in \mathbb{N}^m, |\gamma| = d$. On the other hand, a basis of $[\lambda]$ is given by standard tableaux. Symmetrizing this basis over G_γ the images of basis vectors are sums in which each summand can be obtained from another by applying an element of G_γ. We assign to this sum a basis vector from $\{\lambda\}^{\mathbb{Z}_D^m}$ in the following way. All numbers that can be exchanged with each other under the action of G_γ get replaced with a single number: The first $\gamma_1 D$ numbers are replaced by a 1, the next $\gamma_2 D$ numbers are replaced by a 2, and so on. This results in a tableau with content $D\gamma$. This map is a bijection between basis vectors (and maps zero vectors to zero vectors). This process is very similar to the situation in [46, Sec. 4.3(A)].

For a partition $\varrho \vdash_{\overline{m}} d$ let $\mathfrak{S}_m\varrho \subseteq \mathbb{N}^m$ denote the orbit of ϱ. Note that ϱ is the only partition in its orbit, while the other lists are not in the correct order. Grouping the right-hand side in the previous equation we obtain

$$\bigoplus_{\varrho \vdash_{\overline{m}} d} \left(\bigoplus_{\gamma \in \mathfrak{S}_m \varrho} [\lambda]^{G_\gamma} \right)^{\mathfrak{S}_m},$$

so we can study each ϱ independently.

Let $\mathrm{stab}\varrho \leq \mathfrak{S}_m$ denote the stabilizer of ϱ.

Claim 22.26. $\dim\left(\bigoplus_{\gamma\in\mathfrak{S}_m\varrho}[\lambda]^{G_\gamma}\right)^{\mathfrak{S}_m} = \dim\left([\lambda]^{G_\varrho}\right)^{\mathrm{stab}\varrho}$.

Proof. We construct an isomorphism of vector spaces.

Let $W^\varrho := [\lambda]^{G_\varrho}$ and $W_\varrho := \bigoplus_{\gamma\in\mathfrak{S}_m\varrho} W^\gamma$. Let $\pi_1,\ldots,\pi_r$ be a system of representatives of left cosets for $\mathrm{stab}\varrho \leq \mathfrak{S}_m$ with $\pi_1 = \mathrm{id}$, i. e., $\mathfrak{S}_m = \pi_1\mathrm{stab}\varrho \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} \pi_r\mathrm{stab}\varrho$ and we have $\mathfrak{S}_m\varrho = \{\pi_1\varrho,\ldots,\pi_r\varrho\}$. Therefore we have the decomposition

$$W_\varrho = \bigoplus_{j=1}^{r} \pi_j W^\varrho.$$

Let $\overline{p} : W_\varrho \twoheadrightarrow W^\varrho$ be the projection according to this decomposition. We claim that the restriction

$$p : \left(W_\varrho\right)^{\mathfrak{S}_m} \to (W^\varrho)^{\mathrm{stab}\varrho}$$

is an isomorphism of vector spaces. This then finishes the proof. We verify well-definedness, injectivity, and surjectivity of p.

Well-definedness: The spaces $\pi_1 W^\varrho,\ldots,\pi_r W^\varrho$ are permuted by $\mathfrak{S}_m$. Every $\sigma \in \mathrm{stab}\varrho$ fixes W^ϱ, thus $\sigma v_1 = v_1$ if $v_1 \in W^\varrho$. Thus the map $v = \sum_{j=1}^r v_j \overset{\overline{p}}{\mapsto} v_1$ maps W_ϱ to $(W^\varrho)^{\mathrm{stab}\varrho}$.

Injectivity: If $v \in (W_\varrho)^{\mathfrak{S}_m}$, then $v = \pi v = \sum_j \pi v_j$. Therefore $v_j = \pi_j v_1$. If $p(v) = 0$, then $v_1 = 0$, thus all $v_j = 0$, which proves injectivity.

Surjectivity: Let $v_1 \in (W^\varrho)^{\mathrm{stab}\varrho}$. Set $v_j := \pi_j v_1$ and put $v := \sum_j v_j$. Clearly $p(v) = v_1$. It remains to verify that v is $\mathfrak{S}_m$-invariant.

$$v = \sum_{j=1}^{r} \pi_j v_1 = \sum_{j=1}^{r} \tfrac{1}{|\mathrm{stab}\varrho|} \sum_{\tau\in\mathrm{stab}\varrho} \pi_j \tau v_1 = \tfrac{1}{|\mathrm{stab}\varrho|} \sum_{\pi\in\mathfrak{S}_m} \pi v_1,$$

which is $\mathfrak{S}_m$-invariant. $\square$

We are left with determining $\dim\left([\lambda]^{G_\varrho}\right)^{\mathrm{stab}\varrho}$.

$$\dim\left([\lambda]^{G_\varrho}\right)^{\mathrm{stab}\varrho} = \dim \mathsf{HWV}_\lambda(\{\lambda\} \otimes ([\lambda]^{G_\varrho})^{\mathrm{stab}\varrho}) = \dim \mathsf{HWV}_\lambda((\otimes^{dD} V)^{G_\varrho \rtimes \mathrm{stab}\varrho})$$

$$\begin{aligned}
(\otimes^{dD} V)^{G_\varrho\rtimes\mathrm{stab}\varrho} &= (Sym^{D\varrho_1} V \otimes \cdots \otimes Sym^{D\varrho_m} V)^{\mathrm{stab}\varrho} \\
&= (\bigotimes^{\kappa_1} Sym^D V \otimes \bigotimes^{\kappa_2} Sym^{2D} V \otimes \cdots \otimes \bigotimes^{\kappa_d} Sym^{dD} V)^{\mathrm{stab}\varrho} \\
&= \underbrace{Sym^{\kappa_1} Sym^D V}_{=\bigoplus_{\mu^1}\{\mu^1\}^{\oplus a_{\mu^1}(\kappa_1,D)}} \otimes Sym^{\kappa_2} Sym^{2D} V \otimes \cdots \otimes \underbrace{Sym^{\kappa_d} Sym^{dD} V}_{=\bigoplus_{\mu^d}\{\mu^d\}^{\oplus a_{\mu^d}(\kappa_d,dD)}} \qquad (\dagger)
\end{aligned}$$

where κ is the frequency notation of ϱ. The multiplicity of $\{\mu^i\}$ in $Sym^{\kappa_i} Sym^{iD} V$ is $a_{\mu^i}(\kappa_i, iD)$. Let $c^{\lambda}_{\mu^1,\mu^2,\ldots,\mu^d}$ denote the multiplicity of $\{\lambda\}$ in the tensor product $\{\mu^1\} \otimes \ldots \otimes \{\mu^d\}$. Using distributivity we obtain that the multiplicity of $\{\lambda\}$ in the representation (†) equals

$$\sum_{\substack{\mu^1,\mu^2,\ldots,\mu^d \\ \mu^i \vdash \kappa_i D i}} c^{\lambda}_{\mu^1,\mu^2,\ldots,\mu^d} \prod_{i=1}^{d} a_{\mu^i}(\kappa_i, iD)$$

We conclude

$$\dim\{\lambda\}^H = \sum_{\varrho \vdash_{\overline{m}} d} \sum_{\substack{\mu^1,\mu^2,\ldots,\mu^d \\ \mu^i \vdash \kappa_i D i}} c^{\lambda}_{\mu^1,\mu^2,\ldots,\mu^d} \prod_{i=1}^{d} a_{\mu^i}(\kappa_i, iD).$$

□

22.4.2 From the power sum to the unit tensor

The stabilizer of $\langle n \rangle$ in GL_m^3 is $H := D_m \rtimes \mathfrak{S}_m$, where

$$D_m := \{(\mathrm{diag}(\alpha_1^{(1)}, \ldots, \alpha_m^{(1)}), \ldots, \mathrm{diag}(\alpha_1^{(3)}, \ldots, \alpha_m^{(3)})) \mid \forall i : \alpha_i^{(1)} \alpha_i^{(2)} \alpha_i^{(3)} = 1\},$$

see [20, Prop. 4.1].

As a straightforward generalization of Gay's theorem (Lemma 19.10) we define the *generalized plethysm coefficient* $a_\lambda(\mu, k)$ for a partition $\lambda \vdash mk$, a partition $\mu \vdash m$ and a natural number k via the decomposition

$$\{\lambda\}^{m \times k} = \bigoplus_{\mu \vdash m} [\mu]^{\oplus a_\lambda(\mu,k)}.$$

We obtain the classical plethysm coefficients $a_\lambda(m, k)$ when $\mu = (m)$ is a single row.

Theorem 22.27. $\dim\{\lambda, \lambda', \lambda''\}^H = \sum_{\varrho \vdash_{\overline{m}} d} \sum_{\beta,\beta',\beta''} j_{\beta,\varrho}(\lambda) j_{\beta',\varrho}(\lambda') j_{\beta'',\varrho}(\lambda'') \left(\prod_{i=1}^m k(\beta^i, \beta'^i, \beta''^i)\right)$, *where for κ being the frequency notation of ϱ*

- *the sum for β is over all lists of partitions such that $\beta^i \vdash \kappa_i$ and analogously for β' and β'', and*
- $j_{\beta,\varrho}(\lambda) := \sum_{\substack{\nu^1,\ldots,\nu^m \\ \nu^i \vdash i\kappa_i}} c^{\lambda}_{\nu^1,\ldots,\nu^m} \left(\prod_{i=1}^m a_{\nu^i}(\beta^i, i)\right)$,

Proof. $\{\lambda, \lambda', \lambda''\} = \{\lambda\} \otimes \{\lambda'\} \otimes \{\lambda''\}$.

$\{\lambda, \lambda', \lambda''\}$ has a basis given by triples of tableaux and D_m rescales basis vectors. Thus a vector is invariant if all basis vectors in its support are invariant.

D_m contains the subgroup

$$\{(\mathrm{diag}(\alpha, 1, 1, \ldots, 1), \mathrm{diag}(\alpha^{-1}, 1, 1, \ldots, 1), \mathrm{id})\}$$

and all other such subgroups where α and α^{-1} are both on position i on two different diagonals. A basis vector is invariant under these groups if all three tableaux have the same content. Since

D_m is generated by these groups, this precisely characterizes the invariants: $\{\lambda,\lambda',\lambda''\}^{D_m}$ has as a basis those triples of tableaux in which all three tableaux share the same content $\gamma \in \mathbb{N}^m$, $|\gamma| = d$:

$$\{\lambda,\lambda',\lambda''\}^{D_m} = \bigoplus_{\substack{\gamma\in\mathbb{N}^m,\\ |\gamma|=d}} \{\lambda\}^\gamma \otimes \{\lambda'\}^\gamma \otimes \{\lambda''\}^\gamma,$$

where $\{\lambda\}^\tau$ denotes the vector space of tableaux of shape λ and content τ.

$\bigoplus_{\gamma\in\mathfrak{S}_m\tau}\{\lambda\}^\gamma$ is an $\mathfrak{S}_m$-representation. As seen in the proof for the power sum, we group together with respect to the content:

$$(\{\lambda,\lambda',\lambda''\}^{D_m})^{\mathfrak{S}_m} = \bigoplus_{\varrho \,\underline{\vdash_m}\, d} \Big(\bigoplus_{\gamma\in\mathfrak{S}_m\varrho} \{\lambda\}^\gamma \otimes \{\lambda'\}^\gamma \otimes \{\lambda''\}^\gamma\Big)^{\mathfrak{S}_m}$$

Completely analogously to the proof for the power sum, we can take stabϱ-invariants instead of $\mathfrak{S}_m$-invariants:

$$\dim\Big(\bigoplus_{\gamma\in\mathfrak{S}_m\varrho} \{\lambda\}^\gamma \otimes \{\lambda'\}^\gamma \otimes \{\lambda''\}^\gamma\Big)^{\mathfrak{S}_m} = \dim(\{\lambda\}^\varrho \otimes \{\lambda'\}^\varrho \otimes \{\lambda''\}^\varrho)^{\mathrm{stab}\varrho}$$

We analyze the action of stabϱ separately on each of the three tableau spaces, i. e., we decompose $\{\lambda\}$, $\{\lambda'\}$, and $\{\lambda''\}$ as stabϱ-representations. Once this is done, Kronecker coefficients determine the stabϱ-invariant space dimension.

As seen in the proof for the power sum:

Claim 22.28.

$$\{\lambda\}^\varrho \overset{\mathrm{stab}\varrho\text{-}repr}{\simeq} \bigoplus_{\substack{\beta^1,\dots,\beta^m\\ \beta^i\vdash\kappa_i}} \underbrace{\sum_{\substack{\nu^1,\dots,\nu^m\\ \nu^i\vdash i\kappa_i}} c^\lambda_{\nu^1,\dots,\nu^m} \left(\prod_{i=1}^m a_{\nu^i}(\beta^i,i)\right)}_{=:j_{\beta,\varrho}(\lambda)} [\beta^1]\otimes\cdots\otimes[\beta^m],$$

where κ is the frequency notation of ϱ.

Proof. Recall that $\{\lambda\}^{m\times k} = \bigoplus_{\mu\vdash m} a_\lambda(\mu,k)[\mu]$.

We first prove (*): $\bigotimes^i Sym^j V = (\bigotimes^{ij} V)^{\mathfrak{S}^i_j} = \bigoplus_{\nu\vdash ij}\{\nu\}\otimes[\nu]^{\mathfrak{S}^i_j} = \bigoplus_{\nu\vdash ij}\{\nu\}\otimes\{\nu\}^{i\times j} = \bigoplus_{\nu\vdash ij,\varphi\vdash j} a_\nu(\varphi,j)\{\nu\}\otimes[\varphi]$, where for the last equality we use the generalized Gay's theorem.

Now we can calculate:

$$
\begin{aligned}
\bigoplus_{\lambda\vdash d}\{\lambda\}\otimes\{\lambda\}^{\varrho} &= \bigoplus_{\lambda\vdash d}\{\lambda\}\otimes[\lambda]^{G_\varrho} = (\bigotimes^d V)^{G_\varrho} = Sym^{\varrho_1}V\otimes\cdots\otimes Sym^{\varrho_m}V \\
&= \bigotimes^{\kappa_1} Sym^1 V \quad \otimes \quad \cdots \quad \otimes \quad \bigotimes^{\kappa_d} Sym^d V \\
&\overset{(*)}{=} \bigoplus_{\substack{\nu^1\vdash 1\kappa_1\\ \beta\vdash\kappa_1}} a_{\nu^1}(\beta^1,1)\{\nu^1\}\otimes[\beta^1] \quad \otimes \quad \cdots \quad \otimes \quad \bigoplus_{\substack{\nu^d\vdash d\kappa_d\\ \beta\vdash\kappa_d}} a_{\nu^d}(\beta^d,d)\{\nu^d\}\otimes[\beta^d] \\
&= \bigoplus_{\nu,\beta}(\prod_{i=1}^m a_{\nu^i}(\beta^i,i))(\{\nu^1\}\otimes\cdots\otimes\{\nu^m\})\otimes[\beta^1]\otimes\cdots\otimes[\beta^m].
\end{aligned}
$$

Taking HWVs of weight λ on both sides we obtain

$$
\{\lambda\}^{\varrho} = \bigoplus_{\nu,\beta} c^{\lambda}_{\nu^1,\dots,\nu^m}(\prod_{i=1}^m a_{\nu^i}(\beta^i,i))[\beta^1]\otimes\cdots\otimes[\beta^m].
$$

□

Since the dimension of the $\mathfrak{S}_{\kappa_i}$-invariant space of $[\beta^i]\otimes[\beta'^i]\otimes[\beta''^i]$ is given by the Kronecker coefficient $k(\beta^i,\beta'^i,\beta''^i)$, we obtain:

$$
\dim(\{\lambda\}^{\varrho}\otimes\{\lambda'\}^{\varrho}\otimes\{\lambda''\}^{\varrho})^{\mathrm{stab}\varrho} = \sum_{\beta,\beta',\beta''} j_{\beta,\varrho(\lambda)} j_{\beta',\varrho}(\lambda') j_{\beta'',\varrho}(\lambda'')\left(\prod_{i=1}^m k(\beta^i,\beta'^i,\beta''^i)\right),
$$

where the sum for β is over all lists of partitions such that $\beta^i\vdash\kappa_i$ and analogously for β' and β''. □

The following second proof for (22.1) is taken from the lecture notes [11]:

Corollary 22.29. *As at the beginning of Chapter 22, let $\lambda=\lambda'=\lambda''$ be the hook partition with $3k+1$ boxes and $2k+1$ rows. Then* $\mathrm{mult}_{(\lambda,\lambda',\lambda'')}(\mathsf{GL}^3_{3k}\langle 3k\rangle)=0$.

Proof. We use the formula in Theorem 22.27. Since it has no signs, we can assume (for the sake of contradiction) that the formula yields a is positive result and derive conditions on the partitions that are involved in positive summands.

We use a few standard facts about Littlewood-Richardson coefficients, plethysm coefficients, and Kronecker coefficients, each marked with a †.

First observation: $\nu^1=\beta^1$, because of the plethysm $a_{\nu^1}(\beta^1,1)=\mathrm{mult}_{\nu^1}(\underbrace{S^{\beta^1}(Sym^1V)}_{=\{\beta^1\}})$.

A multi-LR-coefficients can only be positive if all small partitions are contained in the large partition, i. e., the small Young diagrams are subsets of the large Young diagram (†[10]). In our case, all large partitions are hooks, so all ν^i are hooks. Thus also β^1, β'^1, β''^1 are hooks.

Let d be the number of boxes. For a hook ν^1 define the *inner leg length* as $\ell(\nu^1) - 1$. For hook triples with inner leg lengths a_1, a_2, a_3, Kronecker positivity requires (†, see, e. g., [70, Pf. of Thm. 3(4.)]):

$$2d - a_1 - a_2 - a_3 - 2 \geq 0.$$

Thus not all three a_1, a_2, a_3 can be large. Indeed, let $a = \min\{a_1, a_2, a_3\}$, then $2d - 3a - 2 \geq 0$ and thus $a \leq \frac{2d-2}{3}$. In particular this holds for $k(\nu^1, \nu'^1, \nu''^1) = k(\beta^1, \beta'^1, \beta''^1) > 0$. W.l.o.g. ν^1 is the shortest of ν^1, ν'^1, ν''^1. Then

$$\ell(\nu^1) - 1 \leq \frac{2|\nu^1| - 2}{3} = \frac{2}{3}|\nu^1| - \frac{2}{3}$$

and thus

$$\ell(\nu^1) \leq \frac{2}{3}|\nu^1| + \frac{1}{3}.$$

All partitions appearing in $\bigotimes^a Sym^b V$ have at most a rows, as the basis of HWVs is given by semistandard tableaux with content $(b, b, \ldots, b)$. Therefore the positive plethysm coefficients in the formula imply

$$\ell(\nu^i) \leq |\beta^i| = \kappa_i = \frac{\nu^i}{i}$$

Adding up the lengths we obtain

$$\ell(\nu^1) + \cdots + \ell(\nu^\ell) \leq \frac{2}{3}|\nu^1| + \frac{1}{3} + \frac{1}{2}(\underbrace{|\nu^2| + \cdots + |\nu^\ell|}_{=3k+1-|\nu^1|})$$

$$= \frac{2}{3}|\nu^1| + \frac{1}{3} + \frac{3}{2}k + \frac{1}{2} - \frac{1}{2}|\nu^1| = \frac{3}{2}k + \frac{1}{6}|\nu^1| + \frac{5}{6}$$

We now use that for a positive multi-LRC the lengths of the small partitions add up to at least the length of the large partition (†[11]):

$$\ell(\nu^1) + \cdots + \ell(\nu^\ell) \geq \ell(\lambda) = 2k + 1.$$

Therefore

$$\frac{3}{2}k + \frac{5}{6} + \frac{1}{6}|\nu^1| \geq 2k + 1 \Leftrightarrow -\frac{1}{2}k - \frac{1}{6} + \frac{1}{6}|\nu^1| \geq 0 \Leftrightarrow |\nu^1| \geq 3k + 1.$$

[10] $c^{\lambda}_{\nu^1,\ldots,\nu^m}$ equals the multiplicity of $\{\lambda\}$ in the tensor product $\{\nu^1\} \otimes \cdots \{\nu^m\}$. [36, §5.2, Prop. 3] treats the case $m = 2$. Expanding the tensor product $\{\nu^1\} \otimes \cdots \otimes \{\nu^m\}$ via the two-factor Littlewood-Richardson rule show that all $\{\lambda\}$ that occur in $\{\nu^1\} \otimes \cdots \otimes \{\nu^m\}$ must have $\nu^i \subseteq \lambda$. see, e. g., [36, §8.3, Cor. 2(c)] for this interpretation of LR-coefficients.

[11] This can be seen directly from a combinatorial description of the LR-coefficient, see, e. g., [36, §5.2, Prop. 3] and it readily generalizes to more than 2 factors.

Since $|\nu^1| = \kappa_1$, this means that $\varrho^1 = (1^{3k+1})$, but the sum is only over ϱ^1 with at most $3k$ rows. □

Occurrence obstructions for matrix multiplication

The tensor setting is analogous polynomial setting, but the group GL_{n^2} is replaced by $\mathsf{GL}_n \times \mathsf{GL}_n \times \mathsf{GL}_n$.
One can explicitly construct occurrence obstructions that show border rank lower bounds on the matrix multiplication tensor. Obstruction designs help visualizing the arguments.

A Some basic algebraic vocabulary

This appendix contains some basic notions from algebra.

A *monoid* $(G, \cdot, e)$ is set with a binary operation $\cdot\colon G \times G \to G$ and a so-called neutral element $e \in G$ such that the following conditions hold:

1. Associativity: For all $a, b, c \in G$ we have $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
2. Existence of identity: For all $a \in G$ we have $e \cdot a = a = a \cdot e$.

We omit the multiplication dot if there is no possibility of confusion. For example, the set $\mathbb{C}^{n\times n}$ of complex $n \times n$ matrices with binary operation the matrix multiplication is a monoid. A subset of a monoid G which contains the identity element and satisfies associativity is called a *submonoid of* G. For example, the set of upper triangular complex $n \times n$ matrices is a submonoid of the monoid of complex $n \times n$ matrices. A monoid G is called *commutative* if for all $a, b \in G$ we have $ab = ba$. A monoid homomorphism from a monoid $(G, \cdot, e)$ to a monoid $(G', \cdot', e')$ is a map $\varphi\colon G \to G'$ which satisfies $\varphi(g \cdot h) = \varphi(g) \cdot' \varphi(h)$ for all $(g, h) \in G \times G$ and $\varphi(e) = e'$.

A *group* is a monoid in which for each element $a \in G$ we have an element $a^{-1} \in G$ such that $a^{-1} \cdot a = e = a \cdot a^{-1}$. The element a^{-1} is called the *inverse element* of a. For example, the set of invertible (i. e., nonzero determinant) complex $n \times n$ matrices with operation the matrix multiplication is a group, the so called *general linear group* GL_n, where a^{-1} is the matrix inverse. Another example is the *symmetric group on n letters* $\mathfrak{S}_n$, which consists of all bijective maps from the set $\{1, 2, \dots, n\}$ to {1,2,. . . ,n} and the operation is the composition of maps, where a^{-1} is the inverse permutation. A monoid homomorphism between groups is called a *group homomorphism*. A group G is called *abelian* if it is commutative as a monoid. For example the set $(\mathbb{C}, +, 0)$ of complex numbers with addition is an abelian group. Moreover, the set $(\mathbb{C} \setminus \{0\}, \cdot, 1)$ of complex numbers (without zero) with multiplication is an abelian group. Also the set $(\mathbb{C}^{n\times m}, +, 0)$ of $n \times m$ matrices with addition is an abelian group.

A *field* $(\mathbb{F}, +, \cdot, 0, 1)$ is a set with two binary operations and two specific elements that satisfies:

1. $(\mathbb{F}, +, 0)$ and $(\mathbb{F} \setminus \{0\}, \cdot, 1)$ are abelian groups, and
2. distributivity holds, i. e., for all $a, b, c \in \mathbb{F}$ we have $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

For example, the complex numbers $\mathbb{C}$ form a field.

If we do not require the existence of multiplicative inverse elements, then our algebraic structure is called a *ring*, more precisely: A ring $(R, +, \cdot, 0, 1)$ is a set with two binary operations and two neutral elements that satisfies:

1. $(R, +, 0)$ is an abelian group,
2. $(R, \cdot, 1)$ is a monoid,
3. distributivity holds, i. e., for all $a, b, c \in R$ we have $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

If the monoid $(R, \cdot, 1)$ is commutative, then we call the ring R a *commutative ring*. For example, the set $\mathbb{C}[X_1, X_2, \ldots, X_n]$ of polynomials in n variables is a commutative ring. The set $\mathbb{C}^{n\times n}$ of $n \times n$ matrices with matrix multiplication and addition is a *non*commutative ring for $n > 1$. A *ring homomorphism* between two rings R_1 and R_2 is a map from R_1 to R_2 which at the same time is a group homomorphism for the additive structure and a monoid morphism for the multiplicative structure. A bijective ring homomorphism is called a *ring isomorphism*.

An *ideal* I of a commutative ring $(R, +, \cdot, 0, 1)$ is a subset $I \subseteq R$ that forms an abelian group $(I, +)$ with the rings addition and moreover is closed under ring multiplication, i. e., for all $a \in R$, $b \in I$ we have $a \cdot b \in I$. For example, the set of polynomials f in $\mathbb{C}[X_1, X_2, \ldots, X_n]$ that are divisible by X_2 form an ideal of $\mathbb{C}[X_1, X_2, \ldots, X_n]$.

If $\mathbb{F}$ is a field, $\alpha \in \mathbb{F}$, and a is an element in a vector space over $\mathbb{F}$, then we write $\alpha.a$ for the scalar multiplication. A vector space over the complex numbers is also called a $\mathbb{C}$-vector space. A ring $(A, +, \cdot, 0, 1)$ that is also $\mathbb{C}$-vector space (with the same addition) is called a $\mathbb{C}$-*algebra*, if A satisfies $(\alpha.1) \cdot a = \alpha.a$ for all $\alpha \in \mathbb{C}$ and all $a \in A$. If the $\mathbb{C}$-algebra A is a commutative ring, then we call A a *commutative* $\mathbb{C}$-*algebra*. For example the set $\mathbb{C}[X_1, X_2, \ldots, X_n]$ of polynomials in n variables is a commutative $\mathbb{C}$-algebra.

A vector space V is a *direct sum* of linear subspaces $V_i \subseteq V$, written $V = \bigoplus_i V_i$, if the union $\bigcup_i V_i$ spans V and each intersection $V_i \cap \sum_{j\neq i} V_j$ is the zero space. If $V = \bigoplus_i V_i$, then each element in V has a unique representation as a sum of elements from the V_i.

A $\mathbb{C}$-algebra A is called *graded*, if the vector space A is a direct sum $A = \bigoplus_{d\in\mathbb{N}_{\geq 0}} A_d$ such that the ring multiplication satisfies $a \cdot a' \in A_{d+d'}$ for all $a \in A_d$ and $a' \in A_{d'}$. For example the algebra $\mathbb{C}[X_1, X_2, \ldots, X_n]$ is graded as follows: The linear subspace $\mathbb{C}[X_1, X_2, \ldots, X_n]_d$ is spanned by the monomials of degree d, where the degree is the sum of exponents, e. g., the monomial $X_1^2 X_2^3$ has degree 5. We call $\mathbb{C}[X_1, X_2, \ldots, X_n]_d$ the *homogeneous degree d component of* $\mathbb{C}[X_1, X_2, \ldots, X_n]$ and elements of $\mathbb{C}[X_1, X_2, \ldots, X_n]_d$ are said to be *homogeneous of degree d*. For example $X_1^2 X_2^3 - X_1 X_2^3$ is *not* homogeneous. The degree of a nonhomogeneous polynomial is defined to be the maximal degree of its monomials. We define $\mathbb{C}[X_1, X_2, \ldots, X_n]_{\leq d}$ to be the vector space of (not necessarily homogeneous) polynomials of degree at most d. A *homomorphism of* $\mathbb{C}$-*algebras* is a linear map that is a ring homomorphism. An *isomorphism of* $\mathbb{C}$-*algebras* is a linear map that is a ring isomorphism. An *isomorphism of graded* $\mathbb{C}$-*algebras* $f : A \to B$ is defined to be an isomorphism of $\mathbb{C}$-algebras such that the restriction of f to each homogeneous degree i part A_i is a vector space isomorphism to B_i.

References

[1] SCOTT AARONSON: $\mathrm{P} \stackrel{?}{=} \mathrm{NP}$. In JOHN FORBES NASH, JR. AND MICHAEL TH. RASSIAS, editors, *Open Problems in Mathematics*, pp. 1–122. Springer, 2016. [doi:10.1007/978-3-319-32162-2_1] 5

[2] SCOTT AARONSON: $P \stackrel{?}{=} NP$. *Electron. Colloq. Comput. Complexity*, TR17-004, 2017. [ECCC] 5

[3] ALEXANDER ALDER: *Grenzrang und Grenzkomplexität aus algebraischer und topologischer Sicht*. Ph. D. thesis, Zurich University, 1984. 88

[4] ERIC ALLENDER AND FENGMING WANG: On the power of algebraic branching programs of width two. *Comput. Complexity*, 25(1):217–253, 2016. [doi:10.1007/s00037-015-0114-7] 34

[5] WALTER BAUR AND VOLKER STRASSEN: The complexity of partial derivatives. *Theoret. Comput. Sci.*, 22(3):317–330, 1983. [doi:10.1016/0304-3975(83)90110-X] 10

[6] MICHAEL BEN-OR AND RICHARD CLEVE: Computing algebraic formulas using a constant number of registers. *SIAM J. Comput.*, 21(1):54–58, 1992. [doi:10.1137/0221006] 32

[7] DARIO BINI, MILVIO CAPOVANI, GRAZIA LOTTI, AND FRANCESCO ROMANI: $O(n^{2.7799})$ complexity for $n \times n$ matrix multiplication. *Inform. Process. Lett.*, 8(5):234–235, 1979. [doi:10.1016/0020-0190(79)90113-3] 82, 87

[8] MARKUS BLÄSER: Lower bounds for the bilinear complexity of associative algebras. *Comput. Complexity*, 9(2):73–112, 2000. [doi:10.1007/PL00001605] 83

[9] MARKUS BLÄSER: Complete problems for Valiant's class of qp-computable families of polynomials. In *Proc. 7th Internat. Combinatorics and Computing Conf. (COCOON'01)*, pp. 1–10. Springer, 2001. [doi:10.1007/3-540-44679-6_1] 30

[10] MARKUS BLÄSER: *Fast Matrix Multiplication*. Number 5 in Graduate Surveys. Theory of Computing Library, 2013. [doi:10.4086/toc.gs.2013.005] 80, 82

[11] MARKUS BLÄSER AND CHRISTIAN IKENMEYER: Lecture notes for Geometric Complexity Theory 2 (unpublished). Available at U. Warwick. 156

[12] MARKUS BLÄSER, CHRISTIAN IKENMEYER, GORAV JINDAL, AND VLADIMIR LYSIKOV: Generalized matrix completion and algebraic natural proofs. In *Proc. 50th STOC*, pp. 1193–1206. ACM Press, 2018. [doi:10.1145/3188745.3188832] 79

[13] MARKUS BLÄSER, CHRISTIAN IKENMEYER, MEENA MAHAJAN, ANURAG PANDEY, AND NITIN SAURABH: Algebraic branching programs, border complexity, and tangent spaces. In *Proc. 35th Comput. Complexity Conf. (CCC'20)*, pp. 21:1–24. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2020. [doi:10.4230/LIPIcs.CCC.2020.21] 54

[14] RICHARD P. BRENT: The parallel evaluation of general arithmetic expressions. *J. ACM*, 21(2):201–206, 1974. [doi:10.1145/321812.321815] 32

[15] KARL BRINGMANN, CHRISTIAN IKENMEYER, AND JEROEN ZUIDDAM: On algebraic branching programs of small width. *J. ACM*, 65(5):32:1–29, 2018. [doi:10.1145/3209663] 34, 54

[16] PETER BÜRGISSER: *Kombinatorik der Darstellungstheorie symmetrischer Gruppen* (Combinatorics of the representation theory of symmetric groups). Course held in 2006 at Paderborn University. Handwritten course notes by Christian Ikenmeyer, unpublished. 57

[17] PETER BÜRGISSER: *Completeness and Reduction in Algebraic Complexity Theory*. Springer, 2000. [doi:10.1007/978-3-662-04179-6] 30, 52

[18] PETER BÜRGISSER, MATTHIAS CHRISTANDL, AND CHRISTIAN IKENMEYER: Even partitions in plethysms. *J. Algebra*, 328(1):322–329, 2011. [doi:10.1016/j.jalgebra.2010.10.031] 119

[19] PETER BÜRGISSER, MICHAEL CLAUSEN, AND M. AMIN SHOKROLLAHI: *Algebraic Complexity Theory*. Springer, 1997. [doi:10.1007/978-3-662-03338-8] 36, 82

[20] PETER BÜRGISSER AND CHRISTIAN IKENMEYER: Geometric complexity theory and tensor rank. In *Proc. 43rd STOC*, pp. 509–518. ACM Press, 2011. [doi:10.1145/1993636.1993704] 134, 152, 154

[21] PETER BÜRGISSER AND CHRISTIAN IKENMEYER: Explicit lower bounds via geometric complexity theory. In *Proc. 45th STOC*, pp. 141–150. ACM Press, 2013. [doi:10.1145/2488608.2488627] 134

[22] PETER BÜRGISSER AND CHRISTIAN IKENMEYER: Fundamental invariants of orbit closures. *J. Algebra*, 477:390–434, 2017. [doi:10.1016/j.jalgebra.2016.12.035] 121

[23] PETER BÜRGISSER, CHRISTIAN IKENMEYER, AND GRETA PANOVA: No occurrence obstructions in geometric complexity theory. *J. AMS*, 32(1):163–193, 2019. Preliminary version in FOCS'16. [doi:10.1090/jams/908] 74, 106, 109, 121, 122, 124

[24] 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 J. Comput.*, 40(4):1179–1209, 2011. [doi:10.1137/090765328] 6

[25] PRASAD CHAUGULE, NUTAN LIMAYE, AND SHOURYA PANDEY: Variants of the determinant polynomial and VP-completeness. In *Proc. 16th Comp. Sci. Symp. in Russia (CSR'21)*, pp. 31–55. Springer, 2021. [doi:10.1007/978-3-030-79416-3_3] 30, 37

[26] XI CHEN, NEERAJ KAYAL, AND AVI WIGDERSON: Partial derivatives in arithmetic complexity and beyond. *Found. Trends Theor. Comp. Sci.*, 6(1–2):1–138, 2011. [doi:10.1561/0400000043] 152

[27] STEPHEN COOK: The P versus NP problem. In *The Millennium Prize Problems*. Clay Mathematical Institute and American Mathematical Society, 2000. Available at AMS Bookstore. 5

[28] DON COPPERSMITH AND SHMUEL WINOGRAD: On the asymptotic complexity of matrix multiplication. *SIAM J. Comput.*, 11(3):472–492, 1982. [doi:10.1137/0211038] 86

[29] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein: *Introduction to Algorithms*. MIT Press, 3rd edition, 2009. 83

[30] László Csánky: Fast parallel matrix inversion algorithms. *SIAM J. Comput.*, 5(4):618–623, 1976. [doi:10.1137/0205040] 44

[31] Arnaud Durand, Meena Mahajan, Guillaume Malod, Nicolas de Rugy-Altherre, and Nitin Saurabh: Homomorphism polynomials complete for VP. *Chicago J. Theoret. Comp. Sci.*, 2016(3), 2016. Preliminary version in FSTTCS'14. 30, 37

[32] Klim Efremenko, Ankit Garg, Rafael Oliveira, and Avi Wigderson: Barriers for rank methods in arithmetic complexity. In *Proc. 9th Innovations in Theoret. Comp. Sci. Conf. (ITCS'18)*, pp. 1:1–19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.ITCS.2018.1] 135

[33] Ismor Fischer: Sums of like powers of multivariate linear forms. *Math. Magazine*, 67(1):59–61, 1994. [doi:10.2307/2690560] 12

[34] Lance Fortnow: The status of the P versus NP problem. *Commun. ACM*, 52(9):78–86, 2009. [doi:10.1145/1562164.1562186] 5

[35] Georg Ferdinand Frobenius: Über die Darstellung der endlichen Gruppen durch lineare Substitutionen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, pp. 994–1015, 1897. [doi:10.3931/e-rara-18879] 106

[36] William Fulton: *Young Tableaux*. Cambridge Univ. Press, 1996. [doi:10.1017/CBO9780511626241] 95, 157

[37] Ankit Garg, Visu Makam, Rafael Oliveira, and Avi Wigderson: More barriers for rank methods, via a "numeric to symbolic" transfer. In *Proc. 60th FOCS*, pp. 824–844. IEEE Comp. Soc., 2019. [doi:10.1109/FOCS.2019.00054] 135

[38] David A. Gay: Characters of the Weyl group of $SU(n)$ on zero weight spaces and centralizers of permutation representations. *Rocky Mountain J. Math.*, 6(3):449–455, 1976. JSTOR. 114

[39] Dima Grigoriev, Mikhail Muzychuk, and Ilya Ponomarenko: Tensor rank: Matching polynomials and Schur rings. *Found. Computational Math.*, 14(3):457–481, 2014. [doi:10.1007/s10208-014-9190-3] 135

[40] Joshua A. Grochow: *Symmetry and equivalence relations in classical and geometric complexity theory*. Ph. D. thesis, Department of Computer Science, University of Chicago, 2012. Available at author's website. 6

[41] Joshua A. Grochow: Unifying known lower bounds via geometric complexity theory. *Comput. Complexity*, 24(2):393–475, 2015. [doi:10.1007/s00037-015-0103-x] 54

[42] JOSHUA A. GROCHOW, KETAN D. MULMULEY, AND YOUMING QIAO: Boundaries of VP and VNP. In *Proc. 43rd Internat. Colloq. on Automata, Languages, and Programming (ICALP'16)*, pp. 34:1–14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016. [doi:10.4230/LIPIcs.ICALP.2016.34] 54

[43] JOHAN HÅSTAD: Tensor rank is NP-complete. *J. Algorithms*, 11(4):644–654, 1990. [doi:10.1016/0196-6774(90)90014-6] 79

[44] ROGER HOWE: $(\mathrm{GL}_n, \mathrm{GL}_m)$-duality and symmetric plethysm. *Proc. Indian Acad. Sci. Math. Sci.*, 97(1–3):85–109, 1987. [doi:10.1007/BF02837817] 115

[45] JESKO HÜTTENHAIN AND PIERRE LAIREZ: The boundary of the orbit of the 3-by-3 determinant polynomial. *C. R. Math. Acad. Sci. Paris*, 354(9):931–935, 2016. [doi:https://doi.org/10.1016/j.crma.2016.07.002] 23

[46] CHRISTIAN IKENMEYER: *Geometric Complexity Theory, Tensor Rank, and Littlewood-Richardson Coefficients*. Ph. D. thesis, Institute of Mathematics, University of Paderborn, 2012. Online available at U. Paderborn. 6, 113, 134, 142, 152

[47] CHRISTIAN IKENMEYER AND GRETA PANOVA: Rectangular Kronecker coefficients and plethysms in geometric complexity theory. *Adv. Math.*, 319:40–66, 2017. Preliminary version in FOCS'16. [doi:10.1016/j.aim.2017.08.024] 36, 106, 129

[48] HARLAN KADISH AND JOSEPH M. LANDSBERG: Padded polynomials, their cousins, and geometric complexity theory. *Comm. Algebra*, 42(5):2171–2180, 2014. [doi:10.1080/00927872.2012.758268] 107, 120

[49] ALLEN KNUTSON: Peter-Weyl vs. Schur-Weyl theorem. MathOverflow post: `https://mathoverflow.net/questions/267552/peter-weyl-vs-schur-weyl-theorem`, 2017. 108

[50] EMMANUEL KOWALSKI: *An Introduction to the Representation Theory of Groups*. Amer. Math. Soc., 2014. [doi:10.1090/gsm/155] 104

[51] HANSPETER KRAFT: *Geometrische Methoden in der Invariantentheorie*. Friedr. Vieweg und Sohn Verlagsgesellschaft, Braunschweig, 1985. [doi:10.1007/978-3-322-83813-1] 26, 27, 61

[52] JOSEPH M. LANDSBERG: New lower bounds for the rank of matrix multiplication. *SIAM J. Comput.*, 43(1):144–149, 2014. [doi:10.1137/120880276] 83

[53] JOSEPH M. LANDSBERG: Geometric complexity theory: an introduction for geometers. *Annali dell' Università di Ferrara*, 61(1):65–117, 2015. [doi:10.1007/s11565-014-0202-7] 6

[54] JOSEPH M. LANDSBERG: *Geometry and Complexity Theory*. Cambridge Univ. Press, 2017. [doi:10.1017/9781108183192] 105

[55] JOSEPH M. LANDSBERG, LAURENT MANIVEL, AND NICOLAS RESSAYRE: Hypersurfaces with degenerate duals and the Geometric Complexity Theory Program. *Comment. Math. Helvetici*, 88(2):469–484, 2013. [doi:10.4171/CMH/292] 54

[56] JOSEPH M. LANDSBERG AND MATEUSZ MICHAŁEK: A $2n^2 - \log_2(n) - 1$ lower bound for the border rank of matrix multiplication. *Internat. Math. Research Notices*, 2018(15):4722–4733, 2018. [doi:10.1093/imrn/rnx025] 135

[57] JOSEPH M. LANDSBERG AND GIORGIO OTTAVIANI: New lower bounds for the border rank of matrix multiplication. *Theory of Computing*, 11(11):285–298, 2015. [doi:10.4086/toc.2015.v011a011] 135

[58] MICHAEL J. LARSEN AND RICHARD PINK: Determining representations from invariant dimensions. *Inventiones Math.*, 102(2):377–398, 1990. [doi:10.1007/BF01233432] 121

[59] MEENA MAHAJAN AND NITIN SAURABH: Some complete and intermediate polynomials in algebraic complexity theory. *Theory Computing Sys.*, 62(3):622–652, 2018. [doi:10.1007/s00224-016-9740-y] 30, 37

[60] MEENA MAHAJAN AND V. VINAY: Determinant: Combinatorics, algorithms, and complexity. *Chicago J. Theoret. Comp. Sci.*, 1997(5), 1997. [doi:10.4086/cjtcs.1997.005] 44

[61] GUILLAUME MALOD AND NATACHA PORTIER: Characterizing Valiant's algebraic complexity classes. *J. Complexity*, 24(1):16–38, 2008. [doi:10.1016/j.jco.2006.09.006] 38, 40

[62] KETAN D. MULMULEY: On P vs. NP and geometric complexity theory: Dedicated to Sri Ramakrishna. *J. ACM*, 58(2):5:1–26, 2011. [doi:10.1145/1944345.1944346] 6

[63] KETAN D. MULMULEY AND MILIND SOHONI: Geometric complexity theory. I. An approach to the P vs. NP and related problems. *SIAM J. Comput.*, 31(2):496–526, 2001. [doi:10.1137/S009753970038715X] 5

[64] KETAN D. MULMULEY AND MILIND SOHONI: Geometric complexity theory II: Towards explicit obstructions for embeddings among class varieties. *SIAM J. Comput.*, 38(3):1175–1206, 2008. [doi:10.1137/080718115] 5

[65] DAVID B. MUMFORD: *Algebraic Geometry I: Complex Projective Varieties*. Classics in mathematics. Springer, 1995. Reprint of the 1976 edition in Grundlehren der mathematischen Wissenschaften, vol. 221. Available at Springer. 27

[66] KYO NISHIYAMA: Restriction of the irreducible representations of GL_n to the symmetric group $\mathfrak{S}_n$. Unpublished preprint, online available at Kyoto University. 152

[67] ALEXANDER OSTROWSKI: On two problems in abstract algebra connected with Horner's rule. In *Studies in Mathematics and Mechanics presented to Richard von Mises*, pp. 40–48. Academic Press, 1954. [doi:10.1016/B978-1-4832-3272-0.50010-7] 5

[68] KENNETH W. REGAN: Understanding the Mulmuley–Sohoni approach to P vs NP. *Bulletin of the EATCS*, 78:86–99, 2002. Available at author's website. 6

[69] JEFFREY B. REMMEL: A formula for the Kronecker products of Schur functions of hook shapes. *J. Algebra*, 120(1):100–118, 1989. [doi:10.1016/0021-8693(89)90191-9] 135

[70] MERCEDES H. ROSAS: The Kronecker product of Schur functions indexed by two-row shapes or hook shapes. *J. Algebraic Combinatorics*, 14(2):153–173, 2001. [doi:10.1023/A:1011942029902] 135, 157

[71] MARCUS SCHAEFER AND DANIEL ŠTEFANKOVIČ: The complexity of tensor rank. *Theory Computing Sys.*, 62(5):1161–1174, 2018. [doi:10.1007/s00224-017-9800-y] 79

[72] YAROSLAV SHITOV: How hard is tensor rank?, 2016. [arXiv:1611.01559] 79

[73] ZHAO SONG, DAVID P. WOODRUFF, AND PEILIN ZHONG: Relative error tensor low rank approximation. In *Proc. 30th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'19)*, pp. 2772–2789. SIAM, 2019. [doi:10.1137/1.9781611975482.172] 79

[74] RICHARD P. STANLEY: Positivity problems and conjectures in algebraic combinatorics. In *Mathematics: Frontiers and perspectives*, pp. 295–319. Amer. Math. Soc., 2000. Available at author's website. 75

[75] VOLKER STRASSEN: Gaussian elimination is not optimal. *Numerische Mathematik*, 13:354–356, 1969. [doi:10.1007/BF02165411] 80

[76] VOLKER STRASSEN: Polynomials with rational coefficients which are hard to compute. *SIAM J. Comput.*, 3(2):128–149, 1974. [doi:10.1137/0203010] 5, 27

[77] VOLKER STRASSEN: Relative bilinear complexity and matrix multiplication. *J. reine angew. Math.*, 375–376(4):406–443, 1987. Available at *Eu*DML. 90

[78] JOSEPH SWERNOFSKY: Tensor rank is hard to approximate. In *Proc. 21st Internat. Conf. on Approximation Algorithms for Combinat. Opt. Probl. (APPROX'18)*, pp. 26:1–9. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018. [doi:10.4230/LIPIcs.APPROX-RANDOM.2018.26] 79

[79] PATRICE TAUVEL AND RUPERT W. T. YU: *Lie Algebras and Algebraic Groups*. Springer, 2005. [doi:10.1007/b139060] 26

[80] LESLIE G. VALIANT: Completeness classes in algebra. In *Proc. 11th STOC*, pp. 249–261. ACM Press, 1979. [doi:10.1145/800135.804419] 5, 122

[81] LESLIE G. VALIANT: The complexity of computing the permanent. *Theoret. Comput. Sci.*, 8(2):189–201, 1979. [doi:10.1016/0304-3975(79)90044-6] 10

[82] LESLIE G. VALIANT, SVEN SKYUM, STUART J. BERKOWITZ, AND CHARLES RACKOFF: Fast parallel computation of polynomials using few processors. *SIAM J. Comput.*, 12(4):641–644, 1983. [doi:10.1137/0212043] 31

[83] JUN YU: On the dimension datum of a subgroup. *Duke Math. J.*, 165(14):2683–2736, 2016. [doi:10.1215/00127094-3619898] 121

AUTHORS

Markus Bläser
Professor
Department of Computer Science
Saarland University
Saarland Informatics Campus
Saarbrücken, Germany
mblaeser@cs.uni-saarland.de
http://www-cc.cs.uni-saarland.de/mblaeser

Christian Ikenmeyer
Professor
Department of Computer Science and Warwick Mathematics Institute
University of Warwick
Coventry, United Kingdom
christian.ikenmeyer@warwick.ac.uk
https://warwick.ac.uk/fac/sci/maths/people/staff/ikenmeyer/

ABOUT THE AUTHORS

MARKUS BLÄSER is a full professor of computer science at Saarland University, Saarbrücken, Germany, since 2005. Before joining Saarbrücken, he was an assistant professor at ETH Zürich from 2003 to 2005. He obtained his Ph. D. in 1999 from the University of Bonn under the supervision of Arnold Schönhage and his Habilitation from the University of Lübeck under the supervision of Rüdiger Reischuk in 2003. Together with Peter Bro Miltersen, he created the Workshop on Algebraic Complexity Theory series (WACT), while watching a backgammon game between Anna Gál and Kristoffer Arnsfelt Hansen and drinking red wine at the Dagstuhl-Seminar "Computational Complexity of Discrete Problems" in 2011.

CHRISTIAN IKENMEYER has been a full professor in computer science and mathematics at the University of Warwick, United Kingdom, since 2022. He was a senior lecturer at the University of Liverpool from 2019 until 2022, and a visiting assistant professor at Texas A&M University from 2013 until 2016. In Saarbrücken he was a senior researcher at the Max Planck Institute for Informatics and a research group leader at the Max Planck Institute for Software Systems from 2016 to 2019, where he met Markus Bläser. He obtained his Ph. D. from the University of Paderborn under the supervision of Peter Bürgisser in 2012. This survey was written during his appointments in Saarbrücken and Liverpool.