Geometry and Complexity Theory

J.M. Landsberg

Contents

Preface

The purpose of this book is to describe recent applications of algebraic geometry and representation theory to complexity theory. I focus on two central problems: *the complexity of matrix multiplication* and *Valiant's algebraic variants of* **P** *v.* **NP**.

I have attempted to make this book accessible to both computer scientists and geometers, and the exposition as self-contained as possible. The two main goals of this book are to convince computer scientists of the utility of techniques from algebraic geometry and representation theory, and to show geometers beautiful, interesting, and important questions arising in complexity theory.

Computer scientists have made extensive use of tools from mathematics such as combinatorics, graph theory, probability, and especially linear algebra. I hope to show that even elementary techniques from algebraic geometry and representation theory can substantially advance the search for lower, and even upper bounds in complexity theory. For questions such as lower bounds for the complexity of matrix multiplication and Valiant's algebraic variants of **P** v. **NP**, I believe this additional mathematics will be necessary for further advances. I have attempted to make these techniques accessible, introducing them as needed to deal with concrete problems.

For geometers, I expect that complexity theory will be as good a source for questions in algebraic geometry as modern physics has been. Recent work has indicated that subjects such as Fulton-McPherson intersection theory, the Hilbert scheme of points, and the Kempf-Weyman method for computing minimal free resolutions all have something to add to complexity theory. In addition, complexity theory has a way of rejuvenating old questions that had been nearly forgotten but remain beautiful and intriguing: questions of

Hadamard, Darboux, Luroth, and the classical Italian school. At the same time, complexity theory has brought different areas of mathematics together in new ways- combinatorics, representation theory and algebraic geometry all play a role in understanding the coordinate ring of the orbit closure of the determinant.

This book evolved from several classes I have given on the subject: a spring 2013 semester course at Texas A&M, summer courses at: Sculoa Matematica Inter-universitaria, Cortona (July 2012), CIRM, Trento (June 2014), and an IMA summer school at U. Chicago (July 2014), KAIST (August 2015), a fall 2016 semester course at Texas A&M, and most importantly, a fall 2014 semester course at UC Berkeley as part of the semester long program, *Algorithms and Complexity in Algebraic Geometry*, at the Simons Institute for the Theory of Computing.

Overview. To be written

Prerequisites. I have attempted to limit prerequisites to a solid background in linear algebra, although such a reader would have to accept several basic results in algebraic geometry without proof (e.g. Noether normalization). In Chapter 6 some further, but still elementary algebraic geometry is needed, but nothing beyond [**Sha94**] is used. Starting with Chapter 9, some advanced results from algebraic geometry are needed.

Acknowledgments. To be written later

Layout. To be written later

Dependency of chapters. To be written later

Chapter 1

Introduction

A dramatic leap in signal processing occurred in the 1960's with the implementation of the fast Fourier transform, an algorithm that surprised the engineering community with its efficiency.[1] How could one predict when fast, perhaps non-intuitive, algorithms exist? Can we prove when they do not? *Complexity theory* addresses these questions.

This book is concerned with the use of *geometry* in attaining these goals. I focus primarily on two central questions: the complexity of matrix multiplication, and algebraic variants of the famous **P** versus **NP** problem. In the first case, a surprising algorithm exists and it is conjectured that even more amazing algorithms exist. In the second case it is conjectured that no surprising algorithms exist.

1.1. Matrix multiplication

Much of scientific computation is linear algebra, and the basic operation of linear algebra is matrix multiplication. All operations of linear algebra; solving systems of linear equations, computing determinants etc., use matrix multiplication.

1.1.1. The standard algorithm. The standard algorithm for multiplying matrices is row-column multiplication: Let A, B be 2×2 matrices

$$A = \begin{pmatrix} a^1_1 & a^1_2 \\ a^2_1 & a^2_2 \end{pmatrix}, \quad B = \begin{pmatrix} b^1_1 & b^1_2 \\ b^2_1 & b^2_2 \end{pmatrix}.$$

[1]To this day, it is not known if there is an even more efficient algorithm than the FFT. See [**Val77, KLPSMN09, GHIL**].

Remark 1.1.1.1. While computer scientists generally keep all indices down (to distinguish from powers), I use the convention from differential geometry that in a matrix X, the entry in the i-th row and j-th column is labeled x^i_j.

The usual algorithm to calculate the matrix product $C = AB$ is

$$\begin{aligned}
c^1_1 &= a^1_1 b^1_1 + a^1_2 b^2_1,\\
c^1_2 &= a^1_1 b^1_2 + a^1_2 b^2_2,\\
c^2_1 &= a^2_1 b^1_1 + a^2_2 b^2_1,\\
c^2_2 &= a^2_1 b^1_2 + a^2_2 b^2_2.
\end{aligned}$$

It requires 8 multiplications and 4 additions to execute, and applied to $\mathbf{n}\times\mathbf{n}$ matrices, it uses $\mathbf{n}^3$ multiplications and $\mathbf{n}^3 - \mathbf{n}^2$ additions.

This algorithm has been around for a long time.

In 1968, V. Strassen set out to prove the standard algorithm was optimal in the sense that no algorithm using fewer multiplications exists. Since that might be difficult to prove, he set out to show it was true at least for two by two matrices – at least over $\mathbb{Z}_2$. His spectacular failure opened up a whole new area of research:

1.1.2. Strassen's algorithm for multiplying 2×2 matrices using seven scalar multiplications [Str69]. Set

$$\begin{aligned}
I &= (a^1_1 + a^2_2)(b^1_1 + b^2_2),\\
II &= (a^2_1 + a^2_2)b^1_1,\\
III &= a^1_1(b^1_2 - b^2_2)\\
IV &= a^2_2(-b^1_1 + b^2_1)\\
V &= (a^1_1 + a^1_2)b^2_2\\
VI &= (-a^1_1 + a^2_1)(b^1_1 + b^1_2),\\
VII &= (a^1_2 - a^2_2)(b^2_1 + b^2_2),
\end{aligned} \tag{1.1.1}$$

Exercise 1.1.2.1: (1) Show that if $C = AB$, then

$$\begin{aligned}
c^1_1 &= I + IV - V + VII,\\
c^2_1 &= II + IV,\\
c^1_2 &= III + V,\\
c^2_2 &= I + III - II + VI.
\end{aligned}$$

This raises questions:

(1) Can one find an algorithm that uses just six multiplications?

(2) Could Strassen's algorithm have been predicted in advance?

(3) Since it uses more additions, is it actually better in practice?

(4) This algorithm was found by accident and looks ad-hoc. Is there any way to make sense of it? E.g., is there any way to see that it multiplies matrices other than a brute force calculation?

(5) What about algorithms for $\mathbf{n} \times \mathbf{n}$ matrices?

I address the last question first:

1.1.3. Fast multiplication of $\mathbf{n} \times \mathbf{n}$ matrices. In Strassen's algorithm, the entries of the matrices need not be scalars - they could themselves be matrices. Let A, B be 4×4 matrices, and write

$$A = \begin{pmatrix} a^1_1 & a^1_2 \\ a^2_1 & a^2_2 \end{pmatrix}, \quad B = \begin{pmatrix} b^1_1 & b^1_2 \\ b^2_1 & b^2_2 \end{pmatrix}.$$

where a^i_j, b^i_j are 2×2 matrices. One may apply Strassen's algorithm to get the blocks of $C = AB$ in terms of the blocks of A, B performing 7 multiplications of 2×2 matrices. Since one can apply Strassen's algorithm to each block, one can multiply 4×4 matrices using $7^2 = 49$ multiplications instead of the usual $4^3 = 64$.

If A, B are $2^k \times 2^k$ matrices, one may multiply them using 7^k multiplications instead of the usual 8^k. If $\mathbf{n}$ is not a power of two, enlarge the matrices with blocks of zeros to obtain matrices whose size is a power of two. Asymptotically, by recursion and block multiplication one can multiply $\mathbf{n} \times \mathbf{n}$ matrices using approximately $\mathbf{n}^{\log_2(7)} \simeq \mathbf{n}^{2.81}$ arithmetic operations. To see this, let $\mathbf{n} = 2^k$ and write $7^k = (2^k)^a$ so $k \log_2 7 = ak \log_2 2$ so $a = \log_2 7$.

1.1.4. Regarding the number of additions. The number of additions in Strassen's algorithm also grows like $\mathbf{n}^{2.81}$, so this algorithm *is* more efficient in practice when the matrices are large. For any efficient algorithm for matrix multiplication, the total complexity is governed by the number of multiplications, see [**BCS97**, Prop. 15.1]. This is fortuitous because there is a geometric object, *tensor rank*, that counts the number of multiplications in an optimal algorithm (within a factor of two), and thus provides us with a geometric measure of the complexity of matrix multiplication.

Just how large matrices one needs to obtain a substantial savings with Strassen's algorithm (one needs matrices of size about two thousand) and other practical matters are addressed in [**BB**].

1.1.5. An even better algorithm? Regarding question (1) above, one cannot improve upon Strassen's algorithm for 2×2 matrices. This was first shown in [**Win71**]. I will give a proof, using geometry and representation theory, of a stronger statement in §8.3.2. However for $\mathbf{n} > 2$ very little is known, as is discussed below and in Chapters 2-5. It is known that better algorithms than Strassen's exist for $\mathbf{n} \times \mathbf{n}$ matrices when $\mathbf{n}$ is large, even if they are not written down explicitly.

1.1.6. How to predict in advance? The answer to question (2) is yes! In fact it could have been predicted 100 years ago.

Had someone asked Terracini in 1913, he would have been able to predict the existence of something like Strassen's algorithm from geometric considerations alone. Matrix multiplication is a bilinear map (see §1.1.9). Terracini would have been able to tell you, thanks to a simple parameter count (see §2.1.6), that even a general bilinear map $\mathbb{C}^4 \times \mathbb{C}^4 \to \mathbb{C}^4$ can be executed using seven multiplications and thus, fixing any $\epsilon > 0$, one can perform any bilinear map $\mathbb{C}^4 \times \mathbb{C}^4 \to \mathbb{C}^4$ within an error of ϵ using seven multiplications.

1.1.7. Conventions/Notation. In this book, for simplicity, I work exclusively over the complex numbers.

For functions f, g of a real variable x: $f(x) = O(g(x))$ if there exists a constant $C > 0$ and x_0 such that $|f(x)| \leq C|g(x)|$ for all $x \geq x_0$. $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{|f(x)|}{|g(x)|} = 0$, $f(x) = \Omega(g(x))$ if there exists a constant $C > 0$ and x_0 such that $C|f(x)| \geq |g(x)|$ for all $x \geq x_0$, and $f(x) = \Theta(g(x))$ if $f(x) = O(g(x))$ and $f(x) = \Omega(g(x))$.

1.1.8. An astonishing conjecture. The following quantity is the standard measure of the complexity of matrix multiplication:

Definition 1.1.8.1. The *exponent* ω of matrix multiplication is

$$\begin{aligned}\omega := \inf\{h \in \mathbb{R} \mid\ & \mathbf{n} \times \mathbf{n} \text{ matrices may be multiplied using} \\ & O(\mathbf{n}^h) \text{ arithmetic operations}\}\end{aligned}$$

where inf denotes the infimum.

By Theorem 1.1.11.3 below, Strassen's algorithm shows $\omega \leq \log_2(7) < 2.81$, and it is easy to prove $\omega \geq 2$. Determining ω is a central open problem in complexity theory. After Strassen's work it was shown $\omega \leq 2.79$ [**Bin80**] in 1979, then $\omega \leq 2.55$ [**Sch81**] in 1981, then $\omega \leq 2.48$[**Str87**] in 1987 and then $\omega \leq 2.38$ [**CW90**] in 1989, which might have led people in 1990 to think a resolution was near. However, then nothing happened for over twenty years, and the current "world record" of $\omega < 2.373$ [**Wil, Gal, Sto**] is not much of an improvement since 1990. These results are the topic of Chapter 3.

This work has led to the following astounding conjecture:

Conjecture 1.1.8.2. $\omega = 2$.

That is, *it is conjectured that asymptotically, it is nearly just as easy to multiply matrices as it is to add them*!

Although I am unaware of anyone taking responsibility for the conjecture, all computer scientists I have discussed it with expect it to be true.

Since I have no opinion on whether the conjecture should be true or false, I discuss both upper and lower bounds for the complexity of matrix multiplication, focusing on the role of geometry.

1.1.9. Matrix multiplication as a bilinear map. I will use the notation

$$M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle} : \mathbb{C}^{\mathbf{n}\times\mathbf{m}} \times \mathbb{C}^{\mathbf{m}\times\mathbf{l}} \to \mathbb{C}^{\mathbf{n}\times\mathbf{l}}$$

for matrix multiplication of an $\mathbf{n}\times\mathbf{m}$ matrix with an $\mathbf{m}\times\mathbf{l}$ matrix, and write $M_{\langle \mathbf{n}\rangle} = M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$.

Matrix multiplication is a *bilinear map*, that is, for all $X_j, X \in \mathbb{C}^{\mathbf{n}\times\mathbf{m}}$, $Y_j, Y \in \mathbb{C}^{\mathbf{m}\times\mathbf{l}}$ and $a_j, b_j \in \mathbb{C}$,

$$M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}(a_1X_1 + a_2X_2, Y) = a_1M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}(X_1, Y) + a_2M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}(X_2, Y), \quad \text{and}$$
$$M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}(X, b_1Y_1 + b_2Y_2) = b_1M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}(X, Y_1) + b_2M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}(X, Y_2).$$

The set of all bilinear maps $\mathbb{C}^{\mathbf{a}} \times \mathbb{C}^{\mathbf{b}} \to \mathbb{C}^{\mathbf{c}}$ is a vector space. (In our case $\mathbf{a} = \mathbf{nm}$, $\mathbf{b} = \mathbf{ml}$, and $\mathbf{c} = \mathbf{ln}$.) Write $a_1, \dots, a_{\mathbf{a}}$ for a basis of $\mathbb{C}^{\mathbf{a}}$ and similarly for $\mathbb{C}^{\mathbf{b}}, \mathbb{C}^{\mathbf{c}}$. Then $T : \mathbb{C}^{\mathbf{a}} \times \mathbb{C}^{\mathbf{b}} \to \mathbb{C}^{\mathbf{c}}$ is uniquely determined by its action on basis vectors,

$$T(a_i, b_j) = \sum_{k=1}^{\mathbf{c}} t^{ijk} c_k. \tag{1.1.2}$$

That is, the vector space of bilinear maps $\mathbb{C}^{\mathbf{a}} \times \mathbb{C}^{\mathbf{b}} \to \mathbb{C}^{\mathbf{c}}$, which I will denote by $\mathbb{C}^{\mathbf{a}*}\otimes\mathbb{C}^{\mathbf{b}*}\otimes\mathbb{C}^{\mathbf{c}}$, has dimension $\mathbf{abc}$. (The notation is motivated in §2.1.) If we represent a bilinear map by a three dimensional matrix, it may be thought of as eating two column vectors and returning a third column vector.

1.1.10. Ranks of linear maps. I use the notation $\mathbb{C}^{\mathbf{a}}$ for the column vectors of height $\mathbf{a}$ and $\mathbb{C}^{\mathbf{a}*}$ for the row vectors.

Definition 1.1.10.1. A linear map $f : \mathbb{C}^{\mathbf{a}} \to \mathbb{C}^{\mathbf{b}}$ has *rank one* if there exist $\alpha \in \mathbb{C}^{\mathbf{a}*}$ and $w \in \mathbb{C}^{\mathbf{b}}$ such that $f(v) = \alpha(v)w$. (In other words, every rank one matrix is the product of a row vector with a column vector.) In this case I write $f = \alpha\otimes w$. The *rank* of a linear map $h : \mathbb{C}^{\mathbf{a}} \to \mathbb{C}^{\mathbf{b}}$ is the smallest r such that h may be expressed as a sum of r rank one linear maps.

Given an $\mathbf{a}\times\mathbf{b}$ matrix X, one can always change bases, i.e., multiply X on the left by an invertible $\mathbf{a}\times\mathbf{a}$ matrix and on the right by an invertible $\mathbf{b}\times\mathbf{b}$ matrix, to obtain a matrix with some number of 1's along the diagonal and zeros elsewhere. The number of 1's appearing is called the *rank* of the matrix and is the rank of the linear map X determines. In other words, the only property of a linear map $\mathbb{C}^{\mathbf{a}} \to \mathbb{C}^{\mathbf{b}}$ that is invariant under changes of bases is its rank, and for each rank we have a normal form. This is not surprising because the dimension of the space of such linear maps is $\mathbf{ab}$, we have $\mathbf{a}^2$ parameters of changes of bases in $\mathbb{C}^{\mathbf{a}}$ that we can make in a matrix representing the map, and $\mathbf{a}^2 + \mathbf{b}^2 > \mathbf{ab}$. Another way of saying a matrix X has rank at most r is that it is possible to write X as the sum of r rank one matrices.

1.1.11. Tensor rank. For bilinear maps $\mathbb{C}^{\mathbf{a}} \times \mathbb{C}^{\mathbf{b}} \to \mathbb{C}^{\mathbf{c}}$ we are not so lucky as with linear maps, as usually $\mathbf{abc} > \mathbf{a}^2 + \mathbf{b}^2 + \mathbf{c}^2$, i.e., there are fewer free parameters of changes of bases than the number of parameters needed to describe the map. This already indicates why the study of bilinear maps is vastly more complicated than the study of linear maps.

Nonetheless, there are properties of a bilinear map that will not change under a change of basis. The main property we will use is *tensor rank*. It is a generalization of the rank of a linear map. Tensor rank is defined properly in §2.1.3. Informally, a bilinear map T has *tensor rank one* if it can be computed with one multiplication. More precisely, T has tensor rank one if in some coordinate system the multi-dimensional matrix representing it has exactly one nonzero entry. This may be expressed without coordinates:

Definition 1.1.11.1. $T \in \mathbb{C}^{\mathbf{a}*} \otimes \mathbb{C}^{\mathbf{b}*} \otimes \mathbb{C}^{\mathbf{c}}$ has *tensor rank one* if there exist row vectors $\alpha \in \mathbb{C}^{\mathbf{a}*}$, $\beta \in \mathbb{C}^{\mathbf{b}*}$ and a column vector $w \in \mathbb{C}^{\mathbf{c}}$ such that $T(u, v) = \alpha(u)\beta(v)w$. T has *tensor rank* r if it can be written as the sum of r rank one tensors but no fewer, in which case we write $\mathbf{R}(T) = r$. Let $\hat{\sigma}_r^0 = \hat{\sigma}^0_{r,\mathbf{a},\mathbf{b},\mathbf{c}}$ denote the set of bilinear maps in $\mathbb{C}^{\mathbf{a}*} \otimes \mathbb{C}^{\mathbf{b}*} \otimes \mathbb{C}^{\mathbf{c}}$ of tensor rank at most r.

Remark 1.1.11.2. The peculiar notation $\hat{\sigma}_r^0$ will be explained in §4.8.1. To have an idea where it comes from for now: $\sigma_r = \sigma_r(Seg(\mathbb{P}^{\mathbf{a}-1} \times \mathbb{P}^{\mathbf{b}-1} \times \mathbb{P}^{\mathbf{c}-1}))$ is standard notation in algebraic geometry for the r-th secant variety of the Segre variety, which is the object we will study. The hat denotes its cone in affine space and the 0 indicates the subset of this set consisting of tensors of rank at most r.

The following theorem shows that tensor rank is a legitimate measure of complexity:

Theorem 1.1.11.3. *(Strassen* [**Str69**], *also see* [**BCS97**, §15.1] *)* $\mathbf{R}(M_{\langle \mathbf{n} \rangle}) = O(n^{\omega})$.

Our goal is thus to determine, for a given r, whether or not matrix multiplication lies in $\hat{\sigma}_r^0$.

1.1.12. How to use algebraic geometry to prove lower bounds for the complexity of matrix multiplication? *Algebraic geometry* deals with the study of zero sets of polynomials. By a polynomial on the space of bilinear maps $\mathbb{C}^{\mathbf{a}^*}\otimes\mathbb{C}^{\mathbf{b}*}\otimes\mathbb{C}^{\mathbf{c}}$, I mean a polynomial in the coefficients t^{ijk}, i.e., in **abc** variables. Algebraic geometry may be used to prove both upper and lower complexity bounds. For lower bounds:

Plan to show $M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle} \notin \hat{\sigma}_r^0$ via algebraic geometry.

- Find a polynomial P on the space of bilinear maps $\mathbb{C}^{\mathbf{nm}} \times \mathbb{C}^{\mathbf{ml}} \to \mathbb{C}^{\mathbf{nl}}$, such that $P(T) = 0$ for all $T \in \hat{\sigma}_r^0$.
- Show that $P(M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}) \neq 0$.

Chapters 2 and 5 discuss techniques for finding such polynomials, using algebraic geometry and *representation theory*, the study of symmetry in linear algebra.

1.1.13. Representation theory. Representation theory is the systematic study of symmetry in linear algebra. The study of polynomials is facilitated by sorting the polynomials by degree. When the objects one is interested in have symmetry, one can make a finer sorting of polynomials. This finer sorting has been essential for proving lower bounds for the complexity of $M_{\langle \mathbf{n}\rangle}$.

We will frequently be concerned with properties of bilinear maps, tensors, polynomials, etc.. that are invariant under changes of bases. Representation theory will facilitate the exploitation of these properties.

Let V be a complex vector space of dimension $\mathbf{v}$. (I reserve the notation $\mathbb{C}^{\mathbf{v}}$ for the column vectors with their standard basis.) Let $GL(V)$ denote the group of invertible linear maps $V \to V$. If we have fixed a basis of V, this is the group of invertible $\mathbf{v}\times\mathbf{v}$ matrices. If G is a group and $\mu : G \to GL(V)$ is a group homomorphism, we will say G *acts on* V and that V is a G*-module*.

For example the permutation group on n elements $\mathfrak{S}_n$ acts on $\mathbb{C}^n$ by, for a permutation $\sigma \in \mathfrak{S}_n$,

$$\sigma \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_{\sigma^{-1}(1)} \\ \vdots \\ v_{\sigma^{-1}(n)} \end{pmatrix}$$

i.e., the image of $\mathfrak{S}_n$ in GL_n is the set of permutation matrices.

An action is *irreducible* if there does not exist a proper subspace $U \subset V$ such that $\mu(g)u \in U$ for all $u \in U$ and $g \in G$.

The action of $\mathfrak{S}_n$ on $\mathbb{C}^n$ is reducible since the line spanned by $e_1+\cdots+e_n$ is preserved by $\mathfrak{S}_n$. Note that the subspace spanned by $e_1-e_2,\dots,e_1-e_n$ is also preserved by $\mathfrak{S}_n$. Both these $\mathfrak{S}_n$-modules are irreducible. For reasons that will be explained in §8.6, the first is denoted $[n]$ and the second is denoted $[n-1,n]$

For another example of a group action, the group $GL(V)$ acts on the space $\mathrm{End}(V)$ of linear maps $V\to V$, by $\mu_{\mathrm{End}(V)}(g)(f)=g\circ f\circ g^{-1}$, i.e., $\mu_{\mathrm{End}(V)}(g)(f)(v)=g(f(g^{-1}(v)))$. It also acts on the space of bilinear forms $V\times V\to\mathbb{C}$, which I will denote $V^*{\otimes}V^*$, by $\mu_{V^*\otimes V^*}(g)(b)(v,w)=b(gv,gw)$. Note that if we choose a basis of V, then both $\mathrm{End}(V)$ and $V^*{\otimes}V^*$ are represented by the space of $\mathbf{v}\times\mathbf{v}$ matrices. However the group actions are very different. In the first case, the action on a matrix X is $X\mapsto gXg^{-1}$. In the second the action on a matrix Y (so the map is $(v,w)\mapsto v^TYw$) is $Y\mapsto g^TYg$. There is a dramatic difference in the two spaces as $GL(V)$-modules.

The essential point we will use is: the sets we are looking for polynomials on, such as $X=\hat{\sigma}_r^0\subset\mathbb{C}^{\mathbf{abc}}$ are *invariant* under the action of groups:

Definition 1.1.13.1. A set $X\subset V$ is *invariant* under a group $G\subset GL(V)$ if for all $x\in X$ and all $g\in G$, $g(x)\in X$. Let $G_X\subset GL(V)$ denote the *group preserving* X, the largest subgroup of $GL(V)$ under which X is invariant.

When one says that an object has symmetry, it means the object is invariant under the action of a group.

In the case at hand, $X=\hat{\sigma}_r^0\subset V=A{\otimes}B{\otimes}C$. Then $\hat{\sigma}_r^0$ is invariant under the image of the group $GL(A)\times GL(B)\times GL(C)$ in $GL(V)$, i.e., this image lies in $G_{\hat{\sigma}_r^0}$.

Definition 1.1.13.2. For a set $X\subset V$, we will say a polynomial P *vanishes* on X if $P(x)=0$ for all $x\in X$. The set of all polynomials vanishing on X forms an ideal in the space of polynomials on V, called the *ideal* of X and denoted $I(X)$.

If any polynomial P is in the ideal of X, then $g\cdot P$ will also vanish on X for all $g\in G_X$. That is:

The ideal of polynomials vanishing on X is a G_X-module.

This remark is the cornerstone to this book.

1.1.14. How to use algebraic geometry to prove upper bounds for the complexity of matrix multiplication? Based on the above discussion, one could try:

Plan to show $M_{\langle\mathbf{n},\mathbf{m},\mathbf{l}\rangle}\in\hat{\sigma}_r^0$ with algebraic geometry.

- Find a set of polynomials $\{P_j\}$ on the space of bilinear maps $\mathbb{C}^{\mathbf{nm}} \times \mathbb{C}^{\mathbf{ml}} \to \mathbb{C}^{\mathbf{nl}}$ such that $T \in \hat{\sigma}_r^0$ if and only if $P_j(T) = 0$ for all j.
- Show that $P_j(M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}) = 0$ for all j.

This plan has a problem: Consider the set $S = \{(w,z) \in \mathbb{C}^2 \mid z = 0, w \neq 0\}$, whose real picture looks like the z-axis with the origin removed.

Any polynomial $P \in I(S)$, i.e., any P that evaluates to zero at all points of S, will also be zero at the origin.

Exercise 1.1.14.1: (1!) Prove the above assertion.

Just as in this example, the zero set of the polynomials vanishing on $\hat{\sigma}_r^0$ is larger than $\hat{\sigma}_r^0$ when $r > 1$ (see §2.1.5) so one cannot certify membership in $\hat{\sigma}_r^0$ via polynomials.

Definition 1.1.14.2. Define the *Zariski closure* of a set $S \subset V$, denoted $\overline{S}$, to be the set of $u \in V$ such that $P(u) = 0$ for all $P \in I(S)$. A set S is said to be *Zariski closed* or an *algebraic variety* if $S = \overline{S}$, i.e., S is the common zero set of a collection of polynomials.

In the example above, $\overline{S} = \{(w,z) \in \mathbb{C}^2 \mid z = 0\}$.

When $U = \mathbb{C}^{\mathbf{a}*} \otimes \mathbb{C}^{\mathbf{b}*} \otimes \mathbb{C}^{\mathbf{c}}$, let $\hat{\sigma}_r := \overline{\hat{\sigma}_r^0}$ denote the Zariski closure of the set of bilinear maps of tensor rank at most r.

We will see that for almost all $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and r, $\hat{\sigma}_r^0 \subsetneq \hat{\sigma}_r$. The problem with the above plan is that it would only show $M_{\langle \mathbf{n} \rangle} \in \hat{\sigma}_r$.

Definition 1.1.14.3. $T \in \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{b}} \otimes \mathbb{C}^{\mathbf{c}}$ has *border rank* r if $T \in \hat{\sigma}_r$ and $T \notin \hat{\sigma}_{r-1}$. In this case we write $\underline{\mathbf{R}}(T) = r$.

For the study of the exponent of matrix multiplication, we have good luck:

Theorem 1.1.14.4 (Bini [**Bin80**], see §3.2)**.** $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) = O(n^\omega)$.

That is, although we may have $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) < \mathbf{R}(M_{\langle \mathbf{n} \rangle})$, they are not different enough to effect the exponent. In other words, as far as the exponent is concerned, the plan does *not* have a problem.

For $\mathbf{n} = 2$, we will see that $\underline{\mathbf{R}}(M_{\langle 2 \rangle}) = \mathbf{R}(M_{\langle 2 \rangle}) = 7$. It is expected that for $\mathbf{n} > 2$, $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) < \mathbf{R}(M_{\langle \mathbf{n} \rangle})$. For $\mathbf{n} = 3$ we only know $15 \leq \underline{\mathbf{R}}(M_{\langle 3 \rangle}) \leq 20$ and $19 \leq \mathbf{R}(M_{\langle 3 \rangle}) \leq 23$. In general, we know $\mathbf{R}(M_{\langle \mathbf{n} \rangle}) \geq 3\mathbf{n}^2 - o(\mathbf{n})$, see §2.7, and $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \geq 2\mathbf{n}^2 - \lceil \log_2(\mathbf{n}) \rceil - 1$, see §5.4.3.

1.1.15. Symmetry and algorithms. In this subsection I mention three uses of symmetry groups in the study of algorithms.

I first address the question raised in §1.1.2: Can we make sense of (1.1.1)? Just as the set $\hat{\sigma}_r$ has a symmetry group, the point $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$ also has a symmetry group that includes $GL_{\mathbf{l}} \times GL_{\mathbf{m}} \times GL_{\mathbf{n}}$. (Do not confuse this with $GL_{\mathbf{lm}} \times GL_{\mathbf{mn}} \times GL_{\mathbf{nl}}$ which preserves $\hat{\sigma}_r^0$.) If we let this group act on Strassen's algorithm for $M_{\langle 2\rangle}$, in general we get a new algorithm that also computes $M_{\langle 2\rangle}$. But perhaps the algorithm itself has symmetry.

It does, and the first step to seeing the symmetry is to put all three vector spaces on an equal footing. A linear map $f : A \to B$ determines a bilinear form $A \times B^* \to \mathbb{C}$ by $(a, \beta) \mapsto \beta(f(a))$. Similarly, a bilinear map $A \times B \to C$ determines a trilinear form $A \times B \times C^* \to \mathbb{C}$.

Exercise 1.1.15.1: (2!) Show that $M_{\langle \mathbf{n}\rangle}$, considered as a trilinear form, is $(X, Y, Z) \mapsto \mathrm{trace}(XYZ)$ ⊚

Since $\mathrm{trace}(XYZ) = \mathrm{trace}(YZX)$, we see that $G_{M_{\langle \mathbf{n}\rangle}}$ also includes a cyclic $\mathbb{Z}_3$-symmetry. In Chapter 4 we will see that *Strassen's algorithm is invariant under this* $\mathbb{Z}_3$*-symmetry*!

This hints that we might be able to use geometry to help *find* algorithms. This is the topic of Chapter 4.

For tensors or polynomials with continous symmetry, their algorithms come in *families*. So to prove lower bounds, i.e., non-existence of a family of algorithms, one can just prove non-existence of a special member of the family. This is key to the state of the art lower bound for matrix multiplication presented in §5.4.3. The general theory is discussed in §**??**.

A third use of geometry in algorithms is for the *restricted models* discussed below. There one creates a restricted model by imposing symmetry. This has led to the only exponential separation of permanent and determinant in any restricted model, see §7.4.7.

1.2. Separation of algebraic complexity classes

In 1950, John Nash (see [**NR16**, Chap. 1]) sent a letter to the NSA regarding cryptography, conjecturing an exponential increase in mean key computation length with respect to the length of the key. In a 1956 letter to von Neumann (see [**Sip92**, Appendix]) Gödel tried to quantify the apparent difference between intuition and systematic problem solving. Around the same time, researchers in the Soviet Union were trying to determine if "brute force search" was avoidable in solving problems such as the famous traveling salesman problem where there seems to be no fast way to find a solution, but a proposed solution can be easily checked, see [**Tra84**]. (The problem is to determine if there exists a way to visit, say twenty cities traveling less than a thousand miles. If I claim to have an algorithm to do so, you just need to look at my plan and check the distances.) These discussions eventually gave rise to the complexity classes **P**, which models problems admitting a fast algorithm to produce a solution, and **NP** which models problems admitting a fast algorithm to verify a proposed solution. The famous conjecture of Cook, Karp and Levin that these two classes are distinct. See [**Sip92**] for a history of the problem and [**NR16**, Chap. 1] for an up to date survey.

The transformation of this conjecture to a conjecture in geometry goes via algebra:

1.2.1. From complexity to algebra. The **P** v. **NP** conjecture is generally believed to be out of reach at the moment, so there have been weaker conjectures proposed that might be more tractable. One such comes from a standard counting problem discussed in §6.1.1. This variant has the advantage that it admits a clean algebraic formulation that I now discuss.

L. Valiant [**Val79a**] conjectured that a sequence of polynomials that is "easy" to write down should not necessarily admit a fast evaluation. He defined algebraic complexity classes that are now called **VP** and **VNP**, respectively the sequences of polynomials that are "easy" to evaulate, and the sequences that are "easy" to write down (see §6.1.3 for their definitions), and conjectured:

Conjecture 1.2.1.1 (Valiant [**Val79a**])**. VP $\neq$ VNP**.

For the precise relationship between this conjecture and the **P** $\neq$ **NP** conjecture see [**BCS97**, Chap. 21].

Many problems from graph theory, combinatorics, and statistical physics (partition functions) are in **VNP**. A good way to think of **VNP** is as the class of sequences of polynomials that can be written down "explicitly".

Most problems from linear algebra (e.g., inverting a matrix, computing its determinant, multiplying matrices) are in **VP**.

Valiant also showed that a particular polynomial sequence, the *permanent* (perm_n), is *complete* for the class **VNP**, in the sense that **VP** $\neq$ **VNP** if and only if $(\mathrm{perm}_n) \notin$ **VP**. As explained in §6.1.1, the permanent is natural for computer science. Although it is not immediately clear, the permanent is also natural to geometry, see §6.6.2. The formula for the permanent of an $n \times n$ matrix $x = (x^i_j)$ is:

$$\mathrm{perm}_n(x) := \sum_{\sigma \in \mathfrak{S}_n} x^1_{\sigma(1)} \cdots x^n_{\sigma(n)}. \tag{1.2.1}$$

Here $\mathfrak{S}_n$ denotes the group of permutations of $\{1, \dots, n\}$.

How would one show there is no fast algorithm for the permanent? In §6.1.3 we will define *algebraic circuits*, which are a class of algorithms for computing a polynomial, and their *size*, which is a measure of the complexity of the algorithm. Let circuit-size(perm_n) denote the size of the smallest algebraic circuit computing perm_n. Valiant's conjecture 1.2.1.1 may be rephrased as:

Conjecture 1.2.1.2 (Valiant [**Val79a**]). circuit-size(perm_n) *grows faster than any polynomial in* n.

1.2.2. From algebra to algebraic geometry. As with our earlier discussion, one could work as follows:

Let $S^n\mathbb{C}^N$ denote the vector space of all homogeneous polynomials of degree n in N variables, so perm_n is a point of the vector space $S^n\mathbb{C}^{n^2}$. If we write an element of $S^n\mathbb{C}^N$ as $p(y_1,\dots,y_N)=\sum_{1\le i_1\le\cdots\le i_n\le N} c^{i_1,\dots,i_n} y_{i_1}\cdots y_{i_n}$, then we may view the coefficients $c^{i_1,\dots,i_n}$ as coordinates on the vector space $S^n\mathbb{C}^N$. We will look for polynomials on our space of polynomials, that is, polynomials in the coefficients $c^{i_1,\dots,i_n}$.

Plan to show $(\mathrm{perm}_n)\notin$ **VP, or at least bound its circuit size by** r **with algebraic geometry.**

- Find a polynomial P on the space $S^n\mathbb{C}^{n^2}$ such that $P(p)=0$ for all $p\in S^n\mathbb{C}^{n^2}$ with circuit-size(p) $\le r$.
- Show that $P(\mathrm{perm}_n)\neq 0$.

By the discussion above on Zariski closure, this may be a more difficult problem: we are not just trying to exclude perm_n from having a circuit, but we are also requiring it not be "near" to having a small circuit. I return to this issue in §1.2.5 below.

1.2.3. Benchmarks and restricted models. Valiant's conjecture is expected to be extremely difficult, so it is reasonable to work towards partial results. Two types of partial results are as follows: First, one could attempt to prove the conjecture under additional hypotheses. In the complexity literature, the modified conjecture is called a *restricted model*. For an example of a restricted model, one could restrict to circuits which are *formulas* (the underlying graph is a formula, see Remark 6.1.5.2). The definition of a formula coincides with our usual notion of a formula. Numerous restricted models are discused in Chapter 7. Second, one can fix a complexity measure, e.g., circuit-size(perm_n), and prove lower bounds for it. I will refer to such progress as improving *benchmarks*.

In some cases, one can rephrase Conjecture 1.2.1.1 in a restricted model (shallow circuits) at the following cost: instead of needing to prove non-polynomial growth, one needs to prove non-nearly-exponential growth. This is also discussed in Chapter 7.

1.2.4. Another path to algebraic geometry. The permanent resembles one of the most, perhaps *the* most, studied polynomial, the *determinant* of an $n \times n$ matrix $x = (x^i_j)$:

$$\det_n(x) := \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) x^1_{\sigma(1)} \cdots x^n_{\sigma(n)}. \tag{1.2.2}$$

Here $\operatorname{sgn}(\sigma)$ denotes the sign of the permutation σ. The determinant, despite its enormous formula of $n!$ terms, can be computed very quickly, e.g., by Gaussian elimination. (See §6.1.3 for an explicit division free algorithm.) In particular $(\det_n) \in \mathbf{VP}$. It is not known if $\det_n$ is *complete* for $\mathbf{VP}$, that is, whether or not a sequence of polynomials is in $\mathbf{VP}$ if and only if it can be *reduced* to the determinant in the sense made precise below.

Although

$$\operatorname{perm}_2 \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det_2 \begin{pmatrix} a & -b \\ c & d \end{pmatrix},$$

Marcus and Minc [**MM61**], building on work of Pólya and Szegö (see [**Gat87**]), proved that one could not express $\operatorname{perm}_m(y)$ as a size m determinant of a matrix whose entries are affine linear functions of the x^i_j when $m > 2$. This raised the question that perhaps the permanent of an $m \times m$ matrix could be expressed as a slightly larger determinant, which would imply $\mathbf{VP} = \mathbf{VNP}$. More precisely, we say $p(y^1, \ldots, y^M)$ is an *affine linear projection* of $q(x^1, \ldots, x^N)$, if there exist affine linear functions $x^\alpha(y) = x^\alpha(y^1, \ldots, y^M)$ such that $p(y) = q(x(y))$. For example

$$\operatorname{perm}_3(y) = \det_7 \begin{pmatrix} 0 & 0 & 0 & 0 & y^3_3 & y^3_2 & y^3_1 \\ y^1_1 & 1 & & & & & \\ y^1_2 & & 1 & & & & \\ y^1_3 & & & 1 & & & \\ & y^2_2 & y^2_1 & 0 & 1 & & \\ & y^2_3 & 0 & y^2_1 & & 1 & \\ & 0 & y^2_3 & y^2_2 & & & 1 \end{pmatrix}. \tag{1.2.3}$$

This formula is due to B. Grenet [**Gre11**], who also generalized it to express perm_m as a determinant of size $2^m - 1$, see §6.6.3.

Valiant conjectured that one cannot do much better than this:

Definition 1.2.4.1. Let p be a polynomial. Define the *determinantal complexity* of p, denoted $\operatorname{dc}(p)$, to be the smallest n such that p is an affine linear projection of the determinant.

Valiant shows that for any polynomial P, $\mathrm{dc}(P)$ is finite but possibly larger than circuit-size(P), so the following conjecture is possibly weaker than Conjecture 1.2.1.2.

Conjecture 1.2.4.2 (Valiant [**Val79a**])**.** $\mathrm{dc}(\mathrm{perm}_m)$ *grows faster than any polynomial in* m.

The state of the art, obtained with classical differential geometry, is $\mathrm{dc}(\mathrm{perm}_m) \geq \frac{m^2}{2}$, due to Mignon and Ressayre [**MR04**]. An exposition of their result is given in §6.4.

1.2.5. Geometric Complexity Theory. The "Zariski closed" version of Conjecture 1.2.4.2 is the flagship conjecture of *Geometric Complexity Theory* (GCT) and is discussed in Chapters 6 and 8. To state it in a useful form, first rephrase Valiant's conjecture as follows:

Let $\mathrm{End}(\mathbb{C}^{n^2})$ denote the space of all linear maps $\mathbb{C}^{n^2} \to \mathbb{C}^{n^2}$, which acts on $S^n\mathbb{C}^{n^2}$ under the action $L \cdot p(x) := p(L^T(x))$, where x is viewed as a column vector of size n^2, L is an $n^2 \times n^2$ matrix, and T denotes transpose. (The transpose is used so that $L_1 \cdot (L_2 \cdot p) = (L_1 L_2) \cdot p$.) Let

$$\mathrm{End}(\mathbb{C}^{n^2}) \cdot p = \{L \cdot p \mid L \in \mathrm{End}(\mathbb{C}^{n^2})\}.$$

Define an auxiliary variable $\ell \in \mathbb{C}^1$ so $\ell^{n-m}\,\mathrm{perm}_m \in S^n\mathbb{C}^{m^2+1}$. Consider any linear inclusion $\mathbb{C}^{m^2+1} \to \mathbb{C}^{n^2}$ (e.g. with the $Mat_{m\times m}$ in the upper left hand corner and ℓ in the $(m+1)\times(m+1)$ slot and zeros elsewhere in the space of $n \times n$ matrices), so we may consider $\ell^{n-m}\,\mathrm{perm}_m \in S^n\mathbb{C}^{n^2}$. Then

$$\mathrm{dc}(\mathrm{perm}_m) \leq n \Longleftrightarrow \ell^{n-m}\,\mathrm{perm}_m \in \mathrm{End}(\mathbb{C}^{n^2}) \cdot \det{}_n\,. \tag{1.2.4}$$

This situation begins to resemble our matrix multiplication problem: we have an ambient space $S^n\mathbb{C}^{n^2}$ (resp. $(\mathbb{C}^{\mathbf{n}^2})^{\otimes 3}$ for matrix multiplication), a subset $\mathrm{End}(\mathbb{C}^{n^2}) \cdot \det_n$ (resp. $\hat{\sigma}^0_r$, the tensors of rank at most r), and a point $\ell^{n-m}\,\mathrm{perm}_m$ (resp. $M_{\langle \mathbf{n}\rangle}$) and we want to show the point is not in the subset. Note one difference here: the dimension of the ambient space is exponentially large with respect to the dimension of our subset. As before, if we want to separate the point from the subset with polynomials, we are attempting to prove a stronger statement.

Definition 1.2.5.1. For $p \in S^d\mathbb{C}^M$, let $\overline{\mathrm{dc}}(p)$ denote the smallest n such that $\ell^{n-d}p \in \overline{\mathrm{End}(\mathbb{C}^{n^2}) \cdot \det_n}$, the Zariski closure of $\mathrm{End}(\mathbb{C}^{n^2}) \cdot \det_n$. Call $\overline{\mathrm{dc}}$ the *border determinantal complexity* of p.

Conjecture 1.2.5.2. [**MS01**] $\overline{dc}(\mathrm{perm}_m)$ *grows faster than any polynomial in* m.

For this problem, we do not have an analog of Bini's theorem 1.1.14.4 that promises similar asymptotics for the two complexity measures. In this situation Mulmuley [**Mul**] conjectures that there exist sequences of polynomials (p_m) such that $\overline{\mathrm{dc}}(p_m)$ grows like a polynomial in m but $\mathrm{dc}(p_m)$ grows faster than any polynomial. Moreover he speculates that this gap explains why Valiant's conjecture is so difficult.

Representation theory indicates a path towards solving Conjecture 1.2.5.2. To explain the path, introduce the following terminology: a polynomial $p \in S^n\mathbb{C}^N$ is *characterized by its symmetries* if, letting $G_p := \{g \in GL_N \mid g \cdot p = g\}$, for any $q \in S^n\mathbb{C}^N$ with $G_q \supseteq G_p$, one has $p = \lambda q$ for some $\lambda \in \mathbb{C}$.

There are two essential observations:

- $\overline{\mathrm{End}(\mathbb{C}^{n^2}) \cdot \det_n} = \overline{GL_{n^2} \cdot \det_n}$, that is the variety $\overline{\mathrm{End}(\mathbb{C}^{n^2}) \cdot \det_n}$ is an *orbit closure*.
- $\det_n$ and perm_n are *characterized by their symmetries*.

Representation theory (more precisely, the *Peter-Weyl Theorem*, see §8.6), in principle gives a description of the polynomials vanishing on an orbit closure modulo the effect of the boundary. (More precisely, it describes the ring of regular functions on the orbit.) Unfortunately for the problem at hand, this approach, outlined in [**MS01, MS08**] was recently shown [**IP15, BIP16**] to be not viable as proposed. Nevertheless, it has pointed out several paths one could potentially use. For this reason, I explain the approach and the proof of its non-viability in Chapter 8.

*** Mention additional paths, e.g., Kayal, LR, possible comm algebra.....****

Unlike matrix multiplication, progress on Valiant's conjecture and its variants is in its infancy and I do not expect the conjecture to be fully resolved in the near future. To gain insight as to what techniques might work, it will be useful to examine "toy" versions of the problem - these questions are of mathematical significance in their own right, and lead to interesting connections between combinatorics, representation theory and geometry. Chapter 9 is dedicated to one such problem, dating back to Hermite and Hadamard, to determine the ideal of the *Chow variety* of polynomials that decompose into a product of linear forms.

1.3. How to find Hay in a haystack: the problem of explicitness

A "random" bilinear map $b : \mathbb{C}^{\mathbf{m}} \times \mathbb{C}^{\mathbf{m}} \to \mathbb{C}^{\mathbf{m}}$ will have tensor rank $\lceil \frac{\mathbf{m}^2}{2} \rceil$, see §4.8. (In particular, the standard algorithm for matrix multiplication already shows that it is pathological as a tensor as $\mathbf{n}^3 << \frac{(\mathbf{n}^2)^2}{2}$.) Now say someone hands you an explicit bilinear map, how would you determine if has tensor rank $\lceil \frac{\mathbf{m}^2}{2} \rceil$? This is the problem of *finding hay in a haystack*. Our state of the art for this question is so dismal that there is no known explicit bilinear map of tensor rank $\mathbf{3m}$, in fact the highest rank of an explicit tensor known is for matrix multiplication [**Lan14b**]: $\mathbf{R}(M_{\langle \mathbf{n}\rangle}) \geq 3\mathbf{n}^2 - o(\mathbf{n}^2)$. A second explicit sequence $T_{\mathbf{m}} : \mathbb{C}^{\mathbf{m}} \times \mathbb{C}^{\mathbf{m}} \to \mathbb{C}^{\mathbf{m}}$ with $\mathbf{R}(T_{\mathbf{m}}) \geq 3\mathbf{m} - o(\mathbf{m})$ was found in [**Zui15**]. It is a frequently stated open problem to find explicit bilinear maps $T_{\mathbf{m}} : \mathbb{C}^{\mathbf{m}} \times \mathbb{C}^{\mathbf{m}} \to \mathbb{C}^{\mathbf{m}}$ with $\mathbf{R}(T_m) \geq (3+\epsilon)\mathbf{m}$. I discuss the state of the art of this problem and the related border rank problem, where no explicit tensor $T \in \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$ with $\underline{\mathbf{R}}(T) > 2\mathbf{m}$ is known, in Chapter 5.

maybe delete below if not in book - also a lot for just one section in any case

Another famous hay in a haystack problem is *polynomial identity testing* (PIT): given a polynomial, e.g., described by a circuit (or some other recipe), determine if it is identically zero. **refs** A recent approach to this problem, via *hitting sets*, could be of interest to algebraic geometers. I discuss it in §7.12.

Yet another such problem arises in GCT: the problem of *explicit Noether normalization*. The variety $\overline{\mathrm{End}(\mathbb{C}^{n^2}) \cdot \det_n}$ has dimension (see §3.1.5 for the definition of the dimension of a variety) roughly n^4 but it lives in a space of dimension exponentially large with respect to n, namely $\binom{n^2+n-1}{n}$. If one could project this variety isomorphically into a smaller space, say of dimension polynomial in n, that did not destroy the non-inclusion of the point $\ell^{n-m}\operatorname{perm}_m$ then techniques from complexity such as hitting sets might enable a resolution of the problem. If one chooses a random such projection, it will work with probability one. See §7.12 for a discussion.

more on advanced chapters*

Chapter 2

The complexity of Matrix multiplication I: first lower bounds

In this chapter I discuss lower complexity bounds for tensors in general and matrix multiplication in particular. The two basic measures of complexity are *rank* and *border rank*. I begin, in §2.1, by defining tensors and their rank. I motivate the definition of border rank by the discovery by Bini et. al. of approximate algorithms for a reduced matrix multiplication tensor and then give its definition. Next, in §2.2 I give two derivations of Strassen's equations, the classical one due to Strassen, and a more recent one due to Ottaviani that admits generalizations. These generalizations, to *Koszul flattenings*, are described in §2.6 where they are used to show a $2\mathbf{n}^2 - \mathbf{n}$ lower bound for the border rank of $M_{\langle \mathbf{n} \rangle}$. This border rank lower bound is exploited to prove a $3\mathbf{n}^2 - o(\mathbf{n}^2)$ rank lower bound for $M_{\langle \mathbf{n} \rangle}$ in §2.7. The current state of the art is a $2\mathbf{n}^2 - \lceil \log_2(\mathbf{n}) \rceil - 1$ lower bound for the border rank of $M_{\langle \mathbf{n} \rangle}$, which is presented in §5.4.3, as it requires more geometry and representation theory than what is covered in this chapter.

2.1. Matrix multiplication and multi-linear algebra

To better understand matrix multiplication as a bilinear map, I first review basic facts from multi-linear algebra. For more details on this topic, see [**Lan12**, Chap. 2].

2.1.1. Linear algebra without coordinates. In what follows it will be essential to work without bases, so instead of writing $\mathbb{C}^{\mathbf{v}}$, I use V to denote a complex vector space of dimension $\mathbf{v}$.

The *dual space* V^* to a vector space V, is the vector space whose elements are linear maps from V to $\mathbb{C}$:

$$V^* := \{\alpha : V \to \mathbb{C} \mid \alpha \text{ is linear}\}$$

If one is working in bases and represents elements of V by column vectors, then elements of V^* are naturally represented by row vectors and the map $v \mapsto \alpha(v)$ is just row-column matrix multiplication. Given a basis $v_1, \dots, v_{\mathbf{v}}$ of V, it determines a basis $\alpha^1, \dots, \alpha^{\mathbf{v}}$ of V^* by $\alpha^i(v_j) = \delta_{ij}$, called the *dual basis*.

Exercise 2.1.1.1: (1) Assuming V is finite dimensional, write down a canonical isomorphism $V \to (V^*)^*$. ⊚

Let $V^* \otimes W$ denote the vector space of all linear maps $V \to W$. Given $\alpha \in V^*$ and $w \in W$ define a linear map $\alpha \otimes w : V \to W$ by $\alpha \otimes w(v) := \alpha(v)w$. In bases, if α is represented by a row vector and w by a column vector, $\alpha \otimes w$ will be represented by the matrix $w\alpha$. Such a linear map is said to have *rank one*. Define the *rank* of an element $f \in V^* \otimes W$ to be the smallest r such f may be expressed as a sum of r rank one linear maps.

Definition 2.1.1.2. A property of points in a variety $Z \subset V$ containing an infinite number of points is *general* or *holds generally* if the property holds on the complement of a proper subvariety of Z. In particular, a property that is general holds for a randomly chosen point in Z.

A *general point* of a variety $Z \subset V$ is a point not lying on some explicit Zariski closed subset of Z. This subset is often understood from the context and so not mentioned.

Theorem 2.1.1.3 (Fundamental theorem of linear algebra)**.** *Let V, W be finite dimensional vector spaces, let $f : V \to W$ be a linear map, and let A_f be a matrix representing f. Then*

(1)

$$\begin{aligned} \operatorname{rank}(f) &= \dim f(V) \\ &= \dim(\operatorname{span}\{\text{columns of } A_f\}) \\ &= \dim(\operatorname{span}\{\text{rows of } A_f\}) \\ &= \dim V - \dim \ker f. \end{aligned}$$

In particular $\operatorname{rank}(f) \leq \min\{\dim V, \dim W\}$.

(2) *For general* $f \in V^* \otimes W$, $\operatorname{rank}(f) = \min\{\dim V, \dim W\}$.

(3) *If a sequence of linear maps f_t of rank r has a limit f_0, then* $\operatorname{rank}(f_0) \leq r$.

(4) $\operatorname{rank}(f) \leq r$ *if and only if, in any choice of bases, the determinants of all size $r+1$ submatrices of the matrix representing f are zero.*

Note that assertion 4) shows that the set linear maps of rank at most r forms an algebraic variety. Although we take it for granted, it is really miraculous that the fundamental theorem of linear algebra is true. I explain why in §3.1.3.

Exercise 2.1.1.4: (1!) Prove the theorem. ⊚

Many standard notions from linear algebra have coordinate free definitions. For example: A linear map $f : V \to W$ determines a linear map $f^T : W^* \to V^*$ defined by $f^T(\beta)(v) := \beta(f(v))$ for all $v \in V$ and $\beta \in W^*$. Note that this is consistent with the notation $V^* \otimes W \simeq W \otimes V^*$, being interpreted as the space of all linear maps $(W^*)^* \to V^*$, that is, the order we write the factors does not matter. If we work in bases and insist that all vectors are column vectors, the matrix of f^T is just the transpose of the matrix of f.

Exercise 2.1.1.5: (1) Show that we may also consider an element $f \in V^* \otimes W$ as a bilinear map $b_f : V \times W^* \to \mathbb{C}$ defined by $b_f(\beta, v) := \beta(f(v))$.

In the vector space $V^* \otimes V$ there is a unique line such that every vector on the line has the same matrix representative for any choice of basis (and corresponding choice of dual basis). This line is of course $\mathbb{C}\{\mathrm{Id}_V\}$, the scalar multiples of the identity map. Letting $GL(V)$ denote the group of changes of basis in V, we say $\mathbb{C}\{\mathrm{Id}_V\}$ is the unique line in $V^* \otimes V$ *invariant* under the action of $GL(V)$.

Exercise 2.1.1.6: (1) If $v_1, \ldots, v_{\mathbf{v}}$ is a basis of V and $\alpha^1, \ldots, \alpha^{\mathbf{v}}$ is the dual basis of V^*, show that the identity map on V is $\mathrm{Id}_V = \sum_j \alpha^j \otimes v_j$.

Exercise 2.1.1.7: (1) Show that there is a canonical isomorphism $(V^* \otimes W)^* \to V \otimes W^*$ where $\alpha \otimes w(v \otimes \beta) := \alpha(w)\beta(v)$. Now let $V = W$ and let $\mathrm{Id}_V \in V^* \otimes V \simeq (V^* \otimes V)^*$ denote the identity map. What is $\mathrm{Id}_V(f)$ for $f \in V^* \otimes V$? ⊚

2.1.2. Multi-linear maps and tensors. We say $V \otimes W$ defined in §2.1.1 is the *tensor product* of V with W. More generally, for vector spaces $A_1, \ldots, A_n$ define their tensor product $A_1 \otimes \cdots \otimes A_n$ to be the space of n-linear maps $A_1^* \times \cdots \times A_n^* \to \mathbb{C}$, equivalently the space of $(n-1)$-linear maps $A_1^* \times \cdots \times A_{n-1}^* \to A_n$ etc.. When $A_1 = \cdots = A_n = V$, write $V^{\otimes n} = V \otimes \cdots \otimes V$.

Let $a_j \in A_j$ and define an element $a_1 \otimes \cdots \otimes a_n \in A_1 \otimes \cdots \otimes A_n$ to be the n-linear map

$$a_1 \otimes \cdots \otimes a_n(\alpha^1, \ldots, \alpha^n) := \alpha^1(a_1) \cdots \alpha^n(a_n).$$

Exercise 2.1.2.1: Show that if $\{a_j^{s_j}\}$, $1 \leq s_j \leq \mathbf{a}_j$, is a basis of A_j, then $a_1^{s_1} \otimes \cdots \otimes a_n^{s_n}$ is a basis of $A_1 \otimes \cdots \otimes A_n$. In particular $\dim(A_1 \otimes \cdots \otimes A_n) = \mathbf{a}_1 \cdots \mathbf{a}_n$. ⊚

Remark 2.1.2.2. One may identify $A_1 \otimes \cdots \otimes A_n$ with any re-ordering of the factors. When I need to be explicit about this, I will call this identification the *re-ordering isomorphism.*

Example 2.1.2.3 (Matrix multiplication)**.** Let x_α^i, y_u^α, z_i^u respectively be bases of $A = \mathbb{C}^{\mathbf{nm}}$, $B = \mathbb{C}^{\mathbf{ml}}$, $C = \mathbb{C}^{\mathbf{ln}}$, then the standard expression of matrix multiplication as a tensor is

$$M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle} = \sum_{i=1}^{\mathbf{n}} \sum_{\alpha=1}^{\mathbf{m}} \sum_{u=1}^{\mathbf{l}} x_\alpha^i \otimes y_u^\alpha \otimes z_i^u \tag{2.1.1}$$

Exercise 2.1.2.4: (2) Write Strassen's algorithm out as a tensor. ⊚

2.1.3. Tensor rank. An element $T \in A_1 \otimes \cdots \otimes A_n$ is said to have *rank one* if there exist $a_j \in A_j$ such that $T = a_1 \otimes \cdots \otimes a_n$.

We will use the following measure of complexity:

Definition 2.1.3.1. Let $T \in A_1 \otimes \cdots \otimes A_n$. Define the *rank* (or *tensor rank*) of T to be the smallest r such that T may be written as the sum of r rank one tensors. We write $\mathbf{R}(T) = r$. Let $\hat{\sigma}_r^0 \subset A_1 \otimes \cdots \otimes A_n$ denote the set of tensors of rank at most r.

The rank of $T \in A \otimes B \otimes C$ is comparable to all other standard measures of complexity on the space of bilinear maps, see, e.g., [**BCS97**, §14.1].

By (2.1.1) we conclude $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}) \leq \mathbf{nml}$. Strassen's algorithm shows $\mathbf{R}(M_{\langle 2,2,2\rangle}) \leq 7$. Shortly afterwards, Winograd [**Win71**] showed $\mathbf{R}(M_{\langle 2,2,2\rangle}) = 7$.

Recall the notation $M_{\langle \mathbf{n}\rangle} = M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$.

2.1.4. Another spectacular failure. After Strassen's failure to prove the standard algorithm for matrix multiplication was optimal, Bini et. al. [**BLR80**] considered the *reduced matrix multiplication operator*

$$\begin{aligned} M_{\langle 2\rangle}^{red} :=& x_1^1 \otimes (y_1^1 \otimes z_1^1 + y_2^1 \otimes z_1^2) + x_2^1 \otimes (y_1^2 \otimes z_1^1 + y_2^2 \otimes z_1^2) + x_1^2 \otimes (y_1^1 \otimes z_2^1 + y_2^1 \otimes z_2^2) \\ &\in \mathbb{C}^3 \otimes \mathbb{C}^4 \otimes \mathbb{C}^4. \end{aligned}$$

obtained by setting the x_2^2 entry for $M_{\langle 2\rangle}$ to zero. The standard presentation shows $\mathbf{R}(M_{\langle 2\rangle}^{red}) \leq 6$. They attempted to find a rank five expression for $M_{\langle 2\rangle}^{red}$.

They searched for such an expression by computer. Their method was to minimize the norm of $M^{red}_{\langle 2\rangle}$ minus a rank five tensor that varied, and their computer kept on producing rank five tensors with the norm of the difference getting smaller and smaller, but with larger and larger coefficients. Bini (personal communication) told me about how he lost sleep trying to understand what was wrong with his computer code. This went on for some time, when finally he realized *there was nothing wrong with the code*: that the output it produced was a manifestation of the phenomenon Bini named *border rank* [**Bin80**], which was mentioned in the introduction in the context of finding polynomials for upper rank bounds.

The expression for the tensor $M^{red}_{\langle 2\rangle}$ that their computer search found was essentially

$$
\begin{aligned}
M^{red}_{\langle 2\rangle} = \lim_{t\to 0}\frac{1}{t}[&(x^1_2+tx^1_1)\otimes(y^1_2+ty^2_2)\otimes z^2_1\\
&+(x^2_1+tx^1_1)\otimes y^1_1\otimes(z^1_1+tz^1_2)\\
&-x^1_2\otimes y^1_2\otimes((z^1_1+z^2_1)+tz^2_2)\\
&-x^2_1\otimes((y^1_1+y^1_2)+ty^2_1)\otimes z^1_1\\
&+(x^1_2+x^2_1)\otimes(y^1_2+ty^2_1)\otimes(z^1_1+tz^2_2)].
\end{aligned}
\tag{2.1.2}
$$

In what follows I first explain why border rank is needed in the study of tensors and then properly define it.

2.1.5. The Fundamental theorem of linear algebra is false for tensors.

Recall the fundamental theorem of linear algebra from §2.1.1.3.

Theorem 2.1.5.1. *If $T\in\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$ is outside the zero set of a certain finite collection of polynomials (in particular outside a certain set of measure zero), then $\mathbf{R}(T)\geq\lceil\frac{\mathbf{m}^3-1}{3\mathbf{m}-2}\rceil$.*

Tensor rank can jump up (or down) under limits.

Remark 2.1.5.2. Strassen and Lickteig showed that equality holds in Theorem **??** for all $\mathbf{m}\neq 3$ (and when $\mathbf{m}=3$, for most T one has $\mathbf{R}(T)=5$.

An analogous statement holds in any tensor space $A_1\otimes\cdots\otimes A_n$ with $n\geq 3$.

The first assertion is proved in §**??**. To see the second assertion, at least when $r=2$, consider

$$T(t):=\frac{1}{t}[a_1\otimes b_1\otimes c_1-(a_1+ta_2)\otimes(b_1+tb_2)\otimes(c_1+tc_2)]$$

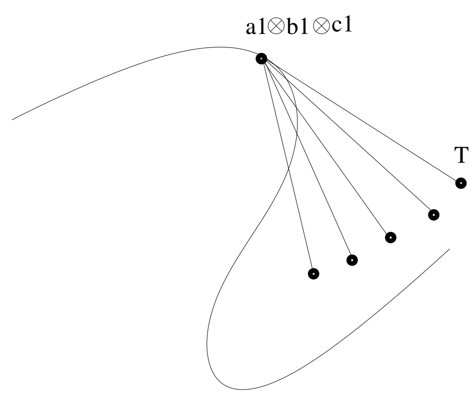
and note that

$$\lim_{t\to 0}T(t)=a_1\otimes b_1\otimes c_2+a_1\otimes b_2\otimes c_1+a_2\otimes b_1\otimes c_1$$

which does not have rank two (exercise).

Remark 2.1.5.3. Physicists like to call the tensor $a_1 \otimes b_1 \otimes c_2 + a_1 \otimes b_2 \otimes c_1 + a_2 \otimes b_1 \otimes c_1$ the *W-state* so I will sometimes denote it T_{WState}

To visualize why rank can jump up while taking limits, consider the following picture, where the curve represents the points of $\hat{\sigma}_1^0$. Points of $\hat{\sigma}_2^0$ (e.g., the dots limiting to the dot labelled T) are those on a secant line to $\hat{\sigma}_1^0$, and the points where the rank jumps up, such at the dot labelled T, are those that lie on a tangent line to $\hat{\sigma}_1^0$. This phenomena fails to occur for matrices because for matrices, every point on a tangent line is also on an honest secant line. Thus in some sense it is a miracle that rank is semi-continuous for matrices.



Our situation regarding tensor rank may be summarized as follows:

- The set $\hat{\sigma}_r^0$ is not closed under taking limits. I will say a set that is closed under taking limits is *Euclidean closed*.
- It is also not *Zariski closed*, i.e., the zero set of all polynomials vanishing on $\hat{\sigma}_r^0$ includes tensors that are of rank greater than r.

The tensors that are honestly "close" to tensors of rank r would be the Euclidean closure, but to deal with polynomials as proposed in §1.1.12-1.1.14, we need to work with the Zariski closure.

Often the Zariski closure is much larger than the Euclidean closure. For example, the Zariski closure of $\mathbb{Z} \subset \mathbb{C}$ is $\mathbb{C}$, while $\mathbb{Z}$ is already closed in the Euclidean topology.

However, for the purposes of proving lower bounds, none of this is an issue, but when we discuss upper bounds, we will need to deal with these problems. For now, I mention that with $\hat{\sigma}_r^0$ we have good luck: the Zariski and Euclidean closures of $\hat{\sigma}_r^0$ coincide, so our apparently different informal uses of the term border rank coincide. I present the proof in §3.1.6.

Exercise 2.1.5.4: (2) Show that the Euclidean closure (i.e., closure under taking limits) of a set is always contained in its Zariski closure. ⊚

2.1.6. Border rank. Generalizing the discussion in §1.1.11, $\hat{\sigma}_r = \hat{\sigma}_{r,A_1\otimes\cdots\otimes A_n}$ denotes the Zariski (and by the above discussion Euclidean) closure of $\hat{\sigma}_r^0$, and the *border rank* of $T \in A_1\otimes\cdots\otimes A_n$, denoted $\underline{\mathbf{R}}(T)$, is the smallest r such that $T \in \hat{\sigma}_r$. By the above discussion, border rank is semi-continuous.

Exercise 2.1.6.1: (1) Write down an explicit tensor of border rank r in $\mathbb{C}^r\otimes\mathbb{C}^r\otimes\mathbb{C}^r$ with rank greater than r. ⊚

Border rank is easier to work with than rank for several reasons. For example, the maximal rank of a tensor in $\mathbb{C}^m\otimes\mathbb{C}^m\otimes\mathbb{C}^m$ is not known in general. In contrast, the maximal border rank is known to be $\lceil \frac{m^3-1}{3m-2} \rceil$ for all $m \neq 3$, and is 5 when $m = 3$ [**Lic85**]. In particular Strassen's algorithm could have been predicted in advance with this knowledge. The method of proof is a differential-geometric calculation that dates back to Terracini in the 1900's [**Ter11**], see §**??** for a discussion.

Exercise 2.1.6.2: (1) Prove that if $T \in A\otimes B\otimes C$ and $T' := T|_{A'\times B'\times C'}$ for some $A' \subseteq A^*$, $B' \subseteq B^*$, $C' \subseteq C^*$, then $\mathbf{R}(T) \geq \mathbf{R}(T')$ and $\underline{\mathbf{R}}(T) \geq \underline{\mathbf{R}}(T')$. ⊚

Exercise 2.1.6.3: (1) Let $T_j \in A_j\otimes B_j\otimes C_j$, $1 \leq j,k,l \leq s$. Consider $T_1 \oplus \cdots \oplus T_s \in (\oplus_j A_j)\otimes(\oplus_k B_k)\otimes(\oplus_l C_l)$ Show that $\mathbf{R}(\oplus_j T_j) \leq \sum_{i=1}^s \mathbf{R}(T_i)$ and that the statement also holds for border rank.

Exercise 2.1.6.4: (1) Let $T_j \in A_j\otimes B_j\otimes C_j$, $1 \leq j,k,l \leq s$. Let $A = \otimes_j A_j$, $B = \otimes_k B_k$, and $C = \otimes_l C_l$, consider $T_1\otimes\cdots\otimes T_s \in A\otimes B\otimes C$. Show that $\mathbf{R}(\otimes_{i=1}^s T_i) \leq \Pi_{i=1}^s \mathbf{R}(T_i)$, and that the statement also holds for border rank.

2.1.7. Our first lower bound. Given $T \in A\otimes B\otimes C$, write $T \in A\otimes(B\otimes C)$ and think of T as a linear map $T_A : A^* \to B\otimes C$.

Proposition 2.1.7.1. $\underline{\mathbf{R}}(T) \geq \mathrm{rank}(T_A)$.

Exercise 2.1.7.2: (1!) Prove Proposition 2.1.7.1. ⊚

Permuting the three factors, we have equations for $\hat{\sigma}_{r,A\otimes B\otimes C}$ for $r \leq \max\{\mathbf{a}-1, \mathbf{b}-1, \mathbf{c}-1\}$, namely the size $r+1$ minors of the linear maps T_A, T_B, T_C.

Definition 2.1.7.3. A tensor $T \in A\otimes B\otimes C$ is *concise* if the maps T_A, T_B and T_C are all injective.

Exercise 2.1.7.4: (1) Find a choice of bases such that

$$M_{\langle \mathbf{n}\rangle A}(A^*) = \begin{pmatrix} x & & \\ & \ddots & \\ & & x \end{pmatrix}$$

where $x = (x^i_j)$ is $\mathbf{n} \times \mathbf{n}$, i.e., the image in the space of $\mathbf{n}^2 \times \mathbf{n}^2$ matrices is block diagonal with all blocks the same.

Exercise 2.1.7.5: (1) Show that $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle}) \geq \mathbf{n}^2$.

Exercise 2.1.7.6: (1) Show $\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},1\rangle}) = \mathbf{mn}$ and $\underline{\mathbf{R}}(M_{\langle \mathbf{m},1,1\rangle}) = \mathbf{m}$.

Exercise 2.1.7.7: (1!) Let $\mathbf{b} = \mathbf{c}$ and assume T_A is injective. Show that if $T(A^*)$ is simultaneously diagonalizable under the action of $GL(B) \times GL(C)$ (i.e., if we take a basis $\alpha^1, \ldots, \alpha^{\mathbf{a}}$ of A^*, there exists $g \in GL(B) \times GL(C)$ such that the elements $g \cdot T(\alpha^1), \ldots, g \cdot T(\alpha^{\mathbf{a}})$ are all diagonal) then $\mathbf{R}(T) \leq \mathbf{b}$, and therefore if $T(A^*)$ is the limit of simultaneously diagonalizable subspaces then $\underline{\mathbf{R}}(T) \leq \mathbf{b}$.

2.2. Strassen's equations

An extensive discussion of Strassen's equations and generalizations appears in [**Lan12**, §7.6].

2.2.1. Beyond the classical equations.

The classical equations just used that $B \otimes C$ is a vector space. To extract more information from T_A, we examine its image in $B \otimes C$, which we will view as a space of linear maps $C^* \to B$. If T is concise and has minimal border rank $\max\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$, the image should be special in some way - how? Assume $\mathbf{b} = \mathbf{c}$ so the image is a space of linear maps $\mathbb{C}^{\mathbf{b}} \to \mathbb{C}^{\mathbf{b}}$ (more precisely a space of linear maps $C^* \to B$). (If $\mathbf{b} < \mathbf{c}$, just restrict to some $\mathbb{C}^{\mathbf{b}} \subset C$.) If $\mathbf{R}(T) = \mathbf{b}$, then $T_A(A^*)$, which I write as $T(A^*)$, will be spanned by $\mathbf{b}$ rank one linear maps.

Lemma 2.2.1.1. *If $\mathbf{a} = \mathbf{b} = \mathbf{c}$ and T_A is injective, then $\mathbf{R}(T) = \mathbf{a}$ if and only if $T(A^*)$ is spanned by $\mathbf{a}$ rank one linear maps.*

Exercise 2.2.1.2: (2!) Prove Lemma 2.2.1.1. ⊚

How can we test if the image is spanned by $\mathbf{b}$ rank one linear maps? If $T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_{\mathbf{a}} \otimes b_{\mathbf{a}} \otimes c_{\mathbf{a}}$ with each set of vectors a basis, then

$$T(A^*) = \left\{ \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_{\mathbf{a}} \end{pmatrix} \mid x_j \in \mathbb{C} \right\},$$

and this is the case for a general rank $\mathbf{a}$ tensor in $\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}$. That is, $T(A^*) \subset B\otimes C$, when T has border rank $\mathbf{a}$ lies in the Zariski closure of the subspaces that, under the action of $GL(B)\times GL(C)$ are simultaneously *diagonalizable* in the sense of Exercise 2.1.7.7. From this perspective our problem becomes: determine polynomials on $A\otimes B\otimes C$ that vanish of the set of T such that $T(A)$ is diagonalizable. (For those familiar with Grassmannians, it is better to say we should look for polynomials on the Grassmannian $G(\mathbf{a}, B\otimes C)$ vanishing on the simultaneously diagonlizable subspaces.)

A set of equations whose zero set is exactly the Zariski closure of the diagonalizable matrices is not known! What follows are *some* equations. (More are given in Chapter 5.) Recall that $B\otimes C = \mathrm{Hom}(C^*, B)$, the space of linear maps from C^* to B. If instead we had $\mathrm{Hom}(B, B) = \mathrm{End}(B)$, the space of linear maps from B to itself, a necessary condition for endomorphisms to be simultaneously diagonalizable is that they must commute, and the algebraic test for a subspace $U \subset \mathrm{End}(B)$ to be abelian is simple: the commutators $[X_i, X_j] := X_iX_j - X_jX_i$ must vanish on a basis $X_1, \dots, X_{\mathbf{u}}$ of U. (Note that commutators only make sense for maps from a vector space to itself.) These degree two equations exactly characterize abelian subspaces. We do not have maps from a vector space to itself, but we can fix the situation if there exists $\alpha \in A^*$ such that $T_A(\alpha) : C^* \to B$ is invertible, as then we could test if the commutators $[T_A(\alpha_1)T_A(\alpha)^{-1}, T_A(\alpha_2)T_A(\alpha)^{-1}]$ are zero. So we now have a test, but it is not expressed in terms of polynomials on $A\otimes B\otimes C$, and we cannot apply it to all tensors. These problems are fixed in §2.5. For now I record what we have so far:

Proposition 2.2.1.3. *Let* $\mathbf{b} = \mathbf{c}$ *and let* $T \in A\otimes B\otimes C$ *be such that there exists* $\alpha \in A^*$ *with* $\mathrm{rank}(T(\alpha)) = \mathbf{b}$, *so* $\underline{\mathbf{R}}(T) \geq \mathbf{b}$. *Use* $T(\alpha)$ *to identify* $B\otimes C$ *with* $\mathrm{End}(B)$. *If* $\underline{\mathbf{R}}(T) = \mathbf{b}$, *then for all* $X_1, X_2 \in T(A^*)T(\alpha)^{-1} \subset \mathrm{End}(B)$, $[X_1, X_2] = 0$.

2.2.2. Strassen's equations: original formulation.

If $T \in A\otimes B\otimes C$ is "close to" having rank $\mathbf{a} = \mathbf{b} = \mathbf{c}$, one expects, using α with $T(\alpha)$ invertible, that $T(A^*)T(\alpha)^{-1} \subset \mathrm{End}(B)$ will be "close to" being abelian. The following theorem makes this precise:

Theorem 2.2.2.1 (Strassen)**.** [**Str83**] *Let* $T \in A\otimes B\otimes C$ *and assume* $\mathbf{b} = \mathbf{c}$. *Assume that there exists* $\alpha \in A^*$ *such that* $\mathrm{rank}(T(\alpha)) = \mathbf{b}$. *Use* $T(\alpha)$ *to identify* $B\otimes C$ *with* $\mathrm{End}(B)$. *Then for all* $X_1, X_2 \in T(A^*)T(\alpha)^{-1} \subset \mathrm{End}(B)$,

$$\underline{\mathbf{R}}(T) \geq \frac{1}{2}\mathrm{rank}([X_1, X_2]) + \mathbf{b}.$$

I prove Theorem 2.2.2.1 for the case of the determinant of $[X_1, X_2]$ in §2.5 below and in general in §5.2.2.

We now have potential tests for border rank for tensors in $\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$ up to $r=\frac{3}{2}\mathbf{m}$, in fact tests for border rank for tensors in $\mathbb{C}^3\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$ up to $r=\frac{3}{2}\mathbf{m}$, since our test only used three vectors from A^*. (I write "potential tests" rather than "polynomial tests" because to write down the commutator we must be able to find an invertible element in $T(A^*)$.)

Strassen uses Theorem 2.2.2.1 to show that $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle})\geq \frac{3}{2}\mathbf{n}^2$:

Exercise 2.2.2.2: (2!) Prove $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle})\geq \frac{3}{2}\mathbf{n}^2$. ⊚

Exercise 2.2.2.3: (2) Show that $\underline{\mathbf{R}}(M^{red}_{\langle 2\rangle})=5$ and for $\mathbf{m}>2$ that $\underline{\mathbf{R}}(M^{red}_{\langle \mathbf{m},2,2\rangle}\geq 3\mathbf{m}-1$, where $M^{red}_{\langle \mathbf{m},2,2\rangle}$ is $M_{\langle \mathbf{m},2,2\rangle}$ with x^1_1 set to zero.

A natural question arises: exchanging the roles of A,B,C we obtain three sets of such equations - are the three sets of equations the same or different? We should have already asked this question for the three types of usual flattenings: are the equations coming from the minors of T_A,T_B,T_C the same or different? It is easy to write down tensors where $\operatorname{rank}(T_A),\operatorname{rank}(T_B),\operatorname{rank}(T_C)$ are distinct, however for 2×2 minors, two sets of them vanishing implies the third does as well, see, §8.3.1, where the question regarding Strassen's equations is answered as well with the help of representation theory.

One can generalize Strassen's equations by taking higher order commutators, see [**LM08**]. These generalizations do give new equations, but they do not give equations for border rank beyond the $\frac{3}{2}\mathbf{b}$ of Strassen's equations.

2.2.3. Coming attractions: border rank bounds beyond Strassen's equations.

The following more complicated expression gives equations for $\hat{\sigma}_r$ for $r>\frac{3}{2}\mathbf{b}$:

Let $T\in\mathbb{C}^5\otimes\mathbb{C}^{\mathbf{b}}\otimes\mathbb{C}^{\mathbf{b}}$, write $T=a_0\otimes X_0+\cdots a_4\otimes X_4$ with $X_j\in B\otimes C$. Assume that $\operatorname{rank}(X_0)=\mathbf{b}$ and choose bases such that $X_0=\mathrm{Id}$. Consider the following $5\mathbf{b}\times 5\mathbf{b}$ matrix:

$$T_A^{\wedge 2}=\begin{pmatrix} 0 & [X_1,X_2] & [X_1,X_3] & [X_1,X_4]\\ [X_2,X_1] & 0 & [X_2,X_3] & [X_2,X_4]\\ [X_3,X_1] & [X_3,X_2] & 0 & [X_3,X_4]\\ [X_4,X_1] & [X_4,X_2] & [X_4,X_3] & 0\end{pmatrix}. \tag{2.2.1}$$

The name $T_A^{\wedge 2}$ is explained in §2.6.1 where the proof of the following proposition also appears.

Proposition 2.2.3.1. [**LO15**] *Let $T\in\mathbb{C}^5\otimes\mathbb{C}^{\mathbf{b}}\otimes\mathbb{C}^{\mathbf{b}}$ be as above. Then $\underline{\mathbf{R}}(T)\geq\frac{\operatorname{rank}T_A^{\wedge 2}}{3}$. If $T\in A\otimes\mathbb{C}^{\mathbf{b}}\otimes\mathbb{C}^{\mathbf{b}}$ with $\mathbf{a}>5$, one obtains the same result for all restrictions of T to $\mathbb{C}^5\otimes\mathbb{C}^{\mathbf{b}}\otimes\mathbb{C}^{\mathbf{b}}$ for any $\mathbb{C}^5\subset A^*$.*

Exercise 2.2.3.2: (2) Show that for $\mathbf{n} \geq 5$, $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \geq \frac{5}{3}\mathbf{n}^2$.

The matrices $[X_1, X_2]$ and $T_A^{\wedge 2}$ are part of a sequence of constructions giving better lower bounds for border rank for tensors. The limits of this method are lower bounds of $2\mathbf{b} - 3$. How can one find such sequences and prove that they give lower bounds for border rank? To do this we will need more language from multi-linear algebra. Our first task will be to generalize the space of skew-symmetric matrices. It will be convienient to generalize symmetric matrices at the same time.

2.3. Symmetric and skew-symmetric tensors

Exercise 2.3.0.1: (1) Let X be a matrix representing a bilinear form on $\mathbb{C}^{\mathbf{m}}$, by $X(v, w) = v^T X w$. Show that if X is a symmetric matrix, then $X(v, w) = X(w, v)$ and if X is a skew-symmetric matrix, then $X(v, w) = -X(w, v)$.

Definition 2.3.0.2. A tensor $T \in V^{\otimes d}$ is said to be *symmetric* if $T(\alpha_1, \dots, \alpha_d) = T(\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(d)})$ for all $\alpha_1, \dots, \alpha_d \in V^*$ and all permutations $\sigma \in \mathfrak{S}_d$, and *skew-symmetric* if $T(\alpha_1, \dots, \alpha_d) = \mathrm{sgn}(\sigma) T(\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(d)})$ for all $\alpha_1, \dots, \alpha_d \in V^*$ and all $\sigma \in \mathfrak{S}_d$. Let $S^d V \subset V^{\otimes d}$ (resp. $\Lambda^d V \subset V^{\otimes d}$) denote the space of symmetric (resp. skew-symmetric) tensors.

The spaces $\Lambda^d V$ and $S^d V$ are independent of a choice of basis in V. In particular, the splitting

$$V^{\otimes 2} = S^2 V \oplus \Lambda^2 V \tag{2.3.1}$$

of the space of matrices into the direct sum of symmetric and skew symmetric matrices is invariant under the *action* of $GL(V)$ given by: for $g \in GL(V)$ and $v \otimes w \in V \otimes V$, $v \otimes w \mapsto gv \otimes gw$.

Introduce the notations:

$$x_1 x_2 \cdots x_k := \frac{1}{k!} \sum_{\sigma \in \mathfrak{S}_k} x_{\sigma(1)} \otimes x_{\sigma(2)} \otimes \cdots \otimes x_{\sigma(k)} \in S^k V,$$

and

$$x_1 \wedge x_2 \wedge \cdots \wedge x_k := \frac{1}{k!} \sum_{\sigma \in \mathfrak{S}_k} \mathrm{sgn}(\sigma) x_{\sigma(1)} \otimes x_{\sigma(2)} \otimes \cdots \otimes x_{\sigma(k)} \in \Lambda^k V,$$

respectively called the *symmetric product* (or simply *product*) of $x_1, \dots, x_k$ and the *wedge product* of $x_1, \dots, x_k$.

If $v_1, \dots, v_{\mathbf{v}}$ is a basis of V, then $v_{i_1} \otimes \cdots \otimes v_{i_d}$ with $i_j \in [\mathbf{v}] := \{1, \dots, \mathbf{v}\}$ is a basis of $V^{\otimes d}$, $v_{i_1} \cdots v_{i_d}$ with $1 \leq i_1 \leq \cdots \leq i_d \leq \mathbf{v}$ is a basis of $S^d V$ and $v_{i_1} \wedge \cdots \wedge v_{i_d}$ with $1 \leq i_1 < \cdots < i_d \leq \mathbf{v}$ is a basis of $\Lambda^d V$. Call these bases *induced bases*. If $x_j = (x_j^1, \dots, x_j^{\mathbf{v}})^T$ in the basis $v_1, \dots, v_{\mathbf{v}}$, then the expression of $x_1 \wedge \cdots \wedge x_k$ in the induced basis is such that the coefficient of $v_{i_1} \wedge \cdots \wedge v_{i_k}$ is

$$\det \begin{pmatrix} x_1^{i_1} & \cdots & x_1^{i_k} \\ & \vdots & \\ x_k^{i_1} & \cdots & x_k^{i_k} \end{pmatrix}.$$

For example, if $V = \mathbb{C}^4$ with basis $e_1, \ldots, e_4$, then $\Lambda^2 V$ inherits a basis $e_1 \wedge e_2, \ldots, e_3 \wedge e_4$. If

$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix}, \ w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix}, \text{ then } v \wedge w = \begin{pmatrix} v_1 w_2 - v_2 w_1 \\ v_1 w_3 - v_3 w_1 \\ v_1 w_4 - v_4 w_1 \\ v_2 w_3 - v_3 w_2 \\ v_2 w_4 - v_4 w_2 \\ v_3 w_4 - v_4 w_3 \end{pmatrix}.$$

Exercise 2.3.0.3: (1) Show that there is a natural map $\Lambda^k V \otimes V \to \Lambda^{k+1} V$ that commutes with the action of $GL(V)$, and more generally there is a natural map $\Lambda^k V \otimes \Lambda^t V \to \Lambda^{k+t} V$.

Exercise 2.3.0.4: (1) Let $k \geq t$ and show that there is a natural map $S^k V^* \otimes S^t V \to S^{k-t} V^*$ that commutes with the action of $GL(V)$. This map has the following interpretation: $S^k V^*$ may be thought of as the space of homogeneous polynomials of degree k on V (to a symmetric tensor T associate the polynomial P_T where $P_T(v) := T(v, \ldots, v)$), and $S^t V$ the homogeneous linear differential operators of order t on the space of polynomials. The map is then $P \otimes D \mapsto D(P)$ where P is a polynomial and D is a differential operator. Sometimes $D(P)$ is denoted $D \lrcorner P$.

Exercise 2.3.0.5: (1) Show that for $k < l$ there is a natural map, $\Lambda^k V^* \otimes \Lambda^l V \to \Lambda^{l-k} V$ that commutes with the action of $GL(V)$. This map is often denoted $\beta \otimes Y \mapsto \beta \lrcorner Y$

Exercise 2.3.0.6: (1) Let $Sym(V) = \oplus_{j=0}^{\infty} S^j V$, $\Lambda^\bullet V = \oplus_{j=0}^{\mathbf{v}} \Lambda^j V$ and $V^{\otimes \bullet} = \oplus_{j=0}^{\infty} V^{\otimes j}$. Show that these spaces are all naturally algebras with the above defined products, respectively called the *symmetric, exterior* and *tensor* algebras. .

2.4. Schur's lemma

I take a short detour into elementary representation theory to prove a lemma everyone should know.

Definition 2.4.0.1. Let W_1, W_2 be vector spaces, let G be a group, and let $\rho_j : G \to GL(W_j)$, $j = 1, 2$ be representations. A *G-module homomorphism*, or *G-module map*, is a linear map $f : W_1 \to W_2$ such that $f(\rho_1(g) \cdot v) =$

$\rho_2(g) \cdot f(v)$ for all $v \in W_1$ and $g \in G$. One also says that f is G-*equivariant*.

One says W_1 and W_2 are *isomorphic G-modules* if there exists a G-module homomorphism $W_1 \to W_2$ that is a linear isomorphism.

For a group G and G-modules V and W, let $\mathrm{Hom}_G(V, W) \subset V^* {\otimes} W$ denote the vector space of G-module homomorphisms $V \to W$.

Exercise 2.4.0.2: (1!!) Show that the image and kernel of a G-module homomorphism are G-modules.

The following easy lemma is central to representation theory:

Lemma 2.4.0.3 (Schur's Lemma)**.** *Let G be a group, let V and W be irreducible G-modules and let $f : V \to W$ be a G-module homomorphism. Then either $f = 0$ or f is an isomorphism. If further $V = W$, then $f = \lambda \,\mathrm{Id}_V$ for some constant λ.*

Exercise 2.4.0.4: (1!!) Prove Schur's Lemma.

We will see numerous examples illustrating the utility of Schur's Lemma. I cannot over-emphasize the importance of this simple Lemma. I use it every day of my mathematical life.

For any group G, G-module M, and irreducible G-module V, the *isotypic component* of V in M is the largest subspace of M isomorphic to $V^{\oplus m_V}$ for some m_V. The integer m_V is called the *multiplicity* of V in M.

2.5. Reformulation and proof of Strassen's equations

We augment the linear map $T_B : B^* \to A{\otimes}C$ by tensoring it with Id_A, to get a linear map

$$\mathrm{Id}_A \otimes T_B : A{\otimes}B^* \to A{\otimes}A{\otimes}C.$$

So far we have done nothing interesting, but by (2.3.1) the target of this map decomposes under the action of $GL(A) \times GL(C)$ as $(\Lambda^2 A{\otimes}C) \oplus (S^2 A{\otimes}C)$, and we may project onto these factors. Write the projections as:

$$T_{BA}^{\wedge} = T_A^{\wedge} : A{\otimes}B^* \to \Lambda^2 A{\otimes}C \text{ and } T_{BA}^{\circ} : A{\otimes}B^* \to S^2 A{\otimes}C. \tag{2.5.1}$$

Exercise 2.5.0.1: (1) Show that if $T = a{\otimes}b{\otimes}c$ is a rank one tensor, then $\mathrm{rank}(T_A^{\wedge}) = \mathbf{a} - 1$ and $\mathrm{rank}(T_{BA}^{\circ}) = \mathbf{a}$.

Exercise 2.5.0.1 implies:

Proposition 2.5.0.2. *If $\underline{\mathbf{R}}(T) \le r$, than $\mathrm{rank}(T_A^{\wedge}) \le r(\mathbf{a}{-}1)$ and $\mathrm{rank}(T_{BA}^{\circ}) \le r\mathbf{a}$.*

The second map will not do any better than the classical equations, but the first, e.g., when $\mathbf{a} = 3$, is a map from a $2\mathbf{b}$ dimensional vector space to a $2\mathbf{c}$ dimensional vector space, so if $\mathbf{b} \leq \mathbf{c}$ we can get border rank bounds up to $\frac{3}{2}\mathbf{b}$.

The first set is Strassen's equations, as I now show. If $\mathbf{a} > 3$, one can choose a three dimensional subspace $A' \subset A^*$ and consider T restricted to $A' \times B^* \times C^*$ to obtain equations. (This is what we did in the case of Strassen's equations where A' was spanned by $\alpha, \alpha', \alpha''$.)

Remark 2.5.0.3. We see that both the classical equations and Strassen's equations are obtained by taking minors of a matrix whose entries are linear combinations of the coefficients of our tensor. Such constructions are part of a long tradition of finding *determinantal equations* for algebraic varieties that is out of the scope of this book. For the experts, given a variety X and a subvariety $Y \subset X$, one way to find defining equations for Y is to find vector bundles E, F over X and a vector bundle map $\phi : E \to F$ such that Y is realized as the *degeneracy locus* of ϕ, that is, the set of points $x \in X$ such that ϕ_x drops rank. Strassen's equations in the partially symmetric case had been discovered by Barth [**Bar77**] in this context. Variants of Strassen's equations date back to Frahm-Toeplitz [**Toe77**] and Aronhold [**Aro58**]. See [**Lan12**, §3.8.5] for a discussion. We will also see in §8.2 and §8.3.1 two different ways of deriving Strassen's equations via representation theory.

Let a_1, a_2, a_3 be a basis of A, with dual basis $\alpha^1, \alpha^2, \alpha^3$ of A^* so $T \in A \otimes B \otimes C$ may be written as $T = a_1 \otimes X_1 + a_2 \otimes X_2 + a_3 \otimes X_3$, where $X_j = T(\alpha_j)$. Then $T_A^\wedge$ will be expressed by a $3\mathbf{b} \times 3\mathbf{b}$ matrix. Ordering the basis of $A \otimes B^*$ by $a_3 \otimes \beta^1, \dots, a_3 \otimes \beta^{\mathbf{b}}, a_2 \otimes \beta^1, \dots, a_2 \otimes \beta^{\mathbf{b}}, a_1 \otimes \beta^1, \dots, a_1 \otimes \beta^{\mathbf{b}}$, and that of $\Lambda^2 A \otimes C$ by $(a_1 \wedge a_2) \otimes c_1, \dots, (a_1 \wedge a_2) \otimes c_{\mathbf{b}}, (a_1 \wedge a_3) \otimes c_1, \dots, (a_1 \wedge a_3) \otimes c_{\mathbf{b}}, (a_2 \wedge a_3) \otimes c_1, \dots, (a_2 \wedge a_3) \otimes c_{\mathbf{b}}$, we obtain the block matrix

$$T_A^\wedge = \begin{pmatrix} 0 & X_1 & -X_2 \\ X_2 & X_3 & 0 \\ X_1 & 0 & X_3 \end{pmatrix}. \tag{2.5.2}$$

Recall the following basic identity about determinants of blocked matrices (see, e.g., [**Pra94**, Thm. 3.1.1]), assuming the block W is invertible:

$$\det \begin{pmatrix} X & Y \\ Z & W \end{pmatrix} = \det(W) \det(X - YW^{-1}Z). \tag{2.5.3}$$

Block (2.5.2) $X = 0$, $Y = (X_1, -X_2)$, $Z = \begin{pmatrix} X_2 \\ X_1 \end{pmatrix}$, $W = \begin{pmatrix} X_3 & 0 \\ 0 & X_3 \end{pmatrix}$. Assume $X_3 = T(\alpha^3)$ is invertible to obtain

$$\det T_A^\wedge = \det(X_3)^2 \det(X_1 {X_3}^{-1} X_2 - X_2 {X_3}^{-1} X_1) \tag{2.5.4}$$

Equation (2.5.4) shows the new formulation is equivalent to the old, at least in the case of maximal rank. (We are only interested in the non-vanishing of the polynomial, not its values, so we can multply the inner matrix on the right by ${X_3}^{-1}$.) Equation (2.5.4) combined with Proposition 2.5.0.2 proves Theorem 2.2.2.1 in this case.

Note that here we have actual polynomials on $A{\otimes}B{\otimes}C$ (the minors of (2.5.2)), whereas in our original formulation of Strassen's equations we did not. To obtain polynomials in the original formulation one uses the adjugate matrix instead of the inverse, see [**Lan12**, §3.8].

2.6. Koszul flattenings

2.6.1. Their definition. The reformulation of Strassen's equations suggests the following generalization: let $\dim A = 2p+1$ and consider

$$T_A^{\wedge p} : B^*{\otimes}\Lambda^p A \to \Lambda^{p+1} A{\otimes}C \tag{2.6.1}$$

obtained by first taking $T_B{\otimes}\operatorname{Id}_{\Lambda^p}A : B^*{\otimes}\Lambda^p A \to \Lambda^p A{\otimes}A{\otimes}C$, and then projecting to $\Lambda^{p+1}A{\otimes}C$ as in Exercise 2.3.0.3.

If $\{a_i\}, \{b_j\}, \{c_k\}$ are bases of A, B, C and $T = \sum_{i,j,k} t^{ijk} a_i{\otimes}b_j{\otimes}c_k$, then

$$T_A^{\wedge p}(\beta{\otimes}f_1 \wedge \cdots \wedge f_p) = \sum_{i,j,k} t^{ijk}\beta(b_j) a_i \wedge f_1 \wedge \cdots \wedge f_p{\otimes}c_k. \tag{2.6.2}$$

The map $T_A^{\wedge p}$ is called a *Koszul flattening.* Note that if $T = a{\otimes}b{\otimes}c$ has rank one, then $\operatorname{rank}(T_A^{\wedge p}) = \binom{2p}{p}$ as the image is $a \wedge \Lambda^p A{\otimes}c$. By linearity of the map $T \mapsto T_A^{\wedge p}$ we conclude:

Proposition 2.6.1.1. [**LO15**] *Let $T \in A{\otimes}B{\otimes}C$ with $\dim A = 2p+1$. Then*

$$\underline{\mathbf{R}}(T) \geq \frac{\operatorname{rank}(T_A^{\wedge p})}{\binom{2p}{p}}.$$

Since the source (resp. target) has dimension $\binom{2p+1}{p}\mathbf{b}$ (resp. $\binom{2p+1}{p+1}\mathbf{c}$), assuming $\mathbf{b} \leq \mathbf{c}$, we potentially obtain equations for $\hat{\sigma}_r$ up to

$$r = \frac{\binom{2p+1}{p}\mathbf{b}}{\binom{2p}{p}} - 1 = \frac{2p+1}{p+1}\mathbf{b} - 1.$$

Just as with Strassen's equations (case $p = 1$), if $\dim A > 2p+1$, one obtains the best bound for these equations by restricting to subspaces of A^* of dimension $2p+1$.

Exercise 2.6.1.2: (2) Show that if $T_A^{\wedge p} : \Lambda^p A{\otimes}B^* \to \Lambda^{p+1}A{\otimes}C$ is injective, then $T_A^{\wedge q} : \Lambda^q A{\otimes}B^* \to \Lambda^{q+1}A{\otimes}C$ is injective for all $q < c$. ⊚

Next we would like to apply our new equations to matrix multiplication. In order to do so, we pause to better understand the matrix multiplication tensor.

2.6.2. The matrix multiplication tensor from an invariant perspective. We have

$$M_{\langle U,V,W\rangle} \in (U^*{\otimes}V){\otimes}(V^*{\otimes}W){\otimes}(W^*{\otimes}U).$$

If we think of matrix multiplication as a bilinear map, the input is a linear map from W to V and a linear map from V to U and the output is their composition, a linear map from W to U, i.e., an element of $W^*{\otimes}U$. If we think of it as a trilinear map, the inputs are three linear maps and the output a number.

Exercise 2.6.2.1: (2!) Show that matrix multiplication

$$M_{\langle U,V,W\rangle} : (U{\otimes}V^*)^* \times (V{\otimes}W^*)^* \to W{\otimes}U^*,$$

when viewed as a trilinear map

$$M_{\langle U,V,W\rangle} : (U^*{\otimes}V)^* \times (V^*{\otimes}W)^* \times (U{\otimes}W^*)^* \to \mathbb{C}.$$

is $(X,Y,Z) \mapsto \mathrm{trace}(XYZ)$. ⊚

Inside the space $V^*{\otimes}V$ of linear maps from V to itself, there is a canonical linear map, namely the identity map Id_V which just sends a vector to itself. If $v_1,\dots,v_{\mathbf{v}}$ is a basis of V with dual basis $v^1,\dots,v^{\mathbf{v}} \in V^*$, then $\mathrm{Id}_V = \sum_j v^j{\otimes}v_j$. One way to characterize the identity map up to scale, is that the line it spans is the unique line in $V^*{\otimes}V$ that is preserved by the action of $GL(V)$, where in matrices, $g\cdot X = gXg^{-1}$, or more invariantly, letting $\rho : GL(V) \to GL(V^*{\otimes}V)$ denote the inclusion map, $[\rho(g)(f)](v) := gf(g^{-1}(v))$.

Exercise 2.6.2.2: (1!) Show that as a tensor $M_{\langle U,V,W\rangle} = \mathrm{Id}_U \otimes \mathrm{Id}_V \otimes \mathrm{Id}_W$. ⊚

Exercise 2.6.2.3: (1) Show that $\mathrm{Id}_V \otimes \mathrm{Id}_W \in V{\otimes}V^*{\otimes}W{\otimes}W^* = (V{\otimes}W){\otimes}(V{\otimes}W)^*$ equals $\mathrm{Id}_{V\otimes W}$.

Exercise 2.6.2.4: (1!) Show that $M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}{\otimes}M_{\langle \mathbf{n}',\mathbf{m}',\mathbf{l}'\rangle} = M_{\langle \mathbf{nn}',\mathbf{mm}',\mathbf{ll}'\rangle}$. ⊚

A fancier proof that $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle}) \geq \mathbf{n}^2$, which will be useful for proving further lower bounds, is as follows: Write $A = U^*{\otimes}V$, $B = V^*{\otimes}W$, $C = W^*{\otimes}U$, so $(M_{\langle \mathbf{n}\rangle})_A : A^* \to B{\otimes}C$ is a map $U{\otimes}V^* \to V^*{\otimes}W{\otimes}W^*{\otimes}U$. This map is, for $f \in A^*$, $f \mapsto f{\otimes}\mathrm{Id}_W$, and thus is clearly injective. In other words, the map is $u{\otimes}\nu \mapsto \sum_k (\nu{\otimes}w_k){\otimes}(w^k{\otimes}u)$, where $w_1,\dots,w_{\mathbf{w}}$ is a basis of W with dual basis $w^1,\dots,w^{\mathbf{w}}$.

2.6.3. Koszul flattenings and matrix multiplication. When I want to emphasize the vector spaces involved, I write $M_{\langle U,V,W\rangle}$ for $M_{\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle}$. When $T = M_{\langle U,V,W\rangle}$, the Koszul flattening map is

$$(M_{\langle U,V,W\rangle})_A^{\wedge p} : V\otimes W^*\otimes\Lambda^p(U^*\otimes V) \to \Lambda^{p+1}(U^*\otimes V)\otimes(W^*\otimes U).$$

The presence of $\mathrm{Id}_W = \mathrm{Id}_{W^*}$ implies the map factors as $(M_{\langle U,V,W\rangle})_A^{\wedge p} = (M_{\langle \mathbf{u},\mathbf{v},1\rangle})_A^{\wedge p}\otimes \mathrm{Id}_{W^*}$, where

(2.6.3)

$$(M_{\langle \mathbf{u},\mathbf{v},1\rangle})_A^{\wedge p} : V\otimes\Lambda^p(U^*\otimes V) \to \Lambda^{p+1}(U^*\otimes V)\otimes U.$$

$$v\otimes(\xi^1\otimes e_1)\wedge\cdots\wedge(\xi^p\otimes e_p) \mapsto \sum_{s=1}^{\mathbf{u}} u_s\otimes(u^s\otimes v)\wedge(\xi^1\otimes e_1)\wedge\cdots\wedge(\xi^p\otimes e_p).$$

where $u_1,\dots,u_{\mathbf{u}}$ is a basis of U with dual basis $u^1,\dots,u^{\mathbf{u}}$ of U^*, so $\mathrm{Id}_U = \sum_{s=1}^{\mathbf{u}} u^s\otimes u_s$.

As discussed above, at first sight, Koszul flattenings could potentially prove a border rank lower bound of $2\mathbf{n}^2-3$ for $M_{\langle \mathbf{n}\rangle}$. However this does not happen, as there is a large kernal for the maps $M_{\langle \mathbf{n}\rangle}^{\wedge p}$ when $p\geq \mathbf{n}$. I first explain why this is the case. Let $\mathbf{u}=\mathbf{v}=\mathbf{n}$.

Let $p=\mathbf{n}$. Then

$$v\otimes(u^1\otimes v)\otimes\cdots\otimes(u^{\mathbf{n}}\otimes v) \mapsto \sum_j (u^j\otimes v)\wedge(u^1\otimes v)\otimes\cdots\otimes(u^{\mathbf{n}}\otimes v)\otimes u_j = 0,$$

so $M_{\langle \mathbf{n}\rangle}^{\wedge \mathbf{n}}$ is not injective. Since the map $M_{\langle \mathbf{u},\mathbf{v},1\rangle})_A^{\wedge p}$ commutes with the action of $GL(U)\times GL(V)$, by Schur's lemma 2.4.0.3, $\ker(M_{\langle \mathbf{n}\rangle}^{\wedge \mathbf{n}}) \subset V\otimes\Lambda^{\mathbf{n}}(U^*\otimes V)\subset V^{\otimes \mathbf{n}+1}\otimes U^{*\otimes \mathbf{n}}$ must be a submodule. It is clearly symmetric in V and skew in U^*, so the kernel must contain the irreducible submodule $\Lambda^{\mathbf{n}}U^*\otimes S^{\mathbf{n}+1}V$.

Exercise 2.6.3.1: (2) Show that $\ker(M_{\langle \mathbf{n},\mathbf{n},1\rangle})_A^{\wedge \mathbf{n}} = (\Lambda^{\mathbf{n}}U^*\otimes S^{\mathbf{n}+1}V)$. ⊚

Now consider the case $p=\mathbf{n}-1$. I claim $(M_{\langle \mathbf{n},\mathbf{n},1\rangle})_A^{\wedge \mathbf{n}-1}$ is injective. The following argument is due to L. Manivel. Say $X_1\otimes v_1+\cdots X_{\mathbf{n}}\otimes v_{\mathbf{n}} \in \ker(M_{\langle \mathbf{n},\mathbf{n},1\rangle})_A^{\wedge \mathbf{n}-1}$, i.e.,

$$\sum_s [X_1\wedge(u^s\otimes v_1)+\cdots+X_{\mathbf{n}}\wedge(u^s\otimes v_{\mathbf{n}})]\otimes u_s = 0.$$

Then for each s, each term in the brackets must be zero.

Lemma 2.6.3.2. *Let A be a vector space, let $X_1,\dots,X_k\in\Lambda^q A$, and let $a_1,\dots,a_k\in A$ be linearly independent. Then if $X_1\wedge a_1+\cdots+X_k\wedge a_k=0$, we may write each $X_j=\sum_{i=1}^k Y_{ij}\wedge a_i$ for some $Y_{ij}\in\Lambda^{q-1}A$.*

Exercise 2.6.3.3: (2) Prove Lemma 2.6.3.2.⊚

Remark 2.6.3.4. This is a special case of the generalized Cartan Lemma, see [**?**, §A.1]. With the aid of representation theory one can more precisely describe the Y_{ji}. (Use the sequence $0 \to S_{2,1^{q-1}}A \to \Lambda^q A \otimes A \to \Lambda^{q+1}A \to 0$.)

In our case, taking $s = 1$, we have $X_j = \sum Y_{j,(1,i)} \wedge (u^1 \otimes a_i)$, so each term in X_j is divisible by $(u^1 \otimes a_i)$ for some i, but then taking $s = 2$, we would also have to have each term in X_j is divisible by $(u^2 \otimes a_l)$ for some l, and continuing, if $p < \mathbf{n}$ we run out of factors, so there cannot be a kernel. In summary:

Proposition 2.6.3.5. *When $p < \mathbf{n}$, the map $(M_{\langle \mathbf{n},\mathbf{n},1\rangle})_A^{\wedge p}$ is injective.*

At this point one would like to say that if some $T^{\wedge p}$ is injective, then restricting to a generic $A' \subset A^*$, the map $T^{\wedge p}|_{\Lambda^p A' \otimes B^*} : \Lambda^p A' \otimes B^* \to \Lambda^{p+1}A' \otimes C$ would still be injective. Unfortunately I do not know how to prove this, because *a priori* $T^{\wedge p}|_{\Lambda^p A' \otimes B^*}$ injects into $[\Lambda^{p+1}A' \otimes C] \oplus [\Lambda^p A' \otimes (A/A') \otimes C]$, and it is not clear to me whether for generic A' it must remain injective when one projects to the first factor. What follows are two proofs that this is indeed the case for $(M_{\langle \mathbf{n},\mathbf{n},1\rangle})_A^{\wedge \mathbf{n}-1}$. The first is combinatorial. It has the advantages that it is elementary and will be used to prove the $2\mathbf{n}^2 - \lceil \log_2 \mathbf{n} \rceil - 1$ lower bound of §5.4.3. The second is geometrical. It has the advantage of being shorter and more elegant.

Theorem 2.6.3.6. [**LO15**] *Let $\mathbf{n} \leq \mathbf{m}$. Then*

$$\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}) \geq \frac{\mathbf{nl}(\mathbf{n}+\mathbf{m}-1)}{\mathbf{m}}.$$

In particular $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle}) \geq 2\mathbf{n}^2 - \mathbf{n}$.

I prove the case $\mathbf{n} = \mathbf{m}$ and leave the general case to the reader. We need to find $A' \subset A^*$ of dimension $2\mathbf{n} - 1$ such that, setting $\tilde{A} = A/A'^{\perp}$, $(M_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A' \otimes B^* \otimes C^*})_{\tilde{A}}^{\wedge \mathbf{n}-1}$ is injective.

First proof. Define the projection

$$\phi : A \to \mathbb{C}^{2\mathbf{n}-1} \tag{2.6.4}$$

$$x^i_j \mapsto e_{i+j-1}. \tag{2.6.5}$$

Let $e_S := e_{s_1} \wedge \cdots \wedge e_{s_{\mathbf{n}-1}}$, where $S = \{s_1, \ldots, s_{\mathbf{n}-1}\} \subset [2\mathbf{n}-1]$ has cardinality $\mathbf{n} - 1$. The map $(M_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A' \otimes B^* \otimes C^*})_{\tilde{A}}^{\wedge \mathbf{n}-1}$ is

$$e_S \otimes v_k \mapsto \sum_j \phi(u^j \otimes v_k) \wedge e_S \otimes u_j = \sum_j e_{j+k-1} \wedge e_S \otimes u_j.$$

Index a basis of the source by pairs (S, k), with $k \in [\mathbf{n}]$, and the target by (P, l) where $P \subset [2\mathbf{n}-1]$ has cardinality $\mathbf{n}$ and $l \in [\mathbf{n}]$.

We will choose an ordering of the basis vectors such that the resulting matrix is upper-triangular. Then we just need to show that each diagonal element of the matrix is nonzero to conclude. Unfortunately the order on (P, l) is a little complicated because e.g., if the l's are ordered sequentially, then to get a diagonal matrix, the P's must be given an order in the opposite direction.

Define an order relation on the target basis vectors as follows: For (P_1, l_1) and (P_2, l_2), set $l = \min\{l_1, l_2\}$, and declare $(P_1, l_1) < (P_2, l_2)$ if and only if

(1) In lexicographic order, the set of l minimal elements of P_1 is strictly after the set of l minimal elements of P_2 (i.e. the smallest element of P_2 is smaller than the smallest of P_1 or they are equal and the second smallest of P_2 is smaller or equal etc. up to l-th), or
(2) the l minimal elements in P_1 and P_2 are the same, and $l_1 < l_2$.
(3) the l minimal elements in P_1 and P_2 are the same, and $l_1 = l_2$, and the set of $\mathbf{n} - l$ tail elements of P_1 are after the set of $\mathbf{n} - l$ tail elements of P_2.

The third ordering is actually irrelevant - any breaking of a tie for the first two will lead to an upper-triangular matrix. Note that $(\{\mathbf{n}, \dots, 2\mathbf{n}-1\}, 1)$ is the unique minimal element for this relation and $([\mathbf{n}], \mathbf{n})$ is the unique maximal element. Note further that

$$e_{\mathbf{n}+1} \wedge \cdots \wedge e_{2\mathbf{n}-1} \otimes u_{\mathbf{n}} \mapsto e_{\mathbf{n}} \wedge \cdots \wedge e_{2\mathbf{n}-1} \otimes v_1$$

i.e., that

$$(\{\mathbf{n}+1, \dots, 2\mathbf{n}-1\}, \mathbf{n}) \mapsto (\{\mathbf{n}, \dots, 2\mathbf{n}-1\}, 1),$$

so $(\{\mathbf{n}+1, \dots, 2\mathbf{n}-1\}, \mathbf{n})$ will be our first basis element for the source. The order for the source is implicitly described in the proof.

The claim will follow by showing that the image is the span of all basis elements (P, l). Work by induction using the relation: the base case that $(\{\mathbf{n}, \dots, 2\mathbf{n}-1\}, 1)$ is in the image has been established. Let (P, l) be any basis element, and assume all (P', l') with $(P', l') < (P, l)$ have been shown to be in the image. Write $P = (p_1, \dots, p_{\mathbf{n}})$ with $p_i < p_{i+1}$. Consider the image of $(P \backslash \{p_l\}, 1 + p_l - l)$ which is

$$\sum_j \phi(u^j \otimes v_{1+p_l-l}) \wedge e_{P\backslash\{p_l\}} \otimes u_j = \sum_{\{j \mid j-l+p_l \notin P\backslash\{p_l\}\}} e_{p_l-l+j} \wedge e_{P\backslash\{p_l\}} \otimes u_j.$$

In particular, taking $j = l$ we see (P, l) is among the summands. If $j < l$, the contribution to the summand is a (P', j) where the first j terms of P' equal the first of P, so by condition (2), $(P', j) < (P, l)$. If $j > l$, the summand is a (P'', j) where the first $l-1$ terms of P and P'' agree, and the l-th terms are respectively p_l and $p_l - l + j$ so by condition (1) $(P'', j) < (P, l)$. $\square$

To illustrate, consider the first seven terms when $\mathbf{n}=3$:

$$(345,1),(345,2),(345,3),(245,1),(235,1),(234,1),(245,2),$$

where the order did not matter for the triple $(245,1),(235,1),(234,1)$. We have

$$\begin{aligned}
(45,3) &\mapsto (345,1)\\
(35,2) &\mapsto (345,2)\\
(34,3) &\mapsto (345,3)\\
(45,2) &\mapsto (245,1)+(345,2)\\
(35,2) &\mapsto (235,2)+(345,3)\\
(34,2) &\mapsto (234,1)\\
(25,3) &\mapsto (245,2)
\end{aligned}$$

Second proof. For this proof we take $\mathbf{u}=\mathbf{n}\leq\mathbf{v}=\mathbf{m}$. Take a vector space E of dimension 2, and fix isomorphisms $U\simeq S^{\mathbf{n}-1}E$, $V\simeq S^{\mathbf{m}-1}E^*$. Let $A'=S^{\mathbf{m}+\mathbf{n}-2}E^*\subset S^{\mathbf{n}-1}E^*\otimes S^{\mathbf{m}-1}E^*=U\otimes V^*$, and set $\tilde{A}=A/A'^{\perp}$. This turns out to be the same projection operator as in the previous proof. Here there is an $SL(E)$-module inclusion $\tilde{A}=S^{\mathbf{m}+\mathbf{n}-2}E\subset A$ because $SL(E)$ is reductive.

Our map is

$$\Lambda^{\mathbf{n}-1}(S^{\mathbf{m}+\mathbf{n}-2}E)\otimes S^{\mathbf{n}-1}E\text{trace }\Lambda^{\mathbf{n}}(S^{\mathbf{m}+\mathbf{n}-2}E)\otimes S^{\mathbf{m}-1}E^*$$

$$Q_1\wedge\cdots\wedge Q_{\mathbf{n}-1}\otimes f\mapsto\sum_{j=0}^{\mathbf{m}-1}(fh^h)\wedge Q_1\wedge\cdots\wedge Q_{\mathbf{n}-1}\otimes h_j$$

where $h^j=x^jy^{\mathbf{m}-j-1}$ and h_j is the dual basis vector.

Recall the contraction map from Exercise 2.3.0.4, for $\alpha\geq\beta$:

$$\begin{aligned}
S^{\alpha}E\times S^{\beta}E^*&\to S^{\alpha-\beta}E\\
(f,g)&\mapsto g\lrcorner f.
\end{aligned}$$

In the case $f=l^{\alpha}$ for some $l\in E$, then $g\lrcorner l^{\alpha}=g(l)l^{\alpha-\beta}$ (here $g(l)$ denotes g, considered as a polynomial, evaluated at the point l), so that $g\lrcorner l^{\alpha}=0$ if and only if l is a root of g.

Consider the transposed map

$$((M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}|_{A'\otimes U^*\otimes V^*})^{\wedge p}_{\tilde{A}})^T:$$

$$S^{\mathbf{m}-1}E^*\otimes\Lambda^{\mathbf{n}}S^{\mathbf{m}+\mathbf{n}-2}E\to S^{\mathbf{n}-1}E\otimes\Lambda^{\mathbf{n}-1}S^{\mathbf{m}+\mathbf{n}-2}E$$

$$g\otimes(f_1\wedge\cdots\wedge f_{\mathbf{n}})\mapsto\sum_{i=1}^{\mathbf{n}}(-1)^{i-1}(g\lrcorner f_i)\otimes f_1\wedge\cdots\hat{f}_i\cdots\wedge f_{\mathbf{n}}.$$

The map $((M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}|_{A'\otimes U^*\otimes V^*})^{\wedge p}_{\tilde{A}})^T$ is surjective: Let $l^{\mathbf{n}-1}\otimes(l_1^{\mathbf{m}+\mathbf{n}-2}\wedge\cdots\wedge l_{\mathbf{n}-1}^{\mathbf{m}+\mathbf{n}-2})\in S^{\mathbf{n}-1}E\otimes\Lambda^{\mathbf{n}-1}S^{\mathbf{m}+\mathbf{n}-2}E$ with $l, l_i\in E$. Such elements span the target so it will be sufficient to show any such element is in the image. Assume first that l is distinct from the l_i. Since $\mathbf{n}\leq\mathbf{m}$, there is a polynomial $g\in S^{\mathbf{m}-1}E^*$ which vanishes on $l_1,\dots,l_{\mathbf{n}-1}$ and is nonzero on l. Then, up to a nonzero scalar, $g\otimes(l_1^{\mathbf{m}+\mathbf{n}-2}\wedge\cdots\wedge l_{\mathbf{n}-1}^{\mathbf{m}+\mathbf{n}-2}\wedge l^{\mathbf{m}+\mathbf{n}-2})$ maps to our element.

The condition that l is distinct from the l_i may be removed by taking limits, as the image of a linear map is closed. $\square$

The above result begs the question: did we fail to get a better bound because this is the best bound Koszul flattenings can give, or is there something pathological about matrix multiplication that prevented the full power of Koszul flattenings? That is, perhaps the Koszul flattenings for $\mathbb{C}^m\otimes\mathbb{C}^m\otimes\mathbb{C}^m$ could be trivial beyond border rank $2m-\sqrt{m}$. This is not the case:

Theorem 2.6.3.7. [**Lan15**] *The maximal minors of the Koszul flattening $T_A^{\wedge p}:\Lambda^p\mathbb{C}^{2p+1}\otimes(\mathbb{C}^{2p+2})^*\to\Lambda^{p+1}\mathbb{C}^{2p+1}\otimes\mathbb{C}^{2p+2}$ give nontrivial equations for $\hat{\sigma}_r\subset\mathbb{C}^{2p+1}\otimes\mathbb{C}^{2p+2}\otimes\mathbb{C}^{2p+2}$, the tensors of border rank at most r in $\mathbb{C}^{2p+1}\otimes\mathbb{C}^{2p+2}\otimes\mathbb{C}^{2p+2}$, up to $r=4p+1$.*

For $\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$, this implies that when $\mathbf{m}$ is even (resp. odd), the equations are nontrivial up to $r=2\mathbf{m}-3$ (resp. $r=2\mathbf{m}-5$).

Exercise 2.6.3.8: (1!!) Prove the theorem. ⊚

2.6.4. Koszul flattenings in coordinates. To prove lower bounds on the rank of matrix multiplication, and to facilitate a comparison with Griesser's equations discussed in §5.2.2, it will be useful to view $T_A^{\wedge p}$ in coordinates. Let $\dim A=2p+1$. Write $T=a_0\otimes X_0+\cdots+a_{2p}\otimes X_{2p}$ where a_j is a basis of A with dual basis α^j and $X_j=T(\alpha^j)$. An expression of $T_A^{\wedge p}$ in bases is as follows: write $a_I:=a_{i_1}\wedge\cdots\wedge a_{i_p}$ for the induced basis elements of Λ^pA, require that the first $\binom{2p}{p-1}$ basis vectors of Λ^pA have $i_1=0$, that the second $\binom{2p}{p}$ do not, and call these multi-indices $0J$ and K. Order the bases of $\Lambda^{p+1}A$ such that the first $\binom{2p}{p+1}$ multi-indices do not have 0, and the second $\binom{2p}{p}$ do, and furthermore that the second set of indices is ordered the same way as K is ordered, only we write $0K$ since a zero index is included. The resulting matrix is of the form

$$\begin{pmatrix}0 & Q\\ \tilde{Q} & R\end{pmatrix}\tag{2.6.6}$$

where this matrix is blocked $(\binom{2p}{p+1}\mathbf{b}, \binom{2p}{p}\mathbf{b}) \times (\binom{2p}{p+1}\mathbf{b}, \binom{2p}{p}\mathbf{b})$,

$$R = \begin{pmatrix} X_0 & & \\ & \ddots & \\ & & X_0 \end{pmatrix},$$

and $Q, \tilde{Q}$ have entries in blocks consisting of $X_1, \dots, X_{2p}$ and zero. Thus if X_0 is of full rank and we change coordinates such that it is the identity matrix, so is R and the determinant equals the determinant of $Q\tilde{Q}$ by (2.5.3). If we order the appearances of the K multi-indices such that the j-th K is the complement of the j-th J in $[2p]$, then $Q\tilde{Q}$ will be skew-symmetric. When $p = 1$, $Q\tilde{Q} = [X_1, X_2]$, and when $p = 2$ we recover the matrix (2.2.1).

In general $Q\tilde{Q}$ is a block skew-symmetric $\binom{2p}{p-1}\mathbf{b} \times \binom{2p}{p-1}\mathbf{b}$ matrix whose block entries are either zero or commutators $[X_i, X_j]$. Each $[X_i, X_j]$ appears (up to sign) $\binom{2p-1}{2}$ times, and each block row and column contain exactly $\binom{2p-1}{2}$ non-zero blocks, so the resulting matrix is very sparse.

2.7. Lower bounds for the rank of matrix multiplication

2.7.1. The results. Most tensors have rank equal to border rank, in the sense that the set of tensors of rank greater than r in $\hat{\sigma}_r$ is a proper subvariety. Matrix multiplication is expected to have larger rank than border rank when $\mathbf{n} > 2$ because of its enormous symmetry group, as explained in Chapter 4.

The key to the rank lower bound is that our proof of the border rank lower bound used equations of relatively low degree because of the factorization $(M_{\langle \mathbf{n} \rangle})_A^{\wedge p} = (M_{\langle \mathbf{n},\mathbf{n},1 \rangle})_A^{\wedge p} \otimes \mathrm{Id}_W$, so we were considering minors of a size $\binom{2\mathbf{n}-1}{\mathbf{n}}\mathbf{n}$ matrix instead of a size $\binom{2\mathbf{n}-1}{\mathbf{n}}\mathbf{n}^2$ matrix. I will show that if a low degree polynomial is nonzero on $M_{\langle \mathbf{n} \rangle}$, and $M_{\langle \mathbf{n} \rangle}$ has an optimal rank decomposition $M_{\langle \mathbf{n} \rangle} = \sum_{j=1}^r a_j \otimes b_j \otimes c_j$, then the polynomial is already zero on a subset of the summands. This is a variant of the *substitution method* discussed in §5.3.

Here is a $3\mathbf{n}^2 - o(\mathbf{n}^2)$ lower bound for $\mathbf{R}(M_{\langle \mathbf{n} \rangle})$ that follows from the method:

Theorem 2.7.1.1. [**Lan14b**] *Let* $p < \mathbf{n} - 1$. *Then*

$$\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{m} \rangle}) \geq \frac{2p+1}{p+1}\mathbf{n}\mathbf{m} + \mathbf{n}^2 - (2p+1)\binom{2p+1}{p}\mathbf{n}.$$

This gives a bound of the form $\mathbf{R}(M_{\langle \mathbf{n} \rangle}) \geq 3\mathbf{n}^2 - o(\mathbf{n}^2)$ by taking, e.g., $p = \log(\log(\mathbf{n}))$.

2.7.2. Proof of Theorem 2.7.1.1. We will need a few facts from algebraic geometry before the proof.

The following standard Lemma, also used in [**Blä03**], appears in this form in [**Lan12**, Lemma 11.5.0.2]:

Lemma 2.7.2.1. *Given a polynomial P of degree d on $\mathbb{C}^{\mathbf{a}}$, there exists a subset of basis vectors $\{e_{i_1}, \dots, e_{i_d}\}$ such that $P\mid_{\langle e_{i_1},\dots,e_{i_d}\rangle}$ is not identically zero.*

In other words, there exists a coordinate subspace $\mathbb{C}^d \subset \mathbb{C}^{\mathbf{a}}$ such that $\mathbb{C}^d \not\subset \mathrm{Zeros}(P)$.

The lemma follows by simply choosing the basis vectors from a degree d monomial that appears in P. For example, Lemma 2.7.2.1 implies that a surface in $\mathbb{P}^3$ defined by a degree two equation cannot contain six lines whose pairwise intersections span $\mathbb{P}^3$.

The proof of the theorem will use a famous algebraic variety, the *Grassmannian*:

$$G(k,V) := \mathbb{P}\{T \in \Lambda^k V \mid \exists v_1,\dots,v_k \in V \text{ such that } T = v_1\wedge\cdots\wedge v_k\} \subset \mathbb{P}\Lambda^k V.$$

The Grassmannian admits the geometric interpretation as the space parametrizing the k-planes through the origin in V via the correspondence $[v_1 \wedge \cdots \wedge v_k] \leftrightarrow \mathrm{span}\{v_1,\dots,v_k\}$.

The following exercise shows that the Grassmannian is indeed an algebraic variety. It can be safely skipped on a first reading.

Exercise 2.7.2.2: (3) The Grassmannian is the zero set of equations parametrized by $\Lambda^{k-2j}V^*\otimes\Lambda^{k+2j}V^*$ for $1 \leq j \leq \min\{\lfloor\frac{\mathbf{v}-k}{2}\rfloor, \lfloor\frac{k}{2}\rfloor\}$ as follows: for $\mu \in \Lambda^{k-2j}V^*$ and $\zeta \in \Lambda^{k+2j}V^*$, recall Exercise 2.3.0.5, and consider $T\lrcorner\zeta \in \Lambda^{2j}V^*$ and $\mu\llcorner T \in \Lambda^{2j}V$. Define $P_{\mu\otimes\zeta}(T) := \langle T\lrcorner\zeta, \mu\llcorner T\rangle$, the evaluation of an element of $\Lambda^{2j}V^*$ on an element of $\Lambda^{2j}V$. Note that these are quadratic equations in the coefficients of T. Show that the zero set of these equations is the Grassmannian. ⊚

Lemma 2.7.2.3. *Let A be given a basis. Given a non-zero homogeneous polynomial of degree d on $\Lambda^k A$ that is not in $I(G(k,A))$ and assume $dk < \dim A$, there exist dk basis vectors of A such that, denoting their dk-dimensional span by $\tilde{A}$, P restricted to $G(k,\tilde{A})$ is not identically zero.*

Proof. Consider the map $f : A^{\times k} \to \hat{G}(k,A)$ given by $(a_1,\dots,a_k) \mapsto a_1 \wedge \cdots \wedge a_k$. Then f is surjective. Take the polynomial P and pull it back by f. Here the *pullback* $f^*(P)$ is defined by $f^*(P)(a_1,\dots,a_k) := P(f(a_1,\dots,a_k))$. The pullback is of degree d in each copy of A. (I.e., fixing $k-1$ of the a_j, it becomes a degree d polynomial in the k-th.) Now apply Lemma 2.7.2.1 k times to see that the pulled back polynomial is not identically zero restricted to the span of these vectors, denoted $\tilde{A}$, and thus P restricted to $\hat{G}(k,\tilde{A})$ is not identically zero. $\square$

Remark 2.7.2.4. The bound in Lemma 2.7.2.3 is sharp, as give A a basis $a_1,\dots,a_{\mathbf{a}}$ and consider the polynomial on $\Lambda^k A$ with coordinates $x^I = x^{i_1}\cdots x^{i_k}$ corresponding to the vector $\sum_I x^I a_{i_1}\wedge\cdots\wedge a_{i_k}$:

$$P = x^{1,\dots,k}x^{k+1,\dots,2k}\cdots x^{(d-1)k+1,\dots,dk}.$$

Then P restricted to $G(k,\langle a_1,\dots,a_{dk}\rangle)$ is non-vanishing but there is no smaller subspace spanned by basis vectors on which it is non-vanishing.

Proof of Theorem 2.7.1.1. Say $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle}) = r$ and write an optimal expression

$$M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle} = \sum_{j=1}^{r} a_j\otimes b_j\otimes c_j. \tag{2.7.1}$$

We will show that the Koszul-flattening equation is already non-zero restricted to a subset of this expression for a judicious choice of $\tilde{A}\subset A$ of dimension $2p+1$ with $p<\mathbf{n}-1$. Then the rank will be at least the border rank bound plus the number of terms not in the subset. Here are the details:

Write $\tilde{A}=A/A'^{\perp}$. Define

$$P_{2p+1} : G(2p+1,A^*)\to\mathbb{C}$$
$$A'\mapsto \det((M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle}|_{A'\otimes B^*\otimes C^*})^{\wedge p}_{\tilde{A}} : \Lambda^p\tilde{A}\otimes B^*\to\Lambda^{p+1}\tilde{A}\otimes C).$$

I claim that P_{2p+1} is not identically zero for all $p\le\mathbf{n}-1$. ***This proof still needs fixing*** To see this we work by downward induction. By the proof of Theorem 2.6.3.6, the claim holds in the case $p=\mathbf{n}-1$. Assume we have proved the claim down to p. Let $\tilde{A}$ have dimension $2p+1$ and assume $P_{2p+1}(\tilde{A})\neq 0$. For each $A'\subset A^*$, write $\tilde{A}=\tilde{A}_1\oplus\tilde{A}_2$ where $\dim(\tilde{A}_1)=2p-1$ and $\dim\tilde{A}_2=2$. By Exercise 2.6.1.2, $\Lambda^{p-1}\tilde{A}\otimes B^*\to\Lambda^p\tilde{A}\otimes C$ is injective. We have

$$\begin{array}{ccc}
\Lambda^{p-1}(\tilde{A}_1\oplus\tilde{A}_2)\otimes B^* & \longrightarrow & \Lambda^{p}(\tilde{A}_1\oplus\tilde{A}_2)\otimes C\\
\| & & \|\\
\Lambda^{p-3}\tilde{A}_1\otimes\Lambda^2\tilde{A}_2\otimes B^* & & \Lambda^{p-2}\tilde{A}_1\otimes\Lambda^2\tilde{A}_2\otimes C\\
\oplus\Lambda^{p-2}\tilde{A}_1\otimes\tilde{A}_2\otimes B^* & \longrightarrow & \oplus\Lambda^{p-1}\tilde{A}_1\otimes\tilde{A}_2\otimes C\\
\oplus\Lambda^{p-1}\tilde{A}_1\otimes B^* & & \oplus\Lambda^{p}\tilde{A}_1\otimes C
\end{array}$$

Since the top horizontal arrow is injective, the bottom must be as well. But only $\Lambda^{p-1}\tilde{A}_1\otimes B^*$ maps to $\Lambda^p\tilde{A}_1\otimes C$ so the map must be an isomorphism.

Now P is a polynomial of degree $\binom{2p+1}{p}\mathbf{nm}>\mathbf{nm}$, so at first sight, e.g., when $\mathbf{m}\sim\mathbf{n}$, Lemma 2.7.2.3 will be of no help because $dk>\dim A=\mathbf{n}^2$, but since

$$(M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle}|_{A'\otimes B^*\otimes C^*})^{\wedge p}_{\tilde{A}} = (M_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A'\otimes V\otimes U^*})^{\wedge p}_{\tilde{A}}\otimes \mathrm{Id}_{W^*},$$

we actually have $P = \tilde{P}^{\mathbf{m}}$, where

$$\tilde{P} : G(2p+1, A) \to \mathbb{C}$$
$$\tilde{A} \mapsto \det((M_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A'\otimes V\otimes U^*})^{\wedge p}_{\tilde{A}} : \Lambda^p \tilde{A}\otimes V \to \Lambda^{p+1}\tilde{A}\otimes U).$$

Hence we may work with $\tilde{P}$ which is of degree $\binom{2p+1}{p}\mathbf{n}$ which will be less than $\mathbf{n}^2$ if p is sufficiently small. Since $(M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle})_A : A^* \to B\otimes C$ is injective, some subset of the a_j forms a basis of A. Lemma 2.7.2.3. implies that there exists a subset of those basis vectors of size $dk = \binom{2p+1}{p}\mathbf{n}(2p+1)$, such that if we restrict to terms of the expression (2.7.1) that use only a_j whose expansion in the fixed basis has nonzero terms from that subset of dk basis vectors, calling the sum of these terms M', we have $\underline{\mathbf{R}}(M') \geq \frac{2p+1}{p+1}\mathbf{nm}$. Let M'' be the sum of the remaining terms in the expression. There are at least $\mathbf{a} - dk = \mathbf{n}^2 - \binom{2p+1}{p}\mathbf{n}(2p+1)$ of the a_j appearing in M'' (the terms corresponding to the complementary basis vectors). Since we assumed we had an optimal expression for $M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle}$, we have

$$\begin{aligned}\mathbf{R}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{m}\rangle}) &= \mathbf{R}(M') + \mathbf{R}(M'')\\ &\geq \frac{2p+1}{p+1}\mathbf{nm} + [\mathbf{n}^2 - (2p+1)\binom{2p+1}{p}\mathbf{n}].\end{aligned}$$

□

2.7.3. Improved lower bounds on the rank. Further lower bounds are obtained by lowering the degree of the polynomial by localizing the equations. An easy such localization is to set $X_0 = \mathrm{Id}$ which reduces the determinant of (2.6.6) to that of (2.2.1) when $p = 2$ and yields a similar reduction of degree in general. Further localizations both reduce the degree and the size of the Grassmannian, both of which improve the error term. The state of the art is:

Theorem 2.7.3.1. [**MR13**] *Let $p \leq \mathbf{n}$ be a natural number. Then*

$$\mathbf{R}(M_{\mathbf{n},\mathbf{n},\mathbf{m}}) \geq (1 + \frac{p}{p+1})\mathbf{nm} + \mathbf{n}^2 - \big(2\binom{2p}{p+1} - \binom{2p-2}{p-1} + 2\big)\mathbf{n}. \tag{2.7.2}$$

When $\mathbf{n} = \mathbf{m}$,

$$\mathbf{R}(M_{\langle \mathbf{n}\rangle}) \geq (3 - \frac{1}{p+1})\mathbf{n}^2 - \big(2\binom{2p}{p+1} - \binom{2p-2}{p-1} + 2\big)\mathbf{n}. \tag{2.7.3}$$

For example, when $p = 1$ one recovers Bläser's bound of $\frac{5}{2}\mathbf{n}^2 - 3\mathbf{n}$. When $p = 3$, the bound (2.7.3) becomes $\frac{11}{4}\mathbf{n}^2 - 26\mathbf{n}$, which improves Bläser's for

$\mathbf{n} \geq 132$. A modification of the method also yields $\mathbf{R}(M_{\langle \mathbf{n} \rangle}) \geq \frac{8}{3}\mathbf{n}^2 - 7\mathbf{n}$. See [**MR13, Lan14b**] for proofs of the modifications of the error terms.

Chapter 3

The complexity of matrix multiplication II: asymptotic upper bounds

This chapter discusses progress towards the astounding conjecture that asymptotically, the complexity of multiplying two $\mathbf{n} \times \mathbf{n}$ matrices is nearly the same as the complexity of adding them. I cover the main advances in upper bounds for the exponent of matrix multiplication beyond Strassen's original discovery in 1969: the 1979 upper bound $\omega < 2.79$ of Bini et. al., the 1981 bound $\omega \leq 2.55$ of Schönhage, the 1987 bound $\omega < 2.48$ of Strassen, and the Coppersmith-Winograd 1990 bound $\omega < 2.38$, emphasizing a geometric perspective. I mention recent "explanations" as to why progress essentially stopped in 1990 from [**?**] and in Chapter 4 I discuss other potential paths for upper bounds.

The exponent ω of matrix multiplication is naturally defined in terms of tensor rank:

$$\omega := \inf\{\tau \in \mathbb{R} \mid \mathbf{R}(M_{\langle \mathbf{n} \rangle}) = O(\mathbf{n}^{\tau})\}.$$

See [**BCS97**, §15.1] for a the proof that tensor rank yields the same exponent as other complexity measures.

The above-mentioned conjecture is that $\omega = 2$. The only general tool for determining tensor rank that I am aware of is the substitution method discussed in §5.3, which is too weak for the purposes of estimating ω. However, as I explain in §3.2, Bini et. al. showed that one may also define the exponent in terms of border rank, namely (see Proposition 3.2.1.11)

$$\omega = \inf\{\tau \in \mathbb{R} \mid \underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) = O(\mathbf{n}^\tau)\},$$

Unfortunately, the state of the art for border rank is also woefully short of what would be needed to determine the exponent.

One bit of good news is that we do not need to work asymptotically to get upper bounds on ω. Theorem 3.2.1.11 states that for all $\mathbf{n}$, $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \geq \mathbf{n}^\omega$.

Another small help is that we may also use rectangular matrix multiplication to prove upper bounds on ω: part of Proposition 3.2.1.11, states that for all $\mathbf{l}, \mathbf{m}, \mathbf{n}$,

$$\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle}) \geq (\mathbf{lmn})^{\frac{\omega}{3}}.$$

In order to make this transition from rank to border rank, we will need a basic result in algebraic geometry. Because of this, I begin, in §3.1 with some basic facts from the subject.

To really improve the situation, one needs further techniques that enable one to avoid dealing with tensors beyond the range we understand. After the work of Bini et. al., *all upper bounds on* ω *are obtained via tensors other than* $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n} \rangle}$.

The next advance in upper bounds, due to Schönhage (Theorem 3.3.3.1) and described in §3.3, is more involved: it says it is sufficient to prove upper bounds on sums of *disjoint* matrix multiplications:

The inequalities regarding ω above are strict, e.g., there does not exist $\mathbf{n}$ with $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle})$ equal to $\mathbf{n}^\omega$. (This does not rule out $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle})$ equal to $2\mathbf{n}^\omega$ for all $\mathbf{n}$.) Thus one can only obtain upper bounds on ω when working with a fixed $\mathbf{n}$. One way to extend the above methods is to find sequences of sums $\oplus_{i=1}^{s(N)} M_{\langle \mathbf{l}_i(N),\mathbf{m}_i(N)\mathbf{n}_i(N) \rangle}$ with the border rank of the sums giving upper bounds on ω. This is one aspect of Strassen's "laser method" described in §3.4. A second new ingredient of his method is that instead of dealing with the sum of a collection of disjoint rectangular matrix multiplications, one looks for a tensor $T \in A \otimes B \otimes C$, that has special combinatorial structure rendering it easy to study, that can be degenerated into a collection of disjoint matrix multiplications. More precisely, to obtain a sequence of disjoint matrix multiplication tensors, one degenerates the tensor powers $T^{\otimes N} \in (A^{\otimes N}) \otimes (B^{\otimes N}) \otimes (C^{\otimes N})$. Strassen's degeneration is in the sense of the $GL(A) \times GL(B) \times GL(C)$-orbit closure of $T^{\otimes N}$.

After Strassen, all other subsequent upper bounds on ω use what I will call *combinatorial restrictions* of $T^{\otimes N}$ for some "simple" tensor T, where entries of a coordinate presentation of $T^{\otimes N}$ are simply set equal to zero. The choice of entries to zero out is subtle. I describe these developments in §3.4.

In addition to combinatorial restrictions, Cohn et. al. exploit a geometric change of basis when a tensor is the multiplication tensor of an algebra (or even more general structures). They use the discrete Fourier transform for finite groups (and more general structures) to show that the multiplication tensor in the Fourier basis (and thus in any basis) has "low" rank, but nevertheless in the standard basis admits a combinatorial restriction to a "large" sum of matrix multiplication tensors. I discuss this approach in §3.5.

The proofs in this chapter make essential use of the property from Exercise 2.6.2.4:

$$M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}\otimes M_{\langle \mathbf{l}',\mathbf{m}',\mathbf{n}'\rangle} = M_{\langle \mathbf{ll}',\mathbf{mm}',\mathbf{nn}'\rangle} \tag{3.0.1}$$

where for tensors $T \in A\otimes B\otimes C$ and $T' \in A'\otimes B'\otimes C'$, $T\otimes T'$ is considered as a tensor in the triple tensor product $(A\otimes A')\otimes(B\otimes B')\otimes(C\otimes C')$.

3.1. Facts and definitions from algebraic geometry

Standard references for this material are [**Har95, Mum95, Sha94**]. The first is very good for examples, the second and third have clean proofs, with the proofs in the second more concise and those in the third more elementary.

3.1.1. Projective varieties. Varieties in a vector space V defined by homogeneous polynomials are invariant under rescaling. For this, and other reasons, it will be convenient to work in projective space $\mathbb{P}V := (V\backslash 0)/\sim$ where $v \sim w$ if and only if $v = \lambda w$ for some $\lambda \in \mathbb{C}\backslash 0$. Write $\pi : V\backslash 0 \to \mathbb{P}V$ for the projection map. For $X \subset \mathbb{P}V$, write $\pi^{-1}(X) \cup \{0\} =: \hat{X} \subset V$, and $\pi(y) = [y]$. If $\hat{X} \subset V$ is a variety, I will also refer to $X \subset \mathbb{P}V$ as a variety. More precisely, the zero set in V of a collection of polynomials on V is called an *affine variety* and the image in $\mathbb{P}V$ of the zero set of a collection of homogeneous polynomials on V is called a *projective variety*. For subsets $Z \subset V$, $\mathbb{P}Z \subset \mathbb{P}V$ denotes its image under π. A variety X is said to be *irreducible* if it is not possible to non-trivially write $X = Y \cup Z$ with Y, Z varieties. If $P \in S^dV^*$ is an irreducible polynomial, then its zero set $\mathrm{Zeros}(P) \subset \mathbb{P}V$ is an irreducible variety, called a *hypersurface of degree d*. For a variety $X \subset \mathbb{P}V$, $I_d(X) := \{P \in S^dV^* \mid X \subset \mathrm{Zeros}(P)\}$ denotes the ideal of X in degree d, and $I(X) = \oplus_d I_d(X) \subset Sym(V^*)$ is the ideal of X.

We will be mostly concerned with varieties in spaces of tensors (for the study of matrix multiplication) and spaces of polynomials (for geometric complexity theory).

3.1.2. Examples of varieties.

(1) Projective space $\mathbb{P}V \subseteq \mathbb{P}V$.

(2) The *Segre variety* of rank one tensors

$$\sigma_1 = Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)$$
$$:= \mathbb{P}\{T \in A_1 \otimes \cdots \otimes A_n \mid \exists a_j \in A_j \text{ such that } T = a_1 \otimes \cdots \otimes a_n\} \subset \mathbb{P}(A_1 \otimes \cdots \otimes A_n).$$

(3) The *Veronese variety*

$$v_d(\mathbb{P}V) = \mathbb{P}\{P \in S^d V \mid P = x^d \text{ for some } x \in V\} \subset \mathbb{P}S^d V.$$

(4) The *Grassmannian*

$$G(k, V) := \mathbb{P}\{T \in \Lambda^k V \mid \exists v_1, \ldots, v_k \in V \text{ such that } T = v_1 \wedge \cdots \wedge v_k\} \subset \mathbb{P}\Lambda^k V.$$

(5) The *Chow variety*

$$Ch_d(V) := \overline{\mathbb{P}\{P \in S^d V \mid \exists v_1, \ldots, v_d \in V \text{ such that } P = v_1 \cdots v_d\}} \subset \mathbb{P}S^d V.$$

By definition, projective space is a variety (the zero set of no equations).

Exercise 3.1.2.1: (2) Show that $Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)$ is the zero set of the size two minors of the *flattenings* $A_j^* \to A_1 \otimes \cdots \otimes \hat{A}_j \otimes \cdots \otimes A_n$, for $1 \leq j \leq n$.

To get equations for $v_d(\mathbb{P}V)$, given $P \in S^d V$, consider the *flattening* $P_{1,d-1} : V^* \to S^{d-1}V$ defined by $\frac{\partial}{\partial v} \mapsto \frac{\partial P}{\partial v}$. For example when $d = 4$, $\mathbf{v} = 2$ and $P = \sum_{i=0}^{4} p_i x^i y^{4-i}$, the matrix representing $P_{1,3}$ is

$$\begin{pmatrix} p_4 & p_3 & p_2 & p_1 \\ p_3 & p_2 & p_1 & p_0 \end{pmatrix} \tag{3.1.1}$$

and $v_4(\mathbb{P}^1)$ is the zero set of the 6 size two minors of this matrix.

Exercise 3.1.2.2: (1) Show that $v_d(\mathbb{P}V)$ is the zero set of the size two minors of the flattening $V^* \to S^{d-1}V$.

We saw equations for the Grassmannian in §2.7.2.

Exercise 3.1.4.2 will show that it is not necessary to take the Zariski closure when defining the Chow variety. Equations for the Chow variety are known, see §9.1.6. However generators of the ideal of the Chow variety are not known explicitly.

3.1.3. Dimension via tangent spaces. Informally, the dimension of a variety is the number of parameters needed to describe it locally. For example, the dimension of $\mathbb{P}V$ is $\mathbf{v} - 1$ because in coordinates on the open neighborhood where $x_1 \neq 0$, points of $\mathbb{P}V$ have a unique expression as $[1, x_2, \ldots, x_{\mathbf{v}}]$, where $x_2, \ldots, x_{\mathbf{v}}$ are free parameters.

I first define dimension of a variety via dimensions of vector spaces. Define the *affine tangent space* to $X \subset \mathbb{P}V$ at $[x] \in X$, $\hat{T}_x \hat{X} = \hat{T}_{[x]} X \subset V$, to be the span of the tangent vectors $x'(0)$ to analytic curves $x(t)$ on $\hat{X}$ with $x(0) = x$, and note that this is independent of the choice of $x \in [x]$. A point

$x \in \hat{X}$ is defined to be a *smooth* point if $\dim \hat{T}_y \hat{X}$ is constant for all y in some neighborhood of x.

The *dimension* of an irreducible variety $\hat{X} \subset V$ is the dimension of the tangent space at a smooth point of $\hat{X}$. If x is a smooth point, $\dim X = \dim \hat{X} - 1 = \dim \hat{T}_x \hat{X} - 1$. If x is not a smooth point, it is called a *singular point* and we let $X_{\text{sing}} \subset X$ denote the singular points of X.

Exercise 3.1.3.1: (2) Show that $\dim\{\det_n = 0\}_{\text{sing}} = n^2 - 4$.

A variety of dimension one is called a *curve*.

If a Zariski open subset of a variety is given parametrically, then one can calculate the tangent space to the variety via the parameter space. For example $\hat{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ may be thought of as the image of the map

$$\begin{aligned} A \times B \times C &\to A \otimes B \otimes C \\ (a, b, c) &\mapsto a \otimes b \otimes c, \end{aligned}$$

so to compute $\hat{T}_{[a\otimes b\otimes c]} Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$, take curves $a(t) \subset A$ with $a(0) = a$ and similarly for B, C, then $\frac{d}{dt}|_{t=0} a(t)\otimes b(t)\otimes c(t) = a'\otimes b\otimes c + a\otimes b'\otimes c + a\otimes b\otimes c'$ by the Leibnitz rule. Since a' can be any vector in A and similarly for b', c' we conclude

$$\hat{T}_{[a\otimes b\otimes c]} Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) = A\otimes b\otimes c + a\otimes B\otimes c + a\otimes b\otimes C.$$

The right hand side spans a space of dimension $\mathbf{a}+\mathbf{b}+\mathbf{c}-2$, so $\dim(Seg(\mathbb{P}A\times \mathbb{P}B \times \mathbb{P}C)) = \mathbf{a}+\mathbf{b}+\mathbf{c}-3$.

I can now pay off two debts: in §2.1.1, I asserted that the fundamental Theorem of linear algebra is something of a miracle, and in Theorem 2.1.5.1 I asserted that a general tensor in $\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$ has tensor rank around $\frac{\mathbf{m}^2}{3}$.

It is straight-forward to compute

$$\hat{T}_{[a_1\otimes b_1\otimes c_1 + a_2\otimes b_2\otimes c_2]}\sigma_2 = \\ \text{span}\{a_1\otimes b_1\otimes c_1' + a_1\otimes b_1'\otimes c_1 + a_1'\otimes b_1\otimes c_1 + a_2\otimes b_2\otimes c_2' + a_2\otimes b_2'\otimes c_2 + a_2'\otimes b_2\otimes c_2\}$$

so that $\dim \sigma_2 \leq 2(\dim(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)) + 2 - 1$ (and equality clearly holds if $\mathbf{a}, \mathbf{b}, \mathbf{c} \geq 3$) and similarly $\dim \sigma_r \leq r(\dim(Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C))+r-1$. The first chance this has to be the entire ambient space is when this number is $\mathbf{abc}-1$. When $\mathbf{a}=\mathbf{b}=\mathbf{c}=\mathbf{m}$, this means $r \geq \frac{\mathbf{m}^3}{3\mathbf{m}-1}$, paying the second debt.

For the first,

$$\begin{aligned} \hat{T}_{[a_1\otimes b_1 + a_2\otimes b_2]}\sigma_{2,A\otimes B} &= \text{span}\{a_1\otimes b_1' + a_1'\otimes b_1 + a_2\otimes b_2' + a_2'\otimes b_2\} \\ &= A\otimes \text{span}\{b_1, b_2\} + \text{span}\{a_1, a_2\}\otimes B \end{aligned}$$

and this space has dimension $2\dim Seg(\mathbb{P}A \times \mathbb{P}B)$, instead of the expected $2\dim Seg(\mathbb{P}A \times \mathbb{P}B) + 1$. This accounts for the semi-continuity of matrix rank which fails for tensor rank: any point on a tangent line, i.e., a point of the form $a'\otimes b + a\otimes b'$ is also transparently on a secant line, i.e., the sum of two rank one matrices.

Exercise 3.1.3.2: (1) Compute $\hat{T}_{[}x^d]v_d(\mathbb{P}V)$.

3.1.4. Noether normalization. Consider the curve $\{xy = 1\} \subset \mathbb{C}^2$:

picture here*

If we project the curve onto the x-axis, we get the set $\{x \in \mathbb{C} \mid x \neq 0\}$, which, as was discussed in §1.1.14, is not Zariski closed.

One of the many wonderful things about projective space is that the projection of an algebraic variety to a hyperplane is still an algebraic variety:

Theorem 3.1.4.1. *If $X \subset \mathbb{P}W$ is a variety, $L \subset W$ is a subspace with $\mathbb{P}L \cap X = \emptyset$, and one considers the projection map $p : W \to W/L$, then $\mathbb{P}p(\hat{X}) \subset \mathbb{P}(W/L)$ is also a variety.*

Theorem 3.1.4.1 is part of the *Noether normalization* theorem (see, e.g., [**Sha94**, §5.4] or [**Mum95**, §2C]). It is proved via elimination theory. In addition to failing in affine space, this projection property fails over $\mathbb{R}$: the surface in $\mathbb{RP}^3$ given by $x^2 + z^2 - y^2 = 0$ when projected from $[1, 0, 0]$ is not a real algebraic variety.

Exercise 3.1.4.2: (1) Show that if $W = V^{\otimes d}$ and L is the $GL(V)$-complement to S^dV in $V^{\otimes d}$, taking $p : V^{\otimes d} \to V^{\otimes d}/L \simeq S^dV$, then $p(Seg(\mathbb{P}V \times \cdots \times \mathbb{P}V)) = Ch_d(V)$. Conclude that the Chow variety is indeed a variety.

For those wishing to understand the projection algebraically, say one projects from a point. Give $\mathbb{P}V$ linear coordinates such that the point is a coordinate point. Then, from the ideal of $X \subset \mathbb{P}V$, eliminate the coordinate from equations to get a new ideal in $\mathbf{v} - 1$ variables. For example, give $S^4\mathbb{C}^2$ coordinates $(p_4, p_3, p_2, p_1, p_0)$ as above and project from p_2. Eliminating p_2 from the equations

$$p_4p_2 - p_3^2, p_4p_1 - p_2p_3, p_4p_0 - p_1p_3, p_3p_1 - p_2^2, p_2p_0 - p_1^2$$

gives the ideal generated by

$$p_4p_0 - p_1p_3, p_3^3 - p_4^2p_1, p_1^3 - p_0^2p_3.$$

Exercise 3.1.4.3: (2) What equations does one get when projecting from p_3? Give a geometric explanation why the answer is different. (A complete answer to this question is beyond what we have covered, I am just asking for some equations.) ⊚

Remark 3.1.4.4. Since the elimination theory doesn't care which point one projects from, one can even project from a point on a variety. The resulting "map" is not defined at the point one projects from, but the Zariski closure of the image of the points where it is defined at is well defined. This is an example of what is called a *rational map*.

Exercise 3.1.4.5: (2) What ideal does one get when projecting $v_4(\mathbb{P}^1)$ from p_4? (A complete answer to this question is beyond what we have covered, I am just asking for some equations.) ⊚

As long as X does not surject onto $\mathbb{P}V/L$, we can continue projecting it to smaller and smaller projective spaces.

If $X \subset \mathbb{P}V$ is a projective variety and $f : X \to Y \subset \mathbb{P}^N$ is given by $N+1$ homogeneous polynomials on V, then f is an example of a *regular map*. (Regular maps are defined in greater generality, essentially maps defined locally by polynomials.)

Exercise 3.1.4.6: (1) Show that if X is irreducible and $f : X \to Y$ is regular, then $f(X)$ is irreducible.

Theorem 3.1.4.1 generalizes to:

Theorem 3.1.4.7. *(see, e.g.,* [**Sha13**, §5.2, Thm. 1.10]*) If X is a projective variety and $f : X \to Y$ is a regular map, then $f(X)$ is Zariski closed.*

In particular, if X is irreducible, then $f(X)$ is an irreducible variety.

3.1.5. Dimension via projection. The dimension of $X \subset \mathbb{P}V$ is also the largest integer n such that there exists a surjective linear projection onto a $\mathbb{P}^n$. In this case the surjective projection $X \to \mathbb{P}(V/\mathbb{C}^c)$ is finite to one. The integer $c = \mathbf{v} - 1 - n$ is called the *codimension* of X in $\mathbb{P}V$. Noether normalization implies that a general linear space $\mathbb{P}L$ will satisfy $\dim(X \cap \mathbb{P}L) = \mathbf{v} - 1 - n - \dim \mathbb{P}L$. Similarly the intersection of X with a general linear space of dimension $c+1$ will be a finite number of points. This number of points is called the *degree* of X.

A consequence of this more algebraic definition of dimension is the following result:

Theorem 3.1.5.1. *Let $X, Y \subset \mathbb{P}^N$ (resp. $X, Y \subset \mathbb{C}^N$) be irreducible projective (resp. affine) varieties.*

Then any non-empty component Z of $X \cap Y$ has $\dim Z \geq \dim X + \dim Y - N$.

Moreover, in the projective case, if $\dim X + \dim Y - N > 0$, then $X \cap Y \neq \emptyset$.

For the proof, see, e.g., [**Sha94**, I.6 Thm. 6].

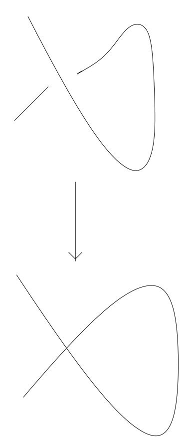
3.1.6. Zariski and Euclidean closure. Recall from §1.1.14.2 that the Zariski closure of a set can be larger than the Euclidean closure. Nevertheless, the following theorem, proved using Noether normalization, shows that in our situation, the competing definitions of closure agree:

Theorem 3.1.6.1. *Let $Z \subset \mathbb{P}V$ be a subset. Then the Euclidean closure of Z is contained in the Zariski closure of Z. If Z contains a Zariski open subset of its Zariski closure, then the two closures coincide. The same assertions hold for subsets $Z \subset V$.*

A proof that uses nothing but Noether normalization is given in [**Mum95**, Thm. 2.33]. I present a proof using the following basic fact: for every irreducible algebraic curve $C \subset \mathbb{P}V$ there exists a smooth algebraic curve $\tilde{C}$ and a surjective algebraic map $\pi : \tilde{C} \to C$ that is one-to-one over the smooth points of C. (More precisely, π is a *finite map* in the sense of algebraic geometry.) See, e.g., [**Sha94**, §II.5, Thms. 3 and 6] for a proof. The curve $\tilde{C}$ is called the *normalization* of C.

The theorem will follow immediately from the following Lemma:

Lemma 3.1.6.2. *Let $Z \subset \mathbb{P}V$ be an irreducible variety and let $Z^0 \subset Z$ be a Zariski open subset. Let $p \in Z \backslash Z^0$. Then there exists an analytic curve $C(t)$ such that $C(t) \in Z^0$ for all $t \neq 0$ and $\lim_{t\to 0} C(t) = p$.*



Proof. Let c be the codimension of Z and take a general linear space $\mathbb{P}L \subset \mathbb{P}V$ of dimension $c+1$ that contains p. Then $\mathbb{P}L \cap Z$ will be a possibly reducible algebraic curve containing p. Take a component C of the curve that contains p. If p is a smooth point of the curve we are done, as we can expand a Taylor series about p. Otherwise take the the normalization $\pi : \tilde{C} \to C$ and a point of $\pi^{-1}(p)$, expand a Taylor series about that point and compose with π to obtain the desired analytic curve. $\square$

3.2. The upper bounds of Bini, Capovani, Lotti, and Romani

3.2.1. Rank, border rank, and the exponent of matrix multiplication.

Proposition 3.2.1.1. [**Bin80**] *For all* $\mathbf{n}$, $\mathbf{R}(M_{\langle \mathbf{n} \rangle}) \geq \mathbf{n}^{\omega}$, *i.e.,* $\omega \leq \frac{\log \mathbf{R}(M_{\langle \mathbf{n} \rangle})}{\log(\mathbf{n})}$.

Proof. By the definitions of the exponent and O, there exists a constant C, such that $C\mathbf{R}(M_{\langle \mathbf{n} \rangle}) \geq \mathbf{n}^{\omega}$ for all $\mathbf{n}$. By (3.0.1) and Exercise 2.1.6.3, $\mathbf{R}(M_{\langle \mathbf{n}^k \rangle}) \leq \mathbf{R}(M_{\langle \mathbf{n} \rangle})^k$. Say $\mathbf{R}(M_{\langle \mathbf{n} \rangle}) = r$. Then $Cr^k \geq (\mathbf{n}^k)^{\omega}$, i.e. $C^{\frac{1}{k}} r \geq \mathbf{n}^{\omega}$. Now let k go to infinity, we get $r \geq \mathbf{n}^{\omega}$. $\square$

Remark 3.2.1.2. The calculation in the proof of Proposition 3.2.1.1 is typical in the upper bound literature and will show up several times in this chapter: one has an initially hazardous constant (in this case C) that gets washed out asymptotically by taking high tensor powers of $M_{\langle \mathbf{n} \rangle}$.

Proposition 3.2.1.3. *For all* $\mathbf{l}, \mathbf{m}, \mathbf{n}$, $(\mathbf{lmn})^{\frac{\omega}{3}} \leq \mathbf{R}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle})$, *i.e.,* $\omega \leq \frac{3 \log \mathbf{R}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle})}{\log(\mathbf{mnl})}$.

Exercise 3.2.1.4: (2) Prove Proposition 3.2.1.3. ⊚

Remark 3.2.1.5. The inequalities in Propositions 3.2.1.1 and 3.2.1.3 are strict, see Theorem 3.3.3.5.

To show that ω may also be defined in terms of border rank, introduce a sequence of ranks that interpolate between rank and border rank.

We say $\mathbf{R}_h(T) \leq r$ if there exists an expression

$$T = \lim_{\epsilon \to 0} \frac{1}{\epsilon^h} (a_1(\epsilon) \otimes b_1(\epsilon) \otimes c_1(\epsilon) + \cdots + a_r(\epsilon) \otimes b_r(\epsilon) \otimes c_r(\epsilon)) \tag{3.2.1}$$

where $a_j(\epsilon), b_j(\epsilon), c_j(\epsilon)$ are analytic functions of ϵ.

Proposition 3.2.1.6. $\underline{\mathbf{R}}(T) \leq r$ *if and only if there exists an* h *such that* $\mathbf{R}_h(T) \leq r$.

Proof. We need to show $\underline{\mathbf{R}}(T) \leq r$ implies there exists an h with $\mathbf{R}_h(T) \leq r$. Since $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is just the product of three projective spaces, every curve in $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is of the form $[a(t) \otimes b(t) \otimes c(t)]$ for some curves $a(t) \subset A$ etc., and if the curve is analytic, the functions $a(t), b(t), c(t)$ can be taken to be analytic as well. Thus every analytic curve in $\sigma_r^0(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ may be written as $[\sum_{j=1}^r a_j(t) \otimes b_j(t) \otimes c_j(t)]$ for some analytic curves $a_j(t) \subset A$ etc. Since the Euclidean and Zariski closures of $\hat{\sigma}_r^0$ agree, we conclude that if $T \in \hat{\sigma}_r$, then $\mathbf{R}_h(T) \leq r$ for h equal to the order of first nonzero term in the Taylor expansion of $\sum_{j=1}^r a_j(t) \otimes b_j(t) \otimes c_j(t)$. $\square$

Remark 3.2.1.7. In the matrix multiplication literature, e.g. [**BCS97**], $\hat{\sigma}_r$ is often defined to be the the set of T with $\mathbf{R}_h(T) \leq r$ for some h. One then must show that this set is algebraically closed.

Proposition 3.2.1.8. *If $\mathbf{R}_h(T) \leq r$, then $\mathbf{R}(T) \leq r\binom{h+2}{2} < rh^2$.*

Proof. Write T as in (3.2.1). Then T is the coefficient of the ϵ^h term of the expression in parentheses. For each summand, there is a contribution of $\sum_{\alpha+\beta+\gamma=h}(\epsilon^\alpha a_\alpha) \otimes (\epsilon^\beta b_\beta) \otimes (\epsilon^\gamma c_\gamma)$ which consists of $\binom{h+2}{2}$ terms. □

Remark 3.2.1.9. In fact $\mathbf{R}(T) \leq rh$, see Proposition 3.5.3.2.

Exercise 3.2.1.10: (1) Show that for $T \in A \otimes B \otimes C$, if $\mathbf{R}_h(T) \leq r$, then $\mathbf{R}_{Nh}(T^{\otimes N}) \leq r^N$ where $T^{\otimes N}$ is considered as an element of the triple tensor product $(A^{\otimes N}) \otimes (B^{\otimes N}) \otimes (C^{\otimes N})$.

Theorem 3.2.1.11. *[Bini, [**Bin80**]] For all $\mathbf{l}, \mathbf{m}, \mathbf{n}$, $\omega \leq \frac{3 \log \underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle})}{\log(\mathbf{mnl})}$. In particular, for all $\mathbf{n}$, $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \geq \mathbf{n}^\omega$.*

Proof. Write $r = \underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l} \rangle})$. Set $N = \mathbf{mnl}$. We have $\mathbf{R}_h(M_{\langle N \rangle}) \leq r^3$ for some h and thus $\mathbf{R}(M_{\langle N^k \rangle}) \leq r^{3k}(hk)^2$, which implies

$$(N^k)^\omega \leq r^{3k}(hk)^2$$

so

$$N^\omega \leq r^3(hk)^{\frac{2}{k}}$$

and letting $k \to \infty$ gives the result. □

3.2.2. Bini et. al's algorithm. Recall from §2.1.4 that $\underline{\mathbf{R}}(M^{red}_{\langle 2 \rangle}) \leq 5$.

Exercise 3.2.2.1: (1) Use that $\underline{\mathbf{R}}(M^{red}_{\langle 2 \rangle}) \leq 5$ to show $\underline{\mathbf{R}}(M_{\langle 2,2,3 \rangle}) \leq 10$. More generally, show that if $\underline{\mathbf{R}}(M^{red}_{\langle \mathbf{m},2,2 \rangle}) = r$ and $\underline{\mathbf{R}}(M^{red}_{\langle \mathbf{m}',2,2 \rangle}) = r'$, then setting $n = m + m' - 1$, $\underline{\mathbf{R}}(M_{\langle n,2,2 \rangle}) \leq r + r'$.⊚

Using Proposition 3.2.1.11 we conclude:

Theorem 3.2.2.2. [**BCRL79**] $\omega < 2.78$.

3.3. Schönhage's upper bounds

The next contribution to upper bounds for the exponent of matrix multiplication was Schönhage's discovery that the border rank of the sum of two tensors in disjoint spaces can be smaller than the sum of the border ranks, and that this failure could be exploited to prove further upper bounds on the exponent. This result enables one to prove upper bounds with tensors that are easier to analyze because of their low border rank. Before giving Schönhage's bounds, I begin with geometric preliminaries on orbit closures.

3.3.1. Orbit closures. Orbit closures will play a central role in our study of GCT. They also play a role in the work of Schönhage and Strassen on matrix multiplication, so I make several remarks in this context here.

When $r \leq \mathbf{a}_i$ for $1 \leq i \leq n$, $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ is an *orbit closure*: Let $a_j^{\alpha_j}$, $1 \leq \alpha_j \leq \mathbf{a}_j$, be a basis of A_j,

$$\begin{aligned}&\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))\\ &= \overline{GL(A_1) \times \cdots \times GL(A_n) \cdot [a_1^1 \otimes \cdots \otimes a_n^1 + \cdots + a_1^r \otimes \cdots \otimes a_n^r]} \subset \mathbb{P}(A_1 \otimes \cdots \otimes A_n).\end{aligned}$$

In particular,

$$\sigma_r(Seg(\mathbb{P}^{r-1} \times \mathbb{P}^{r-1} \times \mathbb{P}^{r-1})) = \overline{GL_r \times GL_r \times GL_r \cdot [M_{\langle 1 \rangle}^{\oplus r}]}. \tag{3.3.1}$$

Exercise 3.3.1.1: (2) Let V be a G-module and let $v, w \in V$. Show that $w \in \overline{G \cdot v}$ if and only if $\overline{G \cdot w} \subseteq \overline{G \cdot v}$.

Proposition 3.3.1.2. *If $T' \in \overline{GL(A) \times GL(B) \times GL(C) \cdot T} \subset A \otimes B \otimes C$, then $\underline{\mathbf{R}}(T') \leq \underline{\mathbf{R}}(T)$.*

Exercise 3.3.1.3: Prove Proposition 3.3.1.2. ⊚

Definition 3.3.1.4. If $T' \in \overline{GL(A) \times GL(B) \times GL(C) \cdot T} \subset A \otimes B \otimes C$, we say T' is a *degeneration* of T.

Consider the orbit closure of the matrix multiplication tensor

$$\overline{GL(A) \times GL(B) \times GL(C) \cdot [M_{\langle U,V,W \rangle}]} \subset \mathbb{P}(A \otimes B \otimes C).$$

Write $M_{\langle 1 \rangle}^{\oplus r} = \sum_{j=1}^r a_j \otimes b_j \otimes c_j \in \mathbb{C}^r \otimes \mathbb{C}^r \otimes \mathbb{C}^r$ where $\{a_j\}, \{b_j\}, \{c_j\}$ are bases. This tensor is sometimes called the *unit tensor*.

By Exercise 3.3.1.1, we may rephrase our characterization of border rank as, taking inclusions $A, B, C \subset \mathbb{C}^r$,

$$\begin{aligned}\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \leq r &\Leftrightarrow [M_{\langle \mathbf{n} \rangle}] \in \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))\\ &\Leftrightarrow \overline{GL_r \times GL_r \times GL_r \cdot [M_{\langle \mathbf{n} \rangle}]} \subset \sigma_r(Seg(\mathbb{P}^{r-1} \times \mathbb{P}^{r-1} \times \mathbb{P}^{r-1}))\\ &\Leftrightarrow \overline{GL_r \times GL_r \times GL_r \cdot [M_{\langle \mathbf{n} \rangle}]} \subset \overline{GL_r \times GL_r \times GL_r \cdot [M_{\langle 1 \rangle}^{\oplus r}]}\end{aligned}$$

3.3.2. Schönhage's example. Recall from Exercise 2.1.7.6 that $\underline{\mathbf{R}}(M_{\langle 1,\mathbf{m},\mathbf{n} \rangle}) = \mathbf{mn}$ and $\underline{\mathbf{R}}(M_{\langle N,1,1 \rangle}) = N$. Recall the notation from §2.1.6 that if $T_1 \in A_1 \otimes B_1 \otimes C_1$ and $T_2 \in A_2 \otimes B_2 \otimes C_2$, we define the tensor $T_1 \oplus T_2 \in (A_1 \oplus A_2) \otimes (B_1 \oplus B_2) \otimes (C_1 \oplus C_2)$. (In Exercise 5.3.1.6 you will show that $\mathbf{R}(M_{\langle 1,\mathbf{m},\mathbf{n} \rangle} \oplus M_{\langle N,1,1 \rangle}) = \mathbf{mn} + N$.)

Theorem 3.3.2.1 (Schönhage [**Sch81**])**.** *Set $N = (\mathbf{n}-1)(\mathbf{m}-1)$. Then*

$$\underline{\mathbf{R}}(M_{\langle 1,\mathbf{m},\mathbf{n} \rangle} \oplus M_{\langle N,1,1 \rangle}) = \mathbf{mn} + 1.$$

Proof. By conciseness, we only need to show $\underline{\mathbf{R}}(M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}\oplus M_{\langle N,1,1\rangle})\leq \mathbf{mn}+1$. Write

$$M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}=\sum_{i=1}^{\mathbf{m}}\sum_{j=1}^{\mathbf{n}}x_i\otimes y_j\otimes z_{i,j},$$
$$M_{\langle N,1,1\rangle}=\sum_{u=1}^{\mathbf{m}-1}\sum_{v=1}^{\mathbf{n}-1}x_{u,v}\otimes y_{u,v}\otimes z.$$

Then

$$\begin{aligned}M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}\oplus M_{\langle N,1,1\rangle}=\lim_{t\to 0}\frac{1}{t^2}[&\sum_{u=1}^{\mathbf{m}-1}\sum_{v=1}^{\mathbf{n}-1}(x_u+tx_{uv})\otimes(y_v+ty_{uv})\otimes(z+t^2z_{uv})\\ &+\sum_{u=1}^{\mathbf{m}-1}x_u\otimes(y_{\mathbf{n}}+t(-\sum_v y_{uv}))\otimes(z+t^2z_{u\mathbf{n}})\\ &+\sum_{v=1}^{\mathbf{n}-1}(x_{\mathbf{m}}+t(-\sum_u x_{uv}))\otimes y_v\otimes(z+t^2z_{\mathbf{m}v})\\ &+x_{\mathbf{m}}\otimes y_{\mathbf{n}}\otimes(z+t^2z_{\mathbf{mn}})-(\sum_i x_i)\otimes(\sum_s y_s)\otimes z].\end{aligned}$$

$\square$

Write $M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}\in A_1\otimes B_1\otimes C_1$ and $M_{\langle N,1,1\rangle}\in A_2\otimes B_2\otimes C_2$. One way to understand the proof is as follows: If one takes a curve in $Seg(\mathbb{P}(A_1\oplus A_2)\times\mathbb{P}(B_1\oplus B_2)\times\mathbb{P}(C_1\oplus C_2))$ with zero-th order terms in $A_1\otimes B_1\otimes C_2$, and takes one derivative, one can have terms in $A_1\otimes B_1\otimes C_1$ and after two derivatives, one can have terms in both $A_1\otimes B_1\otimes C_1$ and $A_2\otimes B_2\otimes C_2$. The zero-th order terms must be arranged to all cancel. Schönhage accomplishes this in the simplest possible way: he takes dimensions sufficiently unbalanced that there are more terms than the dimension of $A_1\otimes B_1\otimes C_2$, so they are linearly dependent and easily arranged to cancel. What is more subtle is the cancellation of the first order terms, whose geometry I leave to the reader to explore.

3.3.3. Schönhage's asymptotic sum inequality. To develop intuition how an upper bound on a sum of matrix multiplications could give an upper bound on a single matrix multiplication, say we knew $\underline{\mathbf{R}}(M_{\langle\mathbf{n}\rangle}^{\oplus s})\leq r$ with $s\leq\mathbf{n}^3$. Then to compute $M_{\langle\mathbf{n}^2\rangle}$ we could write $M_{\langle\mathbf{n}^2\rangle}=M_{\langle\mathbf{n}\rangle}\otimes M_{\langle\mathbf{n}\rangle}$. At worst this is evaluating $\mathbf{n}^3$ disjoint copies of $M_{\langle\mathbf{n}\rangle}$. Now group these $\mathbf{n}^3$ disjoint copies in groups of s and apply the bound to obtain a savings.

Here is the precise statement:

Theorem 3.3.3.1. [**Sch81**] *[Schönhage's asymptotic sum inequality] For all* $\mathbf{l}_i, \mathbf{m}_i, \mathbf{n}_i$, *with* $1 \leq i \leq s$*:*

$$\sum_{i=1}^{s} (\mathbf{m}_i \mathbf{n}_i \mathbf{l}_i)^{\frac{\omega}{3}} \leq \underline{\mathbf{R}}(\bigoplus_{i=1}^{s} M_{\langle \mathbf{m}_i, \mathbf{n}_i, \mathbf{l}_i \rangle}).$$

The main step of the proof, and an outline of the rest of the argument is given below.

Remark 3.3.3.2. A similar result (also proven in [**Sch81**]) holds for the border rank of the multiplication of matrices with some entries equal to zero, where the product $\mathbf{m}_i \mathbf{n}_i \mathbf{l}_i$ is replaced by the number of multiplications in the naïve algorithm for the matrices with zeros.

Here is a special case that isolates the new ingredient (following notes of M. Bläser [**Blä13**]):

Lemma 3.3.3.3.

$$\mathbf{n}^{\omega} \leq \lceil \frac{\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}^{\oplus s})}{s} \rceil.$$

In particular, $s\mathbf{n}^{\omega} \leq \underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}^{\oplus s})$.

Proof. Let $r = \underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}^{\oplus s})$. It is sufficient to show that for all N,

$$\underline{\mathbf{R}}(M_{\langle \mathbf{n}^N \rangle}^{\oplus s}) \leq \lceil \frac{r}{s} \rceil^N s \tag{3.3.2}$$

as then, since trivially $\underline{\mathbf{R}}(M_{\langle \mathbf{n}^N \rangle}^{\oplus s}) \geq \underline{\mathbf{R}}(M_{\langle \mathbf{n}^N \rangle}) \geq (\mathbf{n}^N)^{\omega}$, we have

$$(\mathbf{n}^N)^{\omega} \leq \lceil \frac{r}{s} \rceil^N s$$

i.e.,

$$\mathbf{n}^{\omega} \leq \lceil \frac{r}{s} \rceil s^{\frac{1}{N}}$$

and the result follows letting $N \to \infty$.

We prove (3.3.2) by induction on N. The hypothesis is the case $N = 1$. Assume (3.3.2) holds up to N and observe that

$$M_{\langle \mathbf{n}^{N+1} \rangle}^{\oplus s} = M_{\langle \mathbf{n} \rangle}^{\oplus s} \otimes M_{\langle \mathbf{n}^N \rangle}.$$

Now $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}^{\oplus s}) \leq r$ implies $M_{\langle \mathbf{n} \rangle}^{\oplus s} \in \overline{GL_r^{\times 3} \cdot M_{\langle 1 \rangle}^{\oplus r}}$ by Equation (3.3.1), so $M_{\langle \mathbf{n} \rangle}^{\oplus s} \otimes M_{\langle \mathbf{n}^N \rangle} \in \overline{GL_r^{\times 3} \cdot M_{\langle 1 \rangle}^{\oplus r} \otimes M_{\langle \mathbf{n}^N \rangle}}$. Thus $\underline{\mathbf{R}}(M_{\langle \mathbf{n}^{N+1} \rangle}^{\oplus s}) \leq \underline{\mathbf{R}}(M_{\langle 1 \rangle}^{\oplus r} \otimes M_{\langle \mathbf{n}^N \rangle})$.

Recall that $M_{\langle 1\rangle}^{\oplus t}\otimes M_{\langle \mathbf{n}^N\rangle}=M_{\langle \mathbf{n}^N\rangle}^{\oplus t}$. Now

$$\begin{aligned}\underline{\mathbf{R}}(M_{\langle \mathbf{n}^{N+1}\rangle}^{\oplus s}) &\leq \underline{\mathbf{R}}(M_{\langle \mathbf{n}^N\rangle}^{\oplus r})\\ &\leq \underline{\mathbf{R}}(M_{\langle \mathbf{n}^N\rangle}^{\oplus \lceil \frac{r}{s}\rceil s})\\ &\leq \underline{\mathbf{R}}(M_{\langle 1\rangle}^{\oplus \lceil \frac{r}{s}\rceil}\otimes M_{\langle \mathbf{n}^N\rangle}^{\oplus s})\\ &\leq \underline{\mathbf{R}}(M_{\langle 1\rangle}^{\oplus \lceil \frac{r}{s}\rceil})\underline{\mathbf{R}}(M_{\langle \mathbf{n}^N\rangle}^{\oplus s})\\ &\leq \lceil \frac{r}{s}\rceil(\lceil \frac{r}{s}\rceil^N s)\end{aligned}$$

where the last inequality follows from the induction hypothesis. □

The general case of Theorem 3.3.3.1 essentially follows from the above lemma and arguments used previously: one first takes a high tensor power of the sum, then switches to rank at the price of introducing an h that washes out in the end. The new tensor is a sum of products of matrix multiplications that one converts to a sum of matrix multiplications. One then takes the worst term in the summation and estimates with respect to it (multiplying by the number of terms in the summation), and applies the lemma to conclude.

Corollary 3.3.3.4. [**Sch81**] $\omega < 2.55$.

Proof. Applying Theorem 3.3.3.1 to $\underline{\mathbf{R}}(M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}\oplus M_{\langle(\mathbf{m}-1)(\mathbf{n}-1),1,1\rangle})=\mathbf{mn}+1$ gives

$$(\mathbf{mn})^{\frac{\omega}{3}}+((\mathbf{m}-1)(\mathbf{n}-1))^{\frac{\omega}{3}}\leq \mathbf{mn}+1$$

and taking $\mathbf{m}=\mathbf{n}=4$ gives the result. □

In [**CW82**] they prove that for any tensor T that is a direct sum of disjoint matrix multiplications, if $\mathbf{R}(T)\leq r$, then there exists N such that $\underline{\mathbf{R}}(T\oplus M_{\langle N,1,1\rangle})\leq r+1$. This, combined with our earlier arguments using $\mathbf{R}_h$ to bridge the gap between rank and border rank asymptotically, implies the inequality in Theorem 3.3.3.1 is strict:

Theorem 3.3.3.5. [**CW82**] *For all* $\mathbf{l}_i,\mathbf{m}_i,\mathbf{n}_i$, *with* $1\leq i\leq s$*:*

$$\sum_{i=1}^{s}(\mathbf{m}_i\mathbf{n}_i\mathbf{l}_i)^{\frac{\omega}{3}}<\underline{\mathbf{R}}(\bigoplus_{i=1}^{s}M_{\langle \mathbf{m}_i,\mathbf{n}_i,\mathbf{l}_i\rangle}).$$

In particular, for all $\mathbf{n}$, $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle})>\mathbf{n}^\omega$, so one cannot determine ω from $M_{\langle \mathbf{n}\rangle}$ for any fixed $\mathbf{n}$.

3.4. Strassen's laser method

3.4.1. Introduction. Recall our situation: we don't understand rank or even border rank in the range we would need to prove upper bounds on ω via $M_{\langle \mathbf{n} \rangle}$, so we showed upper bounds on ω could be proved first with rectangular matrix multiplication, then with sums of disjoint matrix multiplications which had the property that the border rank of the sum was less than the sum of the border ranks, and the border rank in each case was determined via an explicit algorithm.

We also saw that to determine the exponent by such methods, one would need to deal with sequences of tensors. Strassen's laser method is based on taking high tensor powers of a fixed tensor, and then degenerating it to a disjoint sum of matrix multiplication tensors. Because it deals with sequences, there is no known obstruction to determining ω exactly via Strassen's method.

Starting with Strassen's method, all attempts to determine ω aim at best for a Pyrrhic victory in the sense that even if ω were determined by these methods, they would not give any indication as to what would be optimally fast matrix multiplication for any given size matrix.

3.4.2. Strassen's tensor. Consider the following tensor

$$T_{STR} = \sum_{j=1}^{q} a_0 \otimes b_j \otimes c_j + a_j \otimes b_0 \otimes c_j \in \mathbb{C}^{q+1} \otimes \mathbb{C}^{q+1} \otimes \mathbb{C}^q. \tag{3.4.1}$$

This is presented as a sum of $2q$ rank one tensors. (And we will see $\mathbf{R}(T_{STR}) = 2q$ in §**??**.) Nevertheless, $\underline{\mathbf{R}}(T_{STR}) = q + 1$. To see why one could expect this, consider the q points $a_0 \otimes b_0 \otimes c_j$. The tensor T_{STR} is a sum of tangent vectors to these q points:

$$T_{STR} = \sum_{j=1}^{q} \lim_{t \to 0} [(a_0 + t a_j) \otimes (b_0 + t b_j) \otimes c_j - a_0 \otimes b_0 \otimes c_j]$$

Note that the sum $\sum_j a_0 \otimes b_0 \otimes c_j$ is also a rank one tensor, which leads one to the expression:

$$\lim_{t \to 0} \sum_{j=1}^{q} (a_0 + t a_j) \otimes (b_0 + t b_j) \otimes c_j - a_0 \otimes b_0 \otimes (c_1 + \cdots + c_q)$$

showing the border rank is at most $q + 1$, but since the tensor is concise, we obtain equality. Geometrically, the original q points all lie on the linear space $[a_0 \otimes b_0 \otimes \mathbb{C}^q] \subset Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$.

Now consider $\tilde{T} := T_{STR}\otimes\sigma(T_{STR})\otimes\sigma^2(T_{STR})$ where σ is a cyclic permutation of the three factors. Group triples of spaces together to consider $\tilde{T} \in \mathbb{C}^{q(q+1)^2}\otimes\mathbb{C}^{q(q+1)^2}\otimes\mathbb{C}^{q(q+1)^2}$. We have the upper bound $\underline{\mathbf{R}}(\tilde{T}) \leq (q+1)^3$.

Write $a_{\alpha\beta\gamma} := a_\alpha\otimes a_\beta\otimes a_\gamma$ and similarly for b's and c's. Then, omitting the $\otimes$'s:

(3.4.2)

$$\tilde{T} = \sum_{i,j,k=1}^{q} (a_{ij0}b_{0jk}c_{i0k} + a_{ijk}b_{0jk}c_{i00} + a_{ij0}b_{00k}c_{ijk} + a_{ijk}b_{00k}c_{ij0}$$
$$+ a_{0j0}b_{ijk}c_{i0k} + a_{0jk}b_{ijk}c_{i00} + a_{0j0}b_{i0k}c_{ijk} + a_{0jk}b_{i0k}c_{ij0})$$

We may think of $\tilde{T}$ as a sum of eight terms, each of which is a $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$ with $\mathbf{lmn} = q^3$, e.g., the first is $\sum_{i,j,k=1}^{q} a_{ij0}b_{0jk}c_{i0k} = M_{\langle q,q,q\rangle}$, the second $M_{\langle q^2,q,1\rangle}$ etc.. (I will say of *volume* q^3.) Were they all disjoint expressions, we could use the asymptotic sum inequality to conclude $8q^\omega \leq (q+1)^3$ and for small q we would see $\omega < 2$. Of course this is not the case, but we can try to zero out some of the variables to keep as many of these eight terms as possible. For example if we set $c_{i00}, b_{00k}, b_{ijk}, c_{ijk}$ all to zero, we are left with two disjoint matrix multiplications and we conclude $2q^\omega \leq (q+1)^3$. This is best when $q = 15$, giving $\omega < 2.816$, which is not so interesting.

At this point enters a new idea: since we are dealing with border rank, we have greater flexibility in degeneration than simply zero-ing out terms. By taking limits, we will be able to keep three terms! To explain this, I need to take another detour regarding orbit closures.

3.4.3. All tensors are degenerations of matrix multiplication.

Theorem 3.4.3.1 (Strassen [**Str87**]). *Set $r = \lfloor \frac{3}{4}\mathbf{n}^2 \rfloor$ and choose a linear embedding $\mathbb{C}^r \subset \mathbb{C}^{\mathbf{n}^2}$. Then*

$$\sigma_r(Seg(\mathbb{P}^{r-1}\times\mathbb{P}^{r-1}\times\mathbb{P}^{r-1})) \subset \overline{GL_{\mathbf{n}^2}\times GL_{\mathbf{n}^2}\times GL_{\mathbf{n}^2}\cdot[M_{\langle \mathbf{n}\rangle}]},$$

i.e.,

$$\overline{GL_r\times GL_r\times GL_r\cdot[M_{\langle 1\rangle}^{\oplus r}]} \subset \overline{GL_{\mathbf{n}^2}\times GL_{\mathbf{n}^2}\times GL_{\mathbf{n}^2}\cdot[M_{\langle \mathbf{n}\rangle}]}.$$

Proof. The proof will be by a very simple degeneration: let $T_A \subset GL(A) = GL_{\mathbf{n}^2}$ denote the diagonal $\mathbf{n}^2\times\mathbf{n}^2$ matrices. I will show

$$M_{\langle 1\rangle}^{\oplus r} \subset \overline{T_A\times T_B\times T_C\cdot M_{\langle \mathbf{n}\rangle}}.$$

Write x_{ij} for a basis of A etc. so $M_{\langle \mathbf{n}\rangle} = \sum_{i,j,k} x_{ij}\otimes y_{jk}\otimes z_{ki}$. We want to kill off as few terms as possible such that in the remaining terms, each basis vector appears in at most one monomial. That is if we have x_{ij} appearing, then there should be a unique $k_0 = k(i,j)$, such that the only term surviving in $\sum_k x_{ij}{\otimes}y_{jk}{\otimes}z_{ki}$ is $x_{ij}{\otimes}y_{jk_0}{\otimes}z_{k_0i}$. We should view this more symmetrically, fixing some integer h and requiring that the only terms appearing are of the form $x_{ij}{\otimes}y_{jk}{\otimes}z_{ki}$ where $i+j+k=h$. To do this, we want curves

$$\begin{aligned} x_{ij} &\mapsto t^{\alpha(i,j)}x_{ij} \\ y_{jk} &\mapsto t^{\beta(j,k)}y_{jk} \\ z_{ki} &\mapsto t^{\gamma(k,i)}z_{ki} \end{aligned}$$

so that $\alpha+\beta+\gamma=0$ when $i+j+k=h$ and $\alpha+\beta+\gamma>0$ when $i+j+k\neq h$, as then

$$\lim_{t\to 0}\sum_{i,j,k=1}^{\mathbf{n}} t^{\alpha(i,j)+\beta(j,k)+\gamma(k,i)}x_{ij}{\otimes}y_{jk}{\otimes}z_{ki} = \sum_{i+j+k=h} x_{ij}{\otimes}y_{jk}{\otimes}z_{ki}.$$

Set $\lambda=i+j+k$. We want something like

$$\alpha+\beta+\gamma=(h-\lambda)^2=h^2-2\lambda h+\lambda^2.$$

Take

$$\begin{aligned} \alpha &= \frac{1}{2}(i^2+j^2)+2ij+(\frac{h}{3}-i-j)h \\ \beta &= \frac{1}{2}(k^2+j^2)+2kj+(\frac{h}{3}-k-j)h \\ \gamma &= \frac{1}{2}(i^2+k^2)+2ik+(\frac{h}{3}-i-k)h. \end{aligned}$$

Exercise 3.4.3.2: (1) Verify that $\alpha+\beta+\gamma=(h-\lambda)^2$.

Exercise 3.4.3.3: (2) Show that the best value of h is $h=\lceil\frac{3\mathbf{n}}{2}\rceil+1$ which yields $r=\lfloor\frac{3}{4}\mathbf{n}^2\rfloor$ to finish the proof.

$\square$

Remark 3.4.3.4. Note that we really are doing a degeneration argument here, in the sense that there are values of i,j,k where one of α,β,γ is negative. To avoid negative terms for the curves in A,B,C, we could add r to each of α,β,γ and then divide the entire entire expression by t^{3r}.

I will call degenerations that only use the diagonal matrices *toric degenerations.*

Corollary 3.4.3.5. *Every tensor in* $\mathbb{C}^{\frac{3}{2}\mathbf{n}}{\otimes}\mathbb{C}^{\frac{3}{2}\mathbf{n}}{\otimes}\mathbb{C}^{\frac{3}{2}\mathbf{n}}$ *arises as a degeneration of* $M_{\langle\mathbf{n}\rangle}$.

Proof. As mentioned in §2.1.6, the maximum border rank of any tensor in $\mathbb{C}^{\frac{3}{2}\mathbf{n}}{\otimes}\mathbb{C}^{\frac{3}{2}\mathbf{n}}{\otimes}\mathbb{C}^{\frac{3}{2}\mathbf{n}}$ is at most $\frac{3}{4}\mathbf{n}^2$, and any tensor of border rank r is a degeneration of a general element of $\sigma_r(Seg(\mathbb{P}^{r-1}\times\mathbb{P}^{r-1}\times\mathbb{P}^{r-1}))$. $\square$

Remark 3.4.3.6. Theorem 3.4.3.1 may be interpreted as saying that one can degenerate $M_{\langle \mathbf{n}\rangle}$ to a tensor that computes $\lfloor \frac{3}{4}\mathbf{n}^2 \rfloor$ independent scalar multiplications. If we have any tensor realized as $M_{\langle \mathbf{n}\rangle}\otimes T$, the same degeneration procedure works to degenerate it to $M_{\langle 1\rangle}^{\oplus r}\otimes T$.

3.4.4. A better bound using the toric degeneration. Now we return to the expression (3.4.2). There are four kinds of A-indices, $ij0$, ijk, $0j0$ and $0jk$. To emphasize this, and to suggest what kind of degeneration to perform, label these with superscripts [11], [21], [12] and [22]. Label each of the B and C indices (which come in four types as well) similarly. We obtain:

$$\tilde{T} = \sum_{i,j,k=1}^{q} (a_{ij0}^{[11]} b_{0jk}^{[11]} c_{i0k}^{[11]} + a_{ijk}^{[21]} b_{0jk}^{[11]} c_{i00}^{[12]} + a_{ij0}^{[11]} b_{00k}^{[12]} c_{ijk}^{[21]} + a_{ijk}^{[21]} b_{00k}^{[12]} c_{ij0}^{[22]}$$
$$+ a_{0j0}^{[12]} b_{ijk}^{[21]} c_{i0k}^{[11]} + a_{0jk}^{[22]} b_{ijk}^{[21]} c_{i00}^{[12]} + a_{0j0}^{[12]} b_{i0k}^{[22]} c_{ijk}^{[21]} + a_{0jk}^{[22]} b_{i0k}^{[22]} c_{ij0}^{[22]}).$$

This expression has the structure of block 2×2 matrix multiplication. Think of it as a sum of q^3 2×2 matrix multiplications. Now use Theorem 3.4.3.1 to degenerate each 2×2 matrix multiplication to a sum of 3 disjoint terms. Namely, following the recipe that the three indices must add to 4, we keep all terms $a^{[s,t]}b^{[t,u]}c^{[u,s]}$ where $s + t + u = 4$, namely we degenerate $\tilde{T}$ to

$$\sum_{i,j,k=1}^{q} a_{ijk}^{[21]} b_{0jk}^{[11]} c_{i00}^{[12]} + a_{ij0}^{[11]} b_{00k}^{[12]} c_{ijk}^{[21]} + a_{0j0}^{[12]} b_{ijk}^{[21]} c_{i0k}^{[11]}$$

and apply the asymptotic sum inequality. We obtain $3q^\omega \le (q + 1)^3$ which gives the best bound on ω when $q = 7$, namely $\omega < 2.642$, which is still not as good as Schönhage's bound.

3.4.5. Strassen's bound. We do better by using the standard trick of this chapter: taking a high tensor power of $\tilde{T}$, as $\tilde{T}^{\otimes N}$ contains $(2^N)^2$ matrix multiplications $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$, all with $\mathbf{lmn} = q^{3N}$, and again by Theorem 3.4.3.1 we may keep $\frac{3}{4}2^{2N}$ of them. The asymptotic sum inequality applied to the degenerated tensor gives

$$\frac{3}{4}2^{2N} q^{N\omega} \le (q + 1)^{3N}.$$

Taking N-th roots and letting N tend to infinity, the $\frac{3}{4}$ goes away and we obtain

$$2^2 q^\omega \le (q + 1)^3.$$

Finally, the case $q = 5$ implies:

Theorem 3.4.5.1. **[Str87]** $\omega < 2.48$.

3.4.6. Asymptotic rank. The above discussion suggests the introduction of yet another complexity measure for tensors: given $T \in A \otimes B \otimes C$, we can consider $T^{\otimes N} \in A^{\otimes N} \otimes B^{\otimes N} \otimes C^{\otimes N}$ and this construction played a central role in Strassen's laser method to prove upper bounds for the complexity of matrix multiplication via auxiliary tensors.

Definition 3.4.6.1. The *asymptotic rank* $\tilde{\mathbf{R}}(T)$ of a tensor $T \in A \otimes B \otimes C$, is

$$\tilde{\mathbf{R}}(T) := \inf_N [\mathbf{R}(T^{\otimes N})]^{\frac{1}{N}}.$$

Exercise 3.4.6.2: (1) Show that in the definition, one can replace the infimum by $\lim_{N\to\infty}$ by using Lemma 3.4.7.2 below.

Exercise 3.4.6.3: (2) Show that $\tilde{\mathbf{R}}(T) \leq \underline{\mathbf{R}}(T)$. ⊚

Since $M_{\langle 2 \rangle}^{\otimes k} = M_{\langle 2^k \rangle}$, we have $\tilde{\mathbf{R}}(M_{\langle 2 \rangle}) = 2^{\omega}$.

Conjecture 3.4.6.4. [**Str91**] *Let $T \in A \otimes B \otimes C$ be concise with $\mathbf{a} = \mathbf{b} = \mathbf{c}$. Then $\tilde{\mathbf{R}}(T) = \mathbf{a}$.*

Note that, taking $T = M_{\langle 2 \rangle}$, this would imply $\omega = 2$.

3.4.7. Degeneracy value. I now formalize what we did to get Strassen's bound. The starting point is if a tensor T degenerates to $\bigoplus_{i=1}^{s} M_{\langle \mathbf{l}_i, \mathbf{m}_i, \mathbf{n}_i \rangle}$, then $\sum_{i=1}^{s} (\mathbf{l}_i \mathbf{m}_i \mathbf{n}_i)^{\frac{\omega}{3}} \leq \underline{\mathbf{R}}(T)$, and more generally we worked with degenerations of $T^{\otimes N}$ as well. Informally define the *degeneracy value* of T to be the best upper bound on ω we can get in this manner. More precisely:

Definition 3.4.7.1. Let $T \in A \otimes B \otimes C$. Fix $N \geq 1$ and $\rho \in [2, 3]$. Define $V_{\rho,N}^{degen}(T)$ to be the maximum of $\sum_{i=1}^{s} (\mathbf{l}_i \mathbf{m}_i \mathbf{n}_i)^{\frac{\rho}{3}}$ over all degenerations of $T^{\otimes N}$ to $\oplus_{i=1}^{s} M_{\langle \mathbf{l}_i, \mathbf{m}_i, \mathbf{n}_i \rangle}$ over all choices of $s, \mathbf{l}_i, \mathbf{m}_i, \mathbf{n}_i$ and define the *degeneracy value* of T to be $V_{\rho}^{degen}(T) := \sup_N V_{\rho,N}^{degen}(T)^{\frac{1}{N}}$.

The asymptotic sum inequality implies $V_{\omega}^{degen}(T) \leq \underline{\mathbf{R}}(T)$, or in other words, if $V_{\rho}^{degen}(T) \geq \underline{\mathbf{R}}(T)$, then $\omega \leq \rho$.

The supremum in the definition can be replaced by a limit, thanks to *Fekete's lemma*, since the sequence $\log(V_{\rho,N}^{degen}(T))$ is super-additive:

Lemma 3.4.7.2 (Fekete's Lemma)**.** *For every super-additive sequence $\{a_n\}_{n=1}^{\infty}$ (i.e. $a_{n+m} \geq a_n + a_m$), the limit $\lim_{n\to\infty} \frac{a_n}{n}$ exists (possibly $+\infty$) and is equal to $\sup \frac{a_n}{n}$.*

Exercise 3.4.7.3: (3) Prove Fekete's Lemma.

Fekete's lemma implies $\frac{1}{N} \log V_{\rho,N}^{degen}(T)$ tends to a limit. See [**?**] for details.

There is also an analogue of the asymptotic sum inequality for degeneracy value:

Theorem 3.4.7.4. $\sum_{i=1}^{s} V_\omega^{degen}(T_i) \leq \underline{\mathbf{R}}(\oplus_{i=1}^{s} T_i)$.

The proof is similar to the proof of the asymptotic sum inequality. It is clear that $V_\omega^{degen}(T_1 \otimes T_2) \geq V_\omega^{degen}(T_1) \otimes V_\omega^{degen}(T_2)$. To show $V_\omega^{degen}(T_1 \oplus T_2) \geq V_\omega^{degen}(T_1) + V_\omega^{degen}(T_2)$ one expands out $V_{\omega,N}^{degen}(T_1 \oplus T_2)$, the result is a sum of products with coefficients, but as with the asymptotic sum inequality, one can essentially just look at the largest term, and as N tends to infinity, the coefficient becomes irrelevant after taking N-th roots.

Thus tensors of low border rank with high degeneracy value give upper bounds on ω. The problem is that we have no systematic way of estimating degeneracy value. For an extreme example, for a given r the tensor of border rank r with the highest degeneracy value is $M_{\langle 1 \rangle}^{\oplus r}$ as all border rank r tensors are degenerations of it.

In subsequent work, researchers restrict to a special type of value that is possible to estimate.

3.4.8. The value of a tensor. Let $\mathrm{End}(A) \times \mathrm{End}(B) \times \mathrm{End}(C)$ act on $A \otimes B \otimes C$ by the action inherited from the $GL(A) \times GL(B) \times GL(C)$ action (not the Lie algebra action). Then for all $X \in \mathrm{End}(A) \times \mathrm{End}(B) \times \mathrm{End}(C)$ and $T \in A \otimes B \otimes C$, we have $\mathbf{R}(X \cdot T) \leq \mathbf{R}(T)$ and $\underline{\mathbf{R}}(X \cdot T) \leq \underline{\mathbf{R}}(T)$ by Exercise 2.1.6.2.

Definition 3.4.8.1. One says T *restricts to* T' if $T' \in \mathrm{End}(A) \times \mathrm{End}(B) \times \mathrm{End}(C) \cdot T$.

Definition 3.4.8.2. For $T \in A \otimes B \otimes C$, $N \geq 1$ and $\rho \in [2,3]$ define $V_{\rho,N}^{restr}(T)$ to be the maximum of $\sum_{i=1}^{s} (\mathbf{l}_i \mathbf{m}_i \mathbf{n}_i)^{\frac{\rho}{3}}$ over all restrictions of $T^{\otimes N}$ to $\oplus_{i=1}^{s} M_{\langle \mathbf{l}_i, \mathbf{m}_i, \mathbf{n}_i \rangle}$ and define the *restriction value* of T to be $V_\rho^{restr}(T) := \sup_N V_{\rho,N}^{restr}(T)^{\frac{1}{N}}$.

I emphasize that the degeneration used by Strassen is more general than restriction.

Coppersmith-Winograd and all subsequent work, use only the following type of restriction:

Definition 3.4.8.3. Let A, B, C be given bases, so write them as $\mathbb{C}^{\mathbf{a}}, \mathbb{C}^{\mathbf{b}}, \mathbb{C}^{\mathbf{c}}$. We say $T \in \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{b}} \otimes \mathbb{C}^{\mathbf{c}}$ *combinatorially restricts to* T' if T restricts to T' by setting some of the coordinates of T to zero.

The condition that $T \in \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{b}} \otimes \mathbb{C}^{\mathbf{c}}$ admits a combinatorial restriction to the matrix multiplication tensor $M_{\langle \mathbf{l}, \mathbf{m}, \mathbf{n} \rangle}$ may be phrased as follows (following [**CU03**]): write a_α, b_β, c_γ for the given bases of A, B, C and write

$T = \sum_{\alpha=1}^{\mathbf{a}} \sum_{\beta=1}^{\mathbf{b}} \sum_{\gamma=1}^{\mathbf{c}} t^{\alpha,\beta,\gamma} a_\alpha \otimes b_\beta \otimes c_\gamma$. Then $T \in \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{b}} \otimes \mathbb{C}^{\mathbf{c}}$ combinatorially restricts to $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$ means that there exist injections

$$\begin{aligned} \alpha &: [\mathbf{l}] \times [\mathbf{m}] \to [\mathbf{a}] \\ \beta &: [\mathbf{m}] \times [\mathbf{n}] \to [\mathbf{b}] \\ \gamma &: [\mathbf{n}] \times [\mathbf{l}] \to [\mathbf{c}] \end{aligned}$$

such that

$$t^{\alpha(i,j'),\beta(j,k'),\gamma(k,i')} = \begin{Bmatrix} 1 & \text{if } i = i',\ j = j',\ k = k' \\ 0 & \text{otherwise} \end{Bmatrix}. \tag{3.4.3}$$

One can similarly phrase combinatorial restriction to a sum of disjoint matrix multiplication tensors.

Definition 3.4.8.4. For $T \in \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{b}} \otimes \mathbb{C}^{\mathbf{c}}$, $N \geq 1$ and $\rho \in [2,3]$ define $V_{\rho,N}(T)$ to be the maximum of $\sum_{i=1}^{s} (\mathbf{l}_i \mathbf{m}_i \mathbf{n}_i)^{\frac{\rho}{3}}$ over all combinatorial restrictions of $T^{\otimes N}$ to $\oplus_{i=1}^{s} M_{\langle \mathbf{l}_i,\mathbf{m}_i,\mathbf{n}_i\rangle}$ and define the *combinatorial value* (or *value* for short, since it is the value used in the literature) of T to be $V_\rho(T) := \lim_{N\to\infty} V_{\rho,N}(T)^{\frac{1}{N}}$. (The limit is shown to exist in [**DS13**].)

Note that the values satisfy $V_\rho^{degen} \geq V_\rho^{restr} \geq V_\rho$. As with all the values we have

- $V_\rho(T)$ is a non-decreasing function of ρ,
- $V_\omega(T) \leq \underline{\mathbf{R}}(T)$.

Thus if $V_\rho(T) \geq \underline{\mathbf{R}}(T)$, then $\omega \leq \rho$.

Combinatorial value can be estimated in principle, as for each N, there are only a finite number of combinatorial restrictions. In practice, the tensor is presented in such a way that there are "obvious" combinatorial degenerations to disjoint matrix multiplication tensors and at first, one optimizes just among these obvious combinatorial degenerations. However, it may be that there are matrix multiplication tensors of the form $\sum_j a_0 \otimes b_j \otimes c_j$ as well as tensors of the form $a_0 \otimes b_k \otimes c_k$ where k is not in the range of j. Then one can *merge* these tensors to $a_0 \otimes (\sum_j b_j \otimes c_j + b_k \otimes c_k)$ to increase value because although formally speaking they were not disjoint, they do not interfere with each other. (The value increases as e.g., $q^\omega + r^\omega < (q+r)^\omega$.) So the actual procedure is to optimize among combinatorial restrictions with merged tensors.

3.4.9. The Coppersmith-Winograd tensors.

Coppersmith and Winograd apply Strassen's laser method, enhanced with merging, but only using combinatorial restrictions to the following two tensors:

The "easy Coppersmith-Winograd tensor":

$$T_{cw} := \sum_{j=1}^{q} a_0\otimes b_j\otimes c_j + a_j\otimes b_0\otimes c_j + a_j\otimes b_j\otimes c_0 \in \mathbb{C}^{q+1}\otimes\mathbb{C}^{q+1}\otimes\mathbb{C}^{q+1} \tag{3.4.4}$$

This tensor has low border rank. To see why, consider the second derivatives of a curve in the Segre: Let $x(t) = a(t)\otimes b(t)\otimes c(t)$, write x' for $x'(0)$ and similarly for all derivatives. Then

$$x'' = (a''\otimes b\otimes c + a\otimes b''\otimes c + a\otimes b\otimes c'') + 2(a'\otimes b'\otimes c + a'\otimes b\otimes c' + a\otimes b'\otimes c')$$

so if we begin with the base point $a_0\otimes b_0\otimes c_0$, each term in the summand for T_{cw} is a term of the second kind. The terms in the first parenthesis are ordinary tangent vectors. Thus take q curves beginning at $a_0\otimes b_0\otimes c_0$, we can cancel out all the terms of the first type with a single vector to obtain the resulting border rank $q+2$ expression:

$$\begin{aligned} T_{cw} = \lim_{t\to 0}\frac{1}{t^2}&\left[\sum_{j=1}^{q}(a_0+ta_j)\otimes(b_0+tb_j)\otimes(c_0+tc_j)\right] \\ &- (a_0+t\sum_j a_j)\otimes(b_0+t\sum_j b_j)\otimes(c_0+t\sum_j c_j) - (q-1)a_0\otimes b_0\otimes c_0. \end{aligned}$$

Exercise 3.4.9.1: (2) Show that $\underline{\mathbf{R}}(T_{cw}) \geq q+2$ so that equality holds.

A slightly more complicated tensor yields even better results: Let

(3.4.5)
$$\begin{aligned} T_{CW} := &\sum_{j=1}^{q}(a_0\otimes b_j\otimes c_j + a_j\otimes b_0\otimes c_j + a_j\otimes b_j\otimes c_0) \\ &+ a_0\otimes b_0\otimes c_{q+1} + a_0\otimes b_{q+1}\otimes c_0 + a_{q+1}\otimes b_0\otimes c_0 \in \mathbb{C}^{q+2}\otimes\mathbb{C}^{q+2}\otimes\mathbb{C}^{q+2} \end{aligned}$$

and call T_{CW} the *Coppersmith-Winograd tensor*

Exercise 3.4.9.2: (2) Show the Coppersmith-Winograd tensor also has border rank $q+2$ by modifying the curves used to obtain T_{cw}. ⊚

Now we suggestively re-label T_{CW} as we did with Strassen's tensor:

(3.4.6)
$$\begin{aligned} T_{CW} := &\sum_{j=1}^{q}(a_0^{[0]}\otimes b_j^{[1]}\otimes c_j^{[1]} + a_j^{[1]}\otimes b_0^{[0]}\otimes c_j^{[1]} + a_j^{[1]}\otimes b_j^{[1]}\otimes c_0^{[0]}) \\ &+ a_0^{[0]}\otimes b_0^{[0]}\otimes c_{q+1}^{[2]} + a_0^{[0]}\otimes b_{q+1}^{[2]}\otimes c_0^{[0]} + a_{q+1}^{[2]}\otimes b_0^{[0]}\otimes c_0^{[0]} \in \mathbb{C}^{q+2}\otimes\mathbb{C}^{q+2}\otimes\mathbb{C}^{q+2} \end{aligned}$$

to see that T_{CW} is the sum of 3 matrix multiplications of volume q^2, and 3 of volume 1, all non-disjoint. To get more interesting matrix multiplications, consider $T_{CW}^{\otimes 2}$, but this time, instead of double superscripts, simply add the superscripts.

$$
\begin{aligned}
T_{CW}^{\otimes 2} = \sum_{i,j=1}^{q} [& a_{00}^{[0]}\otimes b_{ij}^{[2]}\otimes c_{ij}^{[2]} + a_{0j}^{[1]}\otimes b_{i0}^{[1]}\otimes c_{ij}^{[2]} + a_{0j}^{[1]}\otimes b_{ij}^{[2]}\otimes c_{i0}^{[1]} + a_{i0}^{[1]}\otimes b_{0j}^{[1]}\otimes c_{ij}^{[2]} + a_{i0}^{[1]}\otimes b_{ij}^{[2]}\otimes c_{0j}^{[1]} \\
& + a_{ij}^{[2]}\otimes b_{i0}^{[1]}\otimes c_{0j}^{[1]} + a_{ij}^{[2]}\otimes b_{00}^{[0]}\otimes c_{ij}^{[2]} + a_{ij}^{[2]}\otimes b_{ij}^{[2]}\otimes c_{00}^{[1]} + a_{ij}^{[2]}\otimes b_{0j}^{[1]}\otimes c_{i0}^{[1]}] \\
+ \sum_{j=1}^{q} [& a_{0,q+1}^{[2]}\otimes b_{j0}^{[1]}\otimes c_{j0}^{[1]} + a_{q+1,0}^{[2]}\otimes b_{0j}^{[1]}\otimes c_{0j}^{[1]} + a_{q+1,j}^{[3]}\otimes b_{0j}^{[1]}\otimes c_{00}^{[0]} + a_{j,q+1}^{[3]}\otimes b_{j0}^{[1]}\otimes c_{00}^{[0]} \\
& + a_{q+1,j}^{[3]}\otimes b_{00}^{[0]}\otimes c_{0j}^{[1]} + a_{j,q+1}^{[3]}\otimes b_{00}^{[0]}\otimes c_{j0}^{[1]}] \\
+ & a_{q+1,q+1}^{[4]}\otimes b_{00}^{[0]}\otimes c_{00}^{[0]} + a_{00}^{[0]}\otimes b_{q+1,j}^{[3]}\otimes c_{0j}^{[1]} + a_{00}^{[0]}\otimes b_{0j}^{[1]}\otimes c_{q+1,j}^{[3]} \\
+ & a_{00}^{[0]}\otimes b_{q+1,q+1}^{[4]}\otimes c_{00}^{[0]} + a_{00}^{[0]}\otimes b_{00}^{[0]}\otimes c_{q+1,q+1}^{[4]}.
\end{aligned}
$$

Now we have non-disjoint matrix multiplications of volumes q^2, q and 1. Thus when we zero-out terms to get disjoint matrix multiplications in $(T_{CW}^{\otimes 2})^{\otimes N}$, in order to optimize value, we need to weight the q^2 terms more than the q terms etc.

As mentioned above, we can obtain better upper bounds with merging. One needs to make a choice how to merge. Coppersmith and Winogrand group the $\mathbb{C}^{\mathbf{a}^2}$-variables

$$
\begin{aligned}
\mathcal{A}^{[0]} &= \{a_{00}^{[0]}\} \\
\mathcal{A}^{[1]} &= \{a_{i0}^{[1]}, a_{0j}^{[1]}\} \\
\mathcal{A}^{[2]} &= \{a_{q+1,0}^{[2]}, a_{ij}^{[2]}, a_{0,q+1}^{[2]}\} \\
\mathcal{A}^{[3]} &= \{a_{q+1,j}^{[3]}, a_{i,q+1}^{[3]}\} \\
\mathcal{A}^{[4]} &= \{a_{q+1,q+1}^{[4]}\}
\end{aligned}
$$

and similarly for b's and c's. Then

$$
T_{CW}^{\otimes 2} = \sum_{I+J+K=4} \sum_{a\in I, b\in J, c\in K} \mathcal{A}^{[a]}\otimes\mathcal{B}^{[b]}\otimes\mathcal{C}^{[c]}.
$$

Most of these terms are just matrix multiplications, however terms with $1+1+2$ are not:

$$\mathcal{A}^{[1]}\otimes\mathcal{B}^{[1]}\otimes\mathcal{C}^{[2]} = \sum_{i=1}^{q} a^{[1]}_{i0}\otimes b^{[1]}_{i0}\otimes c^{[2]}_{0,q+1} + \sum_{j=1}^{q} a^{[1]}_{0j}\otimes b^{[1]}_{0j}\otimes c^{[2]}_{q+1,0} + \sum_{i,j=1}^{q}[a^{[1]}_{i0}\otimes b^{[1]}_{0j}\otimes c^{[2]}_{ij} + a^{[1]}_{0j}\otimes b^{[1]}_{i0}\otimes c^{[2]}_{ij}].$$

To this term we estimate value using the laser method, i.e., we degenerate tensor powers of $\mathcal{A}^{[1]}\otimes\mathcal{B}^{[1]}\otimes\mathcal{C}^{[2]}$ to disjoint matrix multiplication tensors. Coppersmith and Winograd show that has value at least $2^{\frac{2}{3}}q^{\omega}(q^{3\omega}+2)^{\frac{1}{3}}$.

Now there is an optimization problem to solve, that I briefly discuss below.

Coppersmith and Winograd get their best result of $\omega < 2.3755$ by merging $T_{CW}^{\otimes 2}$ and then optimizing over the various combinatorial restrictions. In subsequent work Stothers [**Sto**], resp. Williams [**Wil**], resp. LeGall [**Gal**] used merging with $T_{CW}^{\otimes 4}$ resp. $T_{CW}^{\otimes 8}$, resp. $T_{CW}^{\otimes 16}$ and $T_{CW}^{\otimes 32}$ leading to the current "world record":

Theorem 3.4.9.3. [**Gal**] $\omega < 2.3728639$.

Ambainis, Filmus and LeGall [**?**] showed that taking higher powers of T_{CW} when $q \geq 5$ cannot be used to prove $\omega < 2.30$ by this method alone. Their argument avoids higher powers by more sophisticated methods to account for when potential merging in higher tensor powers can occur.

Thus one either needs to develop new methods, or find better base tensors.

I discuss the search for better base tensors in §**??**.

3.4.10. How one optimizes in practice. To get an idea of how the optimization procedure works, start with some base tensor T that contains a collection of matrix multiplication tensors $M_{\langle \mathbf{l}_i,\mathbf{m}_i,\mathbf{n}_i\rangle}$, $1 \leq i \leq x$ that are not disjoint. Then $T^{\otimes N}$ will contain matrix multiplication tensors of the form $M_{\langle \mathbf{l}_\mu,\mathbf{m}_\mu,\mathbf{n}_\mu\rangle}$ where $\mathbf{l}_\mu = \mathbf{l}_{\mu_1}\cdots\mathbf{l}_{\mu_N}$ and similarly for $\mathbf{m}_\mu,\mathbf{n}_\mu$, where $\mu_j \in [x]$.

Each matrix multiplication tensor will occur with a certain multiplicity and certain variables. The problem becomes to zero out variables in a way that maximizes the value of what remains. More precisely, for large N, one wants to maximize the sum $\sum_j K_j(\mathbf{l}_{\mu_j}\mathbf{m}_{\mu_j}\mathbf{n}_{\mu_j})^{\frac{\rho}{3}}$ where the surviving matrix multiplication tensors are $M^{\oplus K_j}_{\langle \mathbf{l}_{\mu_j}\mathbf{m}_{\mu_j}\mathbf{n}_{\mu_j}\rangle}$ and disjoint. One then takes the smallest ρ such that $\sum_j K_j(\mathbf{l}_{\mu_j}\mathbf{m}_{\mu_j}\mathbf{n}_{\mu_j})^{\frac{\rho}{3}} \geq \underline{\mathbf{R}}(T)$ and concludes $\omega \leq \rho$. One ingredient is the Salem-Spencer Theorem:

Theorem 3.4.10.1 (Salem and Spencer [**SS42**]). *Given $\epsilon > 0$, there exists $M_\epsilon \simeq 2^{\frac{c}{\epsilon^2}}$ such that for all $M > M_\epsilon$, there is a set B of $M' > M^{1-\epsilon}$ distinct integers $0 < b_1 < b_2 < \cdots < b_{M'} < \frac{M}{2}$ with no three terms in an arithmetic progression, i.e., for $b_i, b_j, b_k \in B$, $b_i + b_j = 2b_k$ if and only if $b_i = b_j = b_k$. In fact no three terms form an arithmetic progression* mod M.

This theorem assures one can get away with only zero-ing out a relatively small number of terms, so in some sense it plays the role of Strassen's degeneration theorem. I state explicitly to emphasize that it is an existence result, not an algorithm. In the general case one assigns probability distributions and optimizes using techniques from probability to determine what percentage of each type gets zero-ed out. I suggest [**CW82**] for the basic idea and [**?**] for the state of the art regarding this optimization.

3.5. The Cohn-Umans program

A conceptually appealing approach to proving upper bounds on ω was initiated by H. Cohn and C. Umans.

Imagine a tensor that comes presented in two different bases. In one, the *cost* of the tensor is clear: it may be written as a sum of small disjoint matrix multiplication tensors. On the other hand, in the other its *value* (in the sense discussed above) is high, because it may be seen to degenerate to good matrix multiplication tensors. Such a situation does arise in practice! It occurs for *structure tensors* for the group algebra of a finite group, as defined below. In one (the "matrix coefficient basis"), one gets an upper bound on the rank of the tensor, and in the other (the "standard basis") there are many potential combinatorial degenerations and one gets a lower bound on the value.

I state the needed representation theory now, and defer proofs of the statements to §8.6. I then present their method.

3.5.1. Structure tensor of an algebra.

Let $\mathcal{A}$ be a finite dimensional *algebra*, i.e. a vector space with a multiplication operation, with basis $a_1, \ldots, a_{\mathbf{a}}$ and dual basis $\alpha^1, \ldots, \alpha^{\mathbf{a}}$. Write $a_i a_j = \sum A^k_{ij} a_k$ for the multiplication in $\mathcal{A}$, where the A^k_{ij} are constants. The multiplication $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$ is bilinear and one defines the corresponding *structure tensor of* $\mathcal{A}$

$$M_{\mathcal{A}} := \sum_{i,j,k} A^k_{ij} \alpha^i \otimes \alpha^j \otimes a_k \in \mathcal{A}^* \otimes \mathcal{A}^* \otimes \mathcal{A}. \tag{3.5.1}$$

For example, $M_{\langle \mathbf{n} \rangle}$ is the structure tensor for the algebra of $\mathbf{n} \times \mathbf{n}$-matrices with operation matrix multiplication.

The group algebra of a finite group. Let G be a finite group and let $\mathbb{C}[G]$ denote the vector space of complex-valued functions on G, called the *group algebra* of G. The following exercise justifies the name:

Exercise 3.5.1.1: (1) Show that if the elements of G are $g_1, \dots, g_r$, then $\mathbb{C}[G]$ has a basis indexed $\delta_{g_1}, \dots, \delta_{g_r}$, where $\delta_{g_i}(g_j) = \delta_{ij}$. Show that $\mathbb{C}[G]$ may be given the structure of an algebra by defining $\delta_{g_i}\delta_{g_j} := \delta_{g_ig_j}$ and extending linearly.

Thus if G is a finite group, then $M_{\mathbb{C}[G]} = \sum_{g,h\in G} \delta_g^*\otimes\delta_h^*\otimes\delta_{gh}$.

Example 3.5.1.2.

$$M_{\mathbb{C}[\mathbb{Z}_m]} = \sum_{0\le i,j<m} \delta_i^*\otimes\delta_j^*\otimes\delta_{i+j \bmod m}.$$

Notice that, introducing coordinates $x_0, \dots, x_{m-1}$ on $\mathbb{C}[\mathbb{Z}_m]$, one obtains a circulant matrix for $M_{\mathbb{C}[\mathbb{Z}_m]}(\mathbb{C}[\mathbb{Z}_m]^*) \subset \mathbb{C}[\mathbb{Z}_m]^*\otimes\mathbb{C}[\mathbb{Z}_m]^*$:

$$M_{\mathbb{C}[\mathbb{Z}_m]}(\mathbb{C}[\mathbb{Z}_m]^*) = \begin{pmatrix} x_0 & x_1 & \cdots & x_{m-1} \\ x_{m-1} & x_0 & x_1 & \cdots \\ \vdots & & \ddots & \\ x_1 & x_2 & \cdots & x_0 \end{pmatrix}.$$

Note that all entries of the matrix are non-zero and filled with basis vectors. This holds in general for the presentation of $\mathbb{C}[G]$ in the standard basis, which makes it useful for combinatorial restrictions.

What are $\underline{\mathbf{R}}(M_{\mathbb{C}[\mathbb{Z}_m]})$ and $\mathbf{R}(M_{\mathbb{C}[\mathbb{Z}_m]})$? The space of circulant matrices forms an abelian subspace, which indicates the rank and border rank might be minimal or nearly minimal among concise tensors. We will determine the rank and border rank of $M_{\mathbb{C}[\mathbb{Z}_m]}$ momentarily via the discrete Fourier transform.

3.5.2. The structure theorem of $\mathbb{C}[G]$.

I give a proof of the following theorem in §8.6.5:

Theorem 3.5.2.1. *Let G be a finite group, then as an algebra,*

$$\mathbb{C}[G] = \bigoplus_i V_i^*\otimes V_i \tag{3.5.2}$$

where the sum is over all the distinct irreducible representations of G. In particular, if $\dim V_i = d_i$, *then*

$$\mathbb{C}[G] \simeq \bigoplus_i Mat_{d_i\times d_i}(\mathbb{C}).$$

3.5.3. The (generalized) discrete Fourier transform. We have two natural expressions for $M_{\mathbb{C}[G]}$, the original presentation in terms of the algebra multiplication in terms of delta functions, the *standard basis*, and the *matrix coefficient* basis in terms of Theorem 3.5.2.1. The change of basis matrix from the standard basis to the matrix coefficient basis is called the (generalized) *Discrete Fourier Transform* (DFT).

Example 3.5.3.1. The classical DFT is the case $G = \mathbb{Z}_m$. The irreducible representations of $\mathbb{Z}_m$ are all one dimensional: $\rho_k : \mathbb{Z}_m \to GL_1$. Let $\sigma \in \mathbb{Z}_m$ be a generator, then $\rho_k(\sigma)v = e^{\frac{2\pi ik}{m}}v$ for $0 \leq k \leq m$. The DFT matrix is

$$(e^{\frac{2\pi i(j+k)}{m}})_{0\leq j,k\leq m-1}.$$

Proposition 3.5.3.2. $\underline{\mathbf{R}}(M_{\mathbb{C}[\mathbb{Z}_m]}) = \mathbf{R}(M_{\mathbb{C}[\mathbb{Z}_m]}) = m$.

Proof. Theorem 3.5.2.1 implies $M_{\mathbb{C}[\mathbb{Z}_m]} = M_{\langle 1\rangle}^{\oplus m}$. □

In the matrix coefficient basis the image is:

$$M_{\mathbb{C}[\mathbb{Z}_m]}(\mathbb{C}[\mathbb{Z}_m]^*) = \begin{pmatrix} y_0 & & & \\ & y_1 & & \\ & & \ddots & \\ & & & y_{m-1} \end{pmatrix}.$$

Exercise 3.5.3.3: (2) Show that if $T \in \hat{\sigma}_r^{0,h}$, then $\mathbf{R}(T) \leq r(h+1)$. ⊚

Exercise 3.5.3.4: (2) Obtain a fast algorithm for multiplying two polynomials in one variable by the method you used to solve the previous exercise. ⊚

Example 3.5.3.5. Consider $\mathfrak{S}_3$. In the standard basis,

$$M_{\mathbb{C}[\mathfrak{S}_3]}(\mathbb{C}[\mathfrak{S}_3]^*) = \begin{pmatrix} x_0 & x_1 & x_2 & x_3 & x_4 & x_5 \\ x_1 & x_0 & x_4 & x_5 & x_2 & x_3 \\ x_2 & x_5 & x_0 & x_4 & x_3 & x_1 \\ x_3 & x_4 & x_5 & x_0 & x_1 & x_2 \\ x_4 & x_3 & x_1 & x_2 & x_5 & x_0 \\ x_5 & x_2 & x_3 & x_1 & x_0 & x_4 \end{pmatrix}.$$

Here I have written an element of $\mathbb{C}[\mathfrak{S}_3]$ as $x_0 \operatorname{Id} + x_1(12) + x_2(13) + x_3(23) + x_4(123) + x_5(132)$. The irreducible representations of $\mathfrak{S}_3$ are the trivial, denoted $[3]$, the sign, denoted $[1,1,1]$ and the two-dimensional standard representation (the complement of the trivial in $\mathbb{C}^3$), which is denoted $[2,1]$. (See §8.6.5 for an explanation of the notation.) Since $\dim[3] = 1$, $\dim[1,1,1] = 1$ and $\dim[2,1] = 2$, by Theorem 3.5.2.1 $M_{\mathbb{C}[\mathfrak{S}_3]} = M_{\langle 1\rangle}^{\oplus 2} \oplus M_{\langle 2\rangle}$, and in the matrix coefficient basis:

$$M_{\mathbb{C}[\mathfrak{S}_3]}(\mathbb{C}[\mathfrak{S}_3]^*) = \begin{pmatrix} y_0 & & & & & \\ & y_1 & & & & \\ & & y_2 & y_3 & & \\ & & y_4 & y_5 & & \\ & & & & y_2 & y_3 \\ & & & & y_4 & y_5 \end{pmatrix}$$

where the blank entries are zero. We conclude $\mathbf{R}(M_{\mathbb{C}[\mathfrak{S}_3]}) \leq 1+1+7=9$.

3.5.4. Upper bounds via finite groups. Here is the main idea:

Use the standard basis to get a lower bound on the value of $M_{\mathbb{C}[G]}$ and the matrix coefficient basis to get an upper bound on its cost.

Say $M_{\mathbb{C}[G]}$ expressed in its standard basis combinatorially restricts to a sum of matrix multiplications, say $\oplus_{j=1}^{s} M_{\langle \mathbf{l}_j,\mathbf{m}_j,\mathbf{n}_j\rangle}$. The standard basis is particularly well suited to combinatorial restrictions because all the coefficients of the tensor in this basis are zero or one, and all the entries of the matrix $M_{\mathbb{C}[G]}(\mathbb{C}[G]^*)$ are nonzero and coordinate elements. (Recall that all the entries of the matrix $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}(A^*)$ are either zero or coordinate elements.) Using the matrix coefficient basis, we see $M_{\mathbb{C}[G]} = \oplus_{u=1}^{q} M_{\langle d_u\rangle}$, where d_u is the dimension of the u-th irreducible representation of G. Thus $\underline{\mathbf{R}}(\oplus_{j=1}^{s} M_{\langle \mathbf{l}_j,\mathbf{m}_j,\mathbf{n}_j\rangle}) \leq \underline{\mathbf{R}}(\oplus_{u=1}^{q} M_{\langle d_u\rangle})$ and $\mathbf{R}(\oplus_{j=1}^{s} M_{\langle \mathbf{l}_j,\mathbf{m}_j,\mathbf{n}_j\rangle}) \leq \mathbf{R}(\oplus_{u=1}^{q} M_{\langle d_u\rangle})$.

The asymptotic sum inequality implies:

Proposition 3.5.4.1. [**CU03, CU13**] *If $M_{\mathbb{C}[G]}$ degenerates to $\oplus_{j=1}^{s} M_{\langle \mathbf{l}_j,\mathbf{m}_j,\mathbf{n}_j\rangle}$ and d_u are the dimensions of the irreducible representations of G, then $\sum_{j=1}^{s}(\mathbf{l}_j\mathbf{m}_j\mathbf{n}_j)^{\frac{\omega}{3}} \leq \mathbf{R}(\oplus_{u=1}^{q} M_{\langle d_u\rangle}) \leq \sum d_u^3$. In fact, $\sum_{j=1}^{s}(\mathbf{l}_j\mathbf{m}_j\mathbf{n}_j)^{\frac{\omega}{3}} \leq \sum d_u^{\omega}$.*

In this section I will denote the standard basis for $\mathbb{C}[G]$ given by the group elements (which I have been denoting δ_{g_i}) simply by g_i.

Basis elements of $\mathbb{C}[G]$ are indexed by elements of G, so our sought-after combinatorial restriction is of the form:

$$\begin{aligned} \alpha &: [\mathbf{l}] \times [\mathbf{m}] \to G \\ \beta &: [\mathbf{m}] \times [\mathbf{n}] \to G \\ \gamma &: [\mathbf{n}] \times [\mathbf{l}] \to G. \end{aligned}$$

Recall the requirement that $t^{\alpha(i,j'),\beta(j,k'),\gamma(k,i')}$ is one if and only if $i=i'$, $j=j'$, $k=k'$, and is otherwise zero. Here, when considering $M_{\mathbb{C}[G]}$ as a trilinear map, we have

$$t^{\alpha,\beta,\gamma} = \begin{cases} 1 & \alpha\beta\gamma = \mathrm{Id} \\ 0 & \text{otherwise} \end{cases}$$

We want that $\alpha(i,j')\beta(j,k')\gamma(k,i') = \mathrm{Id}$ if and only if $i = i'$, $j = j'$, $k = k'$. To simplify the requirement, assume the maps factor to $s_1 : [\mathbf{l}] \to G$, $s_2 : [\mathbf{m}] \to G$, $s_3 : [\mathbf{n}] \to G$, and that $\alpha(i,j') = {s_1}^{-1}(i)s_2(j')$, $\beta(j,k') = {s_2}^{-1}(j)s_3(k')$ and $\gamma(k,i') = {s_3}^{-1}(k)s_1(i')$. Our requirement becomes

$${s_1}^{-1}(i)s_2(j'){s_2}^{-1}(j)s_3(k'){s_3}^{-1}(k)s_1(i') = \mathrm{Id} \Leftrightarrow i = i',\ j = j',\ k = k'.$$

Let S_j denote the image of s_j. Our requirement is summarized in the following definition:

Definition 3.5.4.2. [**CU03**] A triple of subsets $S_1, S_2, S_3 \subset G$ satisfy the *triple product property* if for any $s_j, s_j' \in S_j$, $s_1'{s_1}^{-1}s_2'{s_2}^{-1}s_3'{s_3}^{-1} = \mathrm{Id}$ implies $s_1' = s_1$, $s_2' = s_2$, $s_3' = s_3$.

There is a corresponding simultaneous triple product property when there is a combinatorial restriction to a collection of disjoint matrix multiplication tensors.

Example 3.5.4.3. [**CKSU05**] Let $G = (\mathbb{Z}_N^{\times 3} \times \mathbb{Z}_N^{\times 3}) \rtimes \mathbb{Z}_2$ where $\mathbb{Z}_2$ acts by switching the two factors, so $|G| = 2N^6$. Write elements of G as $[(\omega^i, \omega^j, \omega^k)(\omega^l, \omega^s, \omega^t)\tau^\epsilon]$ where $0 \le i,j,k,s,t,u \le N-1$, ω is a primitive N-th root of unity, τ is a generator of $\mathbb{Z}_2$, and $\epsilon \in \{0,1\}$. Set $\mathbf{l} = \mathbf{m} = \mathbf{n} = 2N(N-1)$. Label the elements of $[\mathbf{n}] = [2N(N-1)]$ by a triple (a,b,ϵ) where $1 \le a \le N-1$, $0 \le b \le N-1$ and $\epsilon \in \{0,1\}$, and define

$$\begin{aligned} s_1 &: [\mathbf{l}] \to G \\ (a,b,\epsilon) &\mapsto [(\omega^a,1,1)(1,\omega^b,1)\tau^\epsilon] \end{aligned}$$

$$\begin{aligned} s_2 &: [\mathbf{m}] \to G \\ (a,b,\epsilon) &\mapsto [(1,\omega^a,1)(1,1,\omega^b)\tau^\epsilon] \end{aligned}$$

$$\begin{aligned} s_3 &: [\mathbf{n}] \to G \\ (a,b,\epsilon) &\mapsto [(1,1,\omega^a)(\omega^b,1,1)\tau^\epsilon]. \end{aligned}$$

As explained in [**CKSU05**], the triple product property indeed holds (there are several cases), so $M_{\mathbb{C}[G]}$ combinatorially restricts to $M_{\langle 2N(N-1)\rangle}$. Now G has $2N^3$ irreducible one dimensional representations and $\binom{N^3}{2}$ irreducible two dimensional representations (see [**CKSU05**]). Thus $\mathbf{R}(M_{\langle 2N(N-1)\rangle}) \le 2N^3 + 8\binom{N^3}{2}$, which is less than $\mathbf{n}^3 = [2N(N-1)]^3$ for all $N \ge 5$. Asymptotically this is about $\frac{7}{16}\mathbf{n}^3$. If one applies Proposition 3.5.4.1 with $N = 17$ (which is optimal), one obtains $\omega < 2.9088$. Note that this does not even exploit Strassen's algorithm, so one actually has $\mathbf{R}(M_{\langle \mathbf{n}\rangle}) \le 2N^3 + 7\binom{N^3}{2}$, however this does not effect the asymptotics. If one could use the failure of additivity for border rank one potentially could do better.

While this is worse than what one would obtain just using Strassen's algorithm (writing $40 = 32 + 8$ and using Strassen in blocks), the algorithm is *different*. In [**CKSU05**] they obtain a bound of $\omega < 2.41$ by such methods, but key lemmas in their proof are almost the same as the key lemmas used by Coopersmith-Winograd in their optimizations.

3.5.5. Further ideas towards upper bounds. The structure tensor of $\mathbb{C}[G]$ had the convenient property that in the standard basis all the coefficients of the tensor are zero or one, and all entries of the matrix $M_{\mathbb{C}[G]}(\mathbb{C}[G]^*)$ are basis vectors. In [**CU13**] they propose looking at combinatorial restrictions of more general structure tensors, where the coefficients can be more general, but vestiges of these properties are preserved. They make the following definition, which is very particular to matrix multiplication:

Definition 3.5.5.1. We say $T \in A \otimes B \otimes C$, given in bases a_α, b_β, c_γ of A, B, C, *combinatorially supports* $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$, if such that, writing $T = \sum t^{\alpha,\beta,\gamma} a_\alpha \otimes b_\beta \otimes c_\gamma$, there exist injections

$$\alpha : [\mathbf{l}] \times [\mathbf{m}] \to [\mathbf{a}]$$
$$\beta : [\mathbf{m}] \times [\mathbf{n}] \to [\mathbf{b}]$$
$$\gamma : [\mathbf{n}] \times [\mathbf{l}] \to [\mathbf{c}]$$

such that $t^{\alpha(i,j'),\beta(j,k')\gamma(k,i')} \neq 0$ if and only if $i = i'$, $j = j'$ and $k = k'$. (Recall that T combinatorially restricts to $M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$ if moreover $t^{\alpha(i,j),\beta(j,k)\gamma(k,i)} = 1$ for all i, j, k.)

T combinatorially supports $M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}$ if there exists a coordinate expression of T such that, upon setting some of the coefficients in the multi-dimensional matrix representing T to zero, one obtains $\mathbf{mnl}$ nonzero entries such that in that coordinate system, matrix multiplication is supported on exactly those $\mathbf{mnl}$ entries. They then proceed to define the *s-rank* of a tensor T', which is the lowest rank of a tensor T that combinatorially supports it. This is a strange concept because *the s-rank of a generic tensor is one*: a generic tensor is combinatorially supported by $T = (\sum_j a_j) \otimes (\sum_k b_k) \otimes (\sum_l c_l)$ where $\{a_j\}$ is a basis of A etc..

Despite this, they show that $\omega \leq \frac{3}{2}\omega_s - 1$ where ω_s is the analog of the exponent of matrix multiplication for s-rank. In particular, $\omega_s = 2$ would imply $\omega = 2$. The idea of the proof is that if T combinatorially supports $M_{\langle \mathbf{n}\rangle}$, then $T^{\otimes 3}$ combinatorially degenerates to $M_{\langle \mathbf{n}\rangle}^{\oplus t}$ with $t = O(\mathbf{n}^{2-o(1)})$. Compare this with the situation when T combinatorially restricts to $M_{\langle \mathbf{n}\rangle}$, then $T^{\otimes 3}$ combinatorially restricts to $M_{\langle \mathbf{n} \rangle} \otimes M_{\langle \mathbf{n}^2 \rangle}$ and thus toric degenerates to $M_{\langle \mathbf{n} \rangle}^{\oplus \lfloor \frac{3}{4} \mathbf{n}^2 \rfloor}$ by Theorem 3.4.3.1.

Chapter 4

The complexity of Matrix multiplication III: explicit decompositions via geometry

One might argue that the exponent of matrix multiplication is unimportant for the world we live in, since ω might not be relevant until the sizes of the matrices are on the order of number of atoms in the known universe. For implementation, it is more important to develop explicit decompositions that provide a savings for matrices of sizes that need to be multiplied in practice. One purpose of this chapter is to construct such decompositions. Another is to gain insight into the asymptotic situation by exploring what symmetry groups occur in decompositions of $M_{\langle \mathbf{n} \rangle}$. I begin in §4.1 by discussing generalities about decompositions: the generalized Comon conjecture posting that optimal decompositions with symmetry exist, a review of Strassen's original decomposition of $M_{\langle 2 \rangle}$ that hints that this is indeed the case, and defining symmetry groups of decompositions. In particular, I point out that decompositions come in *families* essentially parametrized by $G_{M_{\langle \mathbf{n} \rangle}}$, and one gains insight studying the entire family rather than individual members. I then, in §4.2, describe an example, the Waring decomposition of $x_1 \cdots x_n$, where we understand everything, observing that the optimal decomposition has some symmetry and there are near optimal decompositions with "maximal" symmetry. If one could prove either of these properties hold in the context of Valiant's conjecture, it would prove the conjecture, see §7.4.7. If this holds in the context of matrix multiplication, it would simplify the problem considerably. In §4.3 I revisit Strassen's decomposition and give a proof of Burichenko's theorem [**Bur14**] that its symmetry group is as large as one could expect. In §4.4 I briefly describe the alternating least squares method that has been used to find decompositions numerically. In order to exploit symmetry groups, one needs to understand the tensors that are invariant under them. I describe the simple case of cyclic symmetry in §4.5. In order to determine symmetry groups and determine if different decompositions are in the same family, one needs invariants of decompositions. These are studied in §4.6. Two interesting examples of decompositions of $M_{\langle 3\rangle}$ are given in §4.7. Border rank decompositions also have geometry associated with them. In order to describe the geometry, I give some geometric preliminaries, including the definition of secant varieties in general in §4.8. I conclude with two examples of border rank decompositions and their geometry in §4.9.

4.1. Symmetry and decompositions

4.1.1. The Comon conjecture and its generalization. In 2008 there was an AIM workshop, *Geometry and Representation theory of tensors for computer science, statistics and other areas*, that brought together a very diverse group of researchers. Among them was Pierre Comon, an engineer working in signal processing. In signal processing (at least practiced by Comon), one wants to decompose tensors presumed to be of rank r explicitly into a sum of r rank one tensors. Sometimes the relevant tensors are symmetric. At the workshop Comon presented the following conjecture:

Conjecture 4.1.1.1 (P. Comon [**Com02**]). *If $T \in S^d\mathbb{C}^N \subset (\mathbb{C}^N)^{\otimes d}$, then there exists an optimal rank decomposition of T made from symmetric tensors.*

After being greeted with skepticism by algebraic geometers, the community has now embraced this conjecture and generalized it. Recall that $S^d\mathbb{C}^N$ admits the interpretation of the tensors in $(\mathbb{C}^N)^{\otimes d}$ invariant under $\mathfrak{S}_d$, i.e., $S^d\mathbb{C}^n = ((\mathbb{C}^N)^{\otimes d})^{\mathfrak{S}_d}$

Consider a rank decomposition of T, $T = \sum_{j=1}^r t_j$ with $\mathbf{R}(t_j) = 1$. The order of the summands does not matter so it is more natural to consider the set $\mathcal{S} = \{t_1, \dots, t_r\}$, and call $\mathcal{S}$ the rank decomposition of T.

Conjecture 4.1.1.2. *[Generalized Comon Conjecture]* [**BILR**] *Let $T \in (\mathbb{C}^N)^{\otimes d}$ be invariant under some $\Gamma \subset \mathfrak{S}_d$. Then there exists an optimal* *rank decomposition $\mathcal{S}$ of T built from Γ-invariant tensors. I.e., for all $g \in \Gamma$, $g\mathcal{S} = \mathcal{S}$.*

Recall that matrix multiplication $M_{\langle \mathbf{n} \rangle} \in (\mathbb{C}^{\mathbf{n}^2})^{\otimes 3}$ is invariant under the cyclic permutation of factors $\mathbb{Z}_3 \subset \mathfrak{S}_3$ because $\mathrm{trace}(XYZ) = \mathrm{trace}(YZX)$, so Conjecture 4.1.1.2 predicts that there should be optimal rank decompositions $M_{\langle \mathbf{n} \rangle} = \sum_{j=1}^{r} t_j$ where the t_j are permuted among themselves by the $\mathbb{Z}_3$.

4.1.2. Strassen's decomposition. Introduce the notation

$$\langle x \otimes y \otimes z \rangle_{\mathbb{Z}_3} := x \otimes y \otimes z + y \otimes z \otimes x + z \otimes x \otimes y.$$

With this notation, Strassen's algorithm, written as a tensor, is

$$\begin{aligned} M_{\langle 2 \rangle} = & \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes 3} \\ & + \langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \rangle_{\mathbb{Z}_3} \\ & - \langle \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \rangle_{\mathbb{Z}_3}. \end{aligned} \tag{4.1.1}$$

In particular, it is transparently built from cyclic $\mathbb{Z}_3$-invariant tensors. It also looks like it may have further symmetry. To discuss this, we need some language.

4.1.3. Generalities on rank decompositions. Consider $Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_d) \subset \mathbb{P}(A_1 \otimes \cdots \otimes A_d)$. If all the vector spaces have different dimensions, we consider the symmetry group of the cone over the Segre as a subgroup of $GL(A_1) \times \cdots \times GL(A_d)$ (more precisely of $GL(A_1) \times \cdots \times GL(A_d)/(\mathbb{C}^*)^{d-1}$, because if $\lambda_1 \cdots \lambda_d = 1$, then $(\lambda_1 \mathrm{Id}_{A_1}, \ldots, \lambda_d \mathrm{Id}_{A_d}) \in GL(A_1) \times \cdots \times GL(A_d)$ acts trivially). If all dimensions are the same, we consider the symmetry group as a subgroup of $GL(A_1) \times \cdots \times GL(A_d) \rtimes \mathfrak{S}_d$, where the $\mathfrak{S}_d$ acts by permuting the factors after some isomorphism of the A_j has been chosen. One can also consider intermediate cases. For $T \in (\mathbb{C}^N)^{\otimes d}$, let

$$G_T := \{ g \in GL_N^{\times d} \rtimes \mathfrak{S}_d \mid gT = T \},$$

and for $T \in A_1 \otimes \cdots \otimes A_d$ with different dimensions, define

$$G_T := \{ g \in GL(A_1) \times \cdots \times GL(A_d) \mid gT = T \}.$$

For a polynomials $P \in S^d V$, the symmetry group of the cone over the Veronese $v_d(\mathbb{P}V) \subset \mathbb{P}S^d V$ is $GL(V)$, and we write

$$G_P := \{ g \in GL(V) \mid gP = P \}.$$

If $T \in A_1 \otimes \cdots \otimes A_d$ has a rank decomposition $\mathcal{S}$ and a nontrivial symmetry group G_T, then given $g \in G_T$, $g \cdot \mathcal{S} := \{gt_1, \dots, gt_r\}$ is another rank decomposition of T or $\mathcal{S}$.

Definition 4.1.3.1. The *symmetry group of the decomposition* $\mathcal{S}$ is $\Gamma_{\mathcal{S}} := \{g \in G_T \mid g \cdot \mathcal{S} = \mathcal{S}\}$. Let $\Gamma'_{\mathcal{S}} = \Gamma_{\mathcal{S}} \cap (\Pi_j GL(A_j))$.

If T is concise, then $\Gamma_{\mathcal{S}}$ is a finite group because any decomposition of it would have to include a basis of each of each A_j and the subgroup of $GL(A_j)$ preserving a basis is finite (and the scale ambiguity is gone when one quotients by the $(\mathbb{C}^*)^{d-1}$. In particular, rank decompositions of T come in $\dim G_T$-dimensional families It will be useful to study the whole family as a geometric object, as well as looking for convenient members of the family in the sense described below.

A guiding hypothesis of this chapter (for which there is no theoretical justification, but has been true in several cases) is that if T has a large symmetry group, then there will exist optimal decompositions of T with some symmetry and geometry. This even extends to border rank decompositions, as we will see in §4.8.4.

Naïvely, one might think that some decompositions in a family have better symmetry groups than others. Strictly speaking this is not correct:

Proposition 4.1.3.2. *For $g \in G_T$, $\Gamma_{g \cdot \mathcal{S}} = g\Gamma_{\mathcal{S}} g^{-1}$.*

Proof. Let $h \in \Gamma_{\mathcal{S}}$, then $ghg^{-1}(gt_j) = g(ht_j) \in g \cdot \mathcal{S}$ so $\Gamma_{g \cdot \mathcal{S}} \subseteq g\Gamma_{\mathcal{S}_t} g^{-1}$, but the construction is symmetric in $\Gamma_{g \cdot \mathcal{S}}$ and $\Gamma_{\mathcal{S}}$. $\square$

For a polynomial $P \in S^d V$ and a decomposition $P = \ell_1^d + \cdots + \ell_r^d$ for some $\ell_j \in V$ (such is often called a *Waring decomposition*), and $g \in G_P \subset GL(V)$, the same result holds with $\mathcal{S} = \{\ell_1, \dots, \ell_r\}$.

In summary, decompositions come in $\dim(G_T)$-dimensional families, and each member of the family has the same abstract symmetry group.

4.2. Example: the polynomial $x_1 \cdots x_n$

Consider the polynomial $e_{n,n} := x_1 \cdots x_n \in S^n \mathbb{C}^n$ (the n-th elementary symmetric function in n variables). We first determine $G_{e_{n,n}}$: It is clear $T_n^{SL} \rtimes \mathfrak{S}_n \subset G_{e_{n,n}}$, where T_n^{SL} denotes the diagonal matrices with determinant one and $\mathfrak{S}_n$ acts by permuting the basis vectors. We need to determine if the stabilizer is larger. Let $G \in GL_n$. Then

$$g \cdot e_{n,n} = \sum_{j_1, \dots, j_n = 1}^{n} (g_1^{j_1} x_{j_1}) \cdots (g_n^{j_n} x_{j_n}).$$

In order that this be equal to $x_1 \cdots x_n$, by unique factorization of polynomials, there must be a permutation $\sigma \in \mathfrak{S}_n$ such that for each k, we have $\sum_j g^j_k x_j = \lambda_k x_{\sigma(k)}$ for some $\lambda_k \in \mathbb{C}^*$. Composing with the inverse of this permutation we have $g^j_k = \delta^j_k \lambda_j$, and finally we see that we must further have $\lambda_1 \cdots \lambda_n = 1$, which means it is an element of T^{SL}_n, so the original g is an element of $T^{SL}_n \rtimes \mathfrak{S}_n$. Thus $G_{e_{n,n}} = T^{SL}_n \rtimes \mathfrak{S}_n$.

Remark 4.2.0.1. The $GL(V)$-orbit closure of $e_{n,n}$ is the Chow variety $Ch_n(V) \subset \mathbb{P}S^n V$ of §3.1.2 that we will study in detail in Chapter 9.

The optimal Waring decomposition of $x_1 \cdots x_n$, dates back at least to Bochnak and Siciak (1971) [**BS71**], although they say the proof comes from the 1934 Mazur Orlicz paper [**MO34**]. It was proved to be optimal in [**RS11**] (I give the proof in §**??**) is

$$x_1 \cdots x_n = \frac{1}{2^{n-1} n!} \sum_{\substack{\epsilon \in \{-1,1\}^n \\ \epsilon_1 = 1}} \big(\sum_{j=1}^n \epsilon_j x_j \big)^n \Pi^n_{i=1} \epsilon_i, \tag{4.2.1}$$

a sum with 2^{n-1} terms. This decomposition has an $\mathfrak{S}_{n-1}$-symmetry but not an $\mathfrak{S}_n$-symmetry, nor is it preserved by any element of T^{SL}_n. One can obtain an $\mathfrak{S}_n$-invariant expression by doubling the size:

$$x_1 \cdots x_n = \frac{1}{2^n n!} \sum_{\epsilon \in \{-1,1\}^n} \big(\sum_{j=1}^n \epsilon_j x_j \big)^n \Pi^n_{i=1} \epsilon_i, , \tag{4.2.2}$$

because

$$\begin{aligned} (-x_1 + \epsilon_2 x_2 + \cdots + &\epsilon_n x_n)^n (-1)\epsilon_2 \ldots \epsilon_n \\ &= (-1)^n (x_1 + (-\epsilon_2) x_2 + \cdots + (-\epsilon_n) x_n)^n (-1) \epsilon_2 \cdots \epsilon_n \\ &= (x_1 + (-\epsilon_2) x_2 + \cdots + (-\epsilon_n) x_n)^n (-\epsilon_2) \cdots (-\epsilon_n). \end{aligned}$$

From this example we see:

- The optimal decomposition has some symmetry.
- A decomposition with "maximal" symmetry exists that is only slightly larger (within a factor of two).

As we will see in §**??**, if one could prove either of these properties holds in the situation of Valiant's conjecture, then one could prove Valiant's conjecture. In this chapter I will take these as working hypotheses in the search for rank and border rank decompositions of the matrix multiplication tensor.

4.3. Strassen's decomposition revisited

Let $\mathcal{S}tr$ denote the Strassen decomposition of $M_{\langle 2 \rangle}$.

4.3.1. Symmetries of $M_{\langle \mathbf{n}\rangle}$. In order to discuss symmetries of decompositions, we need to determine the symmetry group of the matrix multiplication tensor

$$G_{M_{\langle \mathbf{n}\rangle}} := \{g \in GL_{n^2}^{\times 3} \rtimes \mathfrak{S}_3 \mid g \cdot M_{\langle \mathbf{n}\rangle} = M_{\langle \mathbf{n}\rangle}\}.$$

One may also consider matrix multiplication as a polynomial that happens to be multi-linear, $M_{\langle \mathbf{n}\rangle} \in S^3(A \oplus B \oplus C)$, and consider

$$\tilde{G}_{M_{\langle \mathbf{n}\rangle}} := \{g \in GL(A \oplus B \oplus C) \mid g \cdot M_{\langle \mathbf{n}\rangle} = M_{\langle \mathbf{n}\rangle}\}.$$

Note that $(GL(A) \times GL(B) \times GL(C)) \rtimes \mathfrak{S}_3 \subset GL(A \oplus B \oplus C)$, so $G_{M_{\langle \mathbf{n}\rangle}} \subseteq \tilde{G}_{M_{\langle \mathbf{n}\rangle}}$.

Let $PGL(U)$ denote $GL(U)/\mathbb{C}^*$, where $\mathbb{C}^* = \{\lambda \operatorname{Id}_U \mid \lambda \in \mathbb{C}^*\}$. This group acts on $\mathbb{P}U$, as well as on $U^*{\otimes}U$. The first action is clear, the second because the action of $GL(U)$ on $\alpha{\otimes}u$ is $\alpha g^{-1}{\otimes}gu$ so the scalars times the identity will act trivially.

It is clear that $PGL_{\mathbf{n}} \times PGL_{\mathbf{n}} \times PGL_{\mathbf{n}} \rtimes \mathbb{Z}_3 \subset G_{M_{\langle \mathbf{n}\rangle}}$, the $\mathbb{Z}_3$ because $\operatorname{trace}(XYZ) = \operatorname{trace}(YZX)$. Moreover since $\operatorname{trace}(XYZ) = \operatorname{trace}(Y^TX^TZ^T)$, we have $PGL_n^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2) \subseteq G_{M_{\langle \mathbf{n}\rangle}}$. We emphasize that this $\mathbb{Z}_2$ is not contained in either the $\mathfrak{S}_3$ permuting the factors or the $PGL(A) \times PGL(B) \times PGL(C)$ acting on them. In $\tilde{G}_{M_{\langle \mathbf{n}\rangle}}$ we can also rescale the three factors by non-zero complex numbers λ, μ, ν such that $\lambda\mu\nu = 1$, so we have $(\mathbb{C}^*)^{\times 2} \times PGL_n^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2) \subseteq G_{M_{\langle \mathbf{n}\rangle}}$.

We will be primarily interested in $G_{M_{\langle \mathbf{n}\rangle}}$. The first equality in the following proposition appeared in [**dG78**, Thms. 3.3,3.4] and [**Bur15**, Prop. 4.7] with ad-hoc proofs. The second assertion appeared in [**Ges16**].

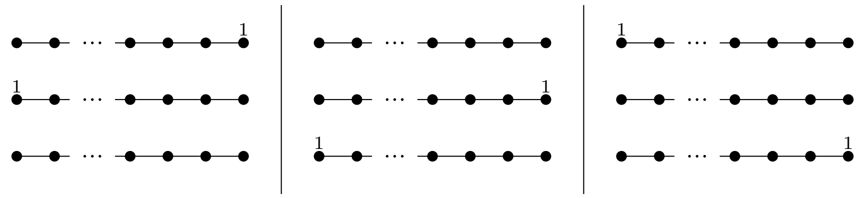
Proposition 4.3.1.1. *$G_{M_{\langle \mathbf{n}\rangle}} = PGL_n^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$ and $\tilde{G}_{M_{\langle \mathbf{n}\rangle}} = (\mathbb{C}^*)^{\times 2} \times PGL_n^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$.*

Remark 4.3.1.2. It would be more natural to write $\tilde{G}_{M_{\langle \mathbf{n}\rangle}} = (GL_n^{\times 3}/\mathbb{C}^*) \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$, but we write it in the above manner to facilitate comparison with $G_{M_{\langle \mathbf{n}\rangle}}$.

A "hands on" elementary proof is possible, see, e.g. [**Bur15**, Prop. 4.7]. For those who know about Dynkin diagrams, here is an elegant proof from [**Ges16**].

Proof. It will be sufficient to show the second equality because the $(\mathbb{C}^*)^{\times 2}$ acts trivially on $A{\otimes}B{\otimes}C$. For polynomials, we use the method of [**BGL14**, Prop. 2.2] adapted to reducible representations. A straight-forward Lie algebra calculation shows the connected component of the identity of $\tilde{G}_{M_{\langle \mathbf{n}\rangle}}$ is $\tilde{G}^0_{M_{\langle \mathbf{n}\rangle}} = (\mathbb{C}^*)^{\times 2} \times PGL_n^{\times 3}$. As was observed in [**BGL14**] the full stabilizer group must be contained in its normalizer $N(\tilde{G}^0_{M_{\langle \mathbf{n}\rangle}})$, see Proposition

8.13.1.1. But the normalizer of $\tilde{G}^0_{M_{\langle \mathbf{n}\rangle}}$ quotiented by $\tilde{G}^0_{M_{\langle \mathbf{n}\rangle}}$ is the automorphism group of the marked Dynkin diagram for $A \oplus B \oplus C$, which in our case is



There are three triples of marked diagrams. Call each column consisting of 3 marked diagrams a group. The automorphism group of the picture is $\mathbb{Z}_3 \rtimes \mathbb{Z}_2$, where the $\mathbb{Z}_2$ may be seen as flipping each diagram, exchanging the first and third diagram in each group, and exchanging the first and second group. The $\mathbb{Z}_3$ comes from cyclically permuting each group and the diagrams within each group. $\square$

4.3.2. The Strassen family. As discussed above, decompositions are best studied in families.

Theorem 4.3.2.1. [**dG78**] *The set of rank seven decompositions of $M_{\langle 2\rangle}$ is the orbit $G_{M_{\langle 2\rangle}} \cdot Str$.*

The proof follows from a careful analysis of every possible decomposition, taking into account that an element $a{\otimes}b{\otimes}c$ is not just a triple of vectors, but a triple of endomorphisms $\mathbb{C}^2 \to \mathbb{C}^2$, and the analysis is via the possible triples of ranks that can appear.

In preparation for studying the Strassen family of decompositions, write

$$u_1 = \begin{pmatrix}1\\0\end{pmatrix},\ u_2 = \begin{pmatrix}0\\1\end{pmatrix},\ u^1 = (1,0)\ u^2 = (0,1) \tag{4.3.1}$$

and set $v_j = w_j = u_j$ and $v^j = w^j = u^j$. Strassen's decomposition becomes

$$\begin{aligned} M_{\langle 2\rangle} =&(v_1u^1 + v_2u^2){\otimes}(w_1v^1 + w_2v^2){\otimes}(u_1w^1 + u_2w^2)\\ &+ \langle v_1u^1{\otimes}w_2(v^1 - v^2){\otimes}(u_1 + u_2)w^2\rangle_{\mathbb{Z}_3}\\ &+ \langle v_2u^2{\otimes}w_1(v^2 - v^1){\otimes}(u_1 + u_2)w^1\rangle_{\mathbb{Z}_3}. \end{aligned} \tag{4.3.2}$$

From this presentation we recover much the entire Strassen family, namely by letting $u_1, u_2,\ v_1, v_2$, and w_1, w_2 be arbitrary bases, with dual basis vectors denoted with superscripts. We obtain a family parametrized by $PGL(U)\times PGL(V)\times PGL(W)$, and since the decomposition (4.3.2) is manifestly $\mathbb{Z}_3$-invariant, the only potential additional decompositions arise from applying a convenient transpose symmetry such as $x{\otimes}y{\otimes}z \mapsto x^T{\otimes}z^T{\otimes}y^T$.

Exercise 4.3.2.2: (2) Show that if we change bases by

$$g_U = \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \in GL(U),\ g_V = \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \in GL(V),\ g_W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in GL(W),$$

Then the new decomposition has four terms fixed by the standard cyclic $\mathbb{Z}_3$. ⊚

Exercise 4.3.2.3: (1) Note that if we set $u_3 = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ then the matrices in Exercise 4.3.2.2 respectively correspond to the permutations $(2,3)$, $(1,3)$ and $(1,2)$. The matrix in the first term of the decomposition that one obtains from Exercise 4.3.2.2 also corresponds to a permutation. Which one?

Exercise 4.3.2.4: (2) Find a change of basis such that the first term in the decomposition of Exercise 4.3.2.2 becomes $\begin{pmatrix} \omega & 0 \\ 0 & \omega^2 \end{pmatrix}^{\otimes 3}$ where $\omega = e^{\frac{2\pi i}{3}}$ and write out the decomposition in this basis.

Under $x \otimes y \otimes z \mapsto x^T \otimes z^T \otimes y^T$, Strassen's decomposition is mapped to:

$$\begin{aligned} M_{\langle 2\rangle} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes 3} \\ &+ \langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix} \rangle_{\mathbb{Z}_3} \\ &- \langle \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix} \rangle_{\mathbb{Z}_3}. \end{aligned} \tag{4.3.3}$$

Notice that this is almost Strassen's decomposition (4.1.1)- just some the signs are wrong. We can "fix" the problem by conjugating all the matrices with

$$g_0 := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Exercise 4.3.2.5: (1) Verify that acting by $g_0^{\times 3} \in PGL(U) \times PGL(V) \times PGL(W)$ takes (4.3.3) to Strassen's decomposition, so acting on Strassen's decomposition by $({g_0}^{-1})^{\times 3}$ takes it to (4.3.3).

Exercise 4.3.2.5 shows that there is a non-standard $\mathbb{Z}_2 \subset PGL_2^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$ contained in $\Gamma_{\mathcal{S}tr}$, namely the convenient transpose symmetry composed with $g_0^{\times 3}$. We also obtain a refinement of deGroote's theorem:

Proposition 4.3.2.6. *The set of rank seven decompositions of $M_{\langle 2\rangle}$ is $PGL_2^{\times 3} \cdot \mathcal{S}tr$.*

With the expression (4.3.2), notice that if we exchange $u_1 \leftrightarrow u_2$ and $u^1 \leftrightarrow u^2$, the decomposition is also preserved by this $\mathbb{Z}_2 \subset PGL_2^{\times 3}$, with orbits (4.3.2) and the exchange of the triples. So we see $\Gamma_{\mathcal{S}tr} \supseteq \mathbb{Z}_2 \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$, where the first $\mathbb{Z}_2$ is diagonally embedded in $PGL_2^{\times 3}$.

Although the above description of the Strassen family of decompositions for $M_{\langle 2\rangle}$ is satisfying, it becomes even more transparent with a projective perspective. With the projective perspective, we will see that $\Gamma_{\mathcal{S}tr}$ is even larger.

4.3.3. $M_{\langle 2\rangle}$ **viewed projectively.** That all rank 7 decompositions of $M_{\langle 2\rangle}$ are obtained via $PGL_2^{\times 3}$ suggests using a projective perspective. The group PGL_2 acts simply transitively on triples of distinct points of $\mathbb{P}^1$. So to fix a decomposition in the family, select a triple of points in each space. I focus on $\mathbb{P}U$. Call the points $[u_1], [u_2], [u_3]$. Then these determine three points in $\mathbb{P}U^*$, $[u^{1\perp}], [u^{2\perp}], [u^{3\perp}]$. Choose representatives u_1, u_2, u_3 satisfying $u_1+u_2+u_3 = 0$. I could have taken any linear relation, it just would introduce coefficients in the decomposition. I take the most symmetric relation to keep all three points on an equal footing. Similarly, fix the scales on the $u^{j\perp}$ by requiring $u^{j\perp}(u_{j-1}) = 1$ and $u^{j\perp}(u_{j+1}) = -1$, where indices are considered mod $\mathbb{Z}_3$, so $u_{3+1} = u_1$ and $u_{1-1} = u_3$.

In comparison with what we had before, letting the old vectors be hatted, we have $\hat{u}_1 = u_1$, $\hat{u}_2 = u_2$, $\hat{u}^1 = u^{2\perp}$, and $\hat{u}^2 = -u^{1\perp}$. The effect is to make the symmetries of the decomposition more transparent. Our identifications of the ordered triples $\{u_1, u_2, u_3\}$ and $\{v_1, v_2, v_3\}$ exactly determine a linear isomorphism $a_0 : U \to V$, and similarly for the other pairs of vector spaces. Note that $a_0 = v_j \otimes u^{j+1\perp} + v_{j+1} \otimes u^{j+2\perp}$ for any $j = 1, 2, 3$.

Then

$$
\begin{aligned}
M_{\langle 2\rangle} &= a_0 \otimes b_0 \otimes c_0 \\
&\quad + \langle (v_1 u^{2\perp}) \otimes (w_3 v^{1\perp}) \otimes (u_2 w^{3\perp}) \rangle_{\mathbb{Z}_3} \\
&\quad + \langle (v_1 u^{3\perp}) \otimes (w_2 v^{1\perp}) \otimes (u_3 w^{2\perp}) \rangle_{\mathbb{Z}_3}.
\end{aligned}
\tag{4.3.4}
$$

Here, to make the terms shifted by $\mathbb{Z}_3$ live in the proper space, one must act by a_0, b_0, c_0 appropriately, e.g., to shift $v_1 u^{2\perp}$ to the second slot, one takes $b_0(v_1) a_0^T (u^{2\perp})$.

With this presentation, the diagonally embedded $\mathfrak{S}_3 \subset PGL_2^{\times 3}$ acting by permuting the indices transparently preserves the decomposition, with two orbits, the fixed point $a_0 \otimes b_0 \otimes c_0$ and the orbit of $(v_1 u^{2\perp}) \otimes (w_3 v^{1\perp}) \otimes (u_2 w^{3\perp})$. The action on each of U, V, W is the standard irreducible representation $[2, 1]$.

We now see $\Gamma_{\mathcal{S}tr} \supseteq \mathfrak{S}_3 \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$, with $\mathfrak{S}_3 \subset \Gamma'_{\mathcal{S}tr}$.

4.3.4. Symmetries of decompositions of $M_{\langle \mathbf{n}\rangle}$**.** Let $M_{\langle \mathbf{n}\rangle} = \sum_{j=1}^r t_j$ be a rank decomposition for $M_{\langle \mathbf{n}\rangle}$ and write $t_j = a_j \otimes b_j \otimes c_j$. Let $\mathbf{r}_j :=$

$(\mathrm{rank}(a_j), \mathrm{rank}(b_j), \mathrm{rank}(c_j))$, and let $\tilde{\mathbf{r}}_j$ denote the unordered triple. The following proposition is clear:

Proposition 4.3.4.1. *Let $\mathcal{S}$ be a rank decomposition of $M_{\langle \mathbf{n} \rangle}$. Partition $\mathcal{S}$ by un-ordered rank triples into disjoint subsets: $\{\tilde{\mathcal{S}}_{1,1,1}, \tilde{\mathcal{S}}_{1,1,2}, \dots, \tilde{\mathcal{S}}_{n,n,n}\}$. Then $\Gamma'_{\mathcal{S}}$ preserves each $\tilde{\mathcal{S}}_{s,t,u}$. The same holds for order rank triples $\mathcal{S}_{s,t,u}$.*

We can say more about rank one elements: If $a \in U^* \otimes V$ and $\mathrm{rank}(a) = 1$, then there are unique points $[\mu] \in \mathbb{P}U^*$ and $[v] \in \mathbb{P}V$ such that $[a] = [\mu \otimes v]$. So given a decomposition $\mathcal{S}$ of $M_{\langle \mathbf{n} \rangle}$, define $\mathcal{S}_{U^*} \subset \mathbb{P}U^*$ and $\mathcal{S}_U \subset \mathbb{P}U$ to correspond to the U^* and U elements appearing in $\mathcal{S}_{1,1,1}$. Then $\Gamma'_{\mathcal{S}}$ preserves $\mathcal{S}_U$ and $\mathcal{S}_{U^*}$.

We will say a decomposition has a *transpose-like $\mathbb{Z}_2$ invariance* if it is invariant under a $\mathbb{Z}_2$ such as $x \otimes y \otimes z \mapsto x^T \otimes z^T \otimes y^T$ composed with an element of $PGL(U) \times PGL(V) \times PGL(W)$.

Exercise 4.3.4.2: (1) Show that if a decomposition of $M_{\langle \mathbf{n} \rangle}$ is cyclic $\mathbb{Z}_3$-invariant and also has a transpose-like $\mathbb{Z}_2$-invariance, then $\mathcal{S}_U$ and $\mathcal{S}_{U^*}$ have the same cardinality.

4.3.5. Symmetries of $\mathcal{S}tr$. In the case of Strassen's decomposition $\mathcal{S}tr_U$ is a configuration of three points in $\mathbb{P}^1$, so *a priori* we must have the projection of $\Gamma'_{\mathcal{S}tr}$ onto $PGL(U)$ is contained in $\mathfrak{S}_3$. If we restrict to the subfamily of decompositions where there is a standard cyclic $\mathbb{Z}_3$-symmetry, there is just one PGL_2 and we have $\Gamma'_{\mathcal{S}tr} \subseteq \mathfrak{S}_3$. Recall that this is no loss of generality as the full symmetry group is the same for all decompositions in the family. We conclude $\Gamma_{\mathcal{S}tr} \subseteq \mathfrak{S}_3 \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$ and: ***previous paragraph may need clarification***

Theorem 4.3.5.1. [**Bur14**] *The symmetry group $\Gamma_{\mathcal{S}tr}$ of Strassen's decomposition of $M_{\langle 2 \rangle}$ is $\mathfrak{S}_3 \times (\mathbb{Z}_3 \rtimes \mathbb{Z}_2) \subset PGL_2^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2) = G_{M_{\langle 2 \rangle}}$.*

Remark 4.3.5.2. One can prove Strassen's decomposition is actually matrix multiplication without checking directly simply by the group invariance, see [**CHI**$^+$].

4.4. Alternating least squares (ALS) approach to decompositions

Let A,B,C respectively have bases $\{e_i\}$, $\{f_j\}$, $\{g_k\}$. Given a tensor $T = \sum_{i=1}^{\mathbf{a}} \sum_{j=1}^{\mathbf{b}} \sum_{k=1}^{\mathbf{c}} t^{ijk} e_i \otimes f_j \otimes g_k \in A \otimes B \otimes C$, say we have reason to believe it has rank at most r. To find a rank r expression we could work as follows: For $1 \leq u \leq r$, write $a_u = \sum_i X^i_u e_i$, $b_u = \sum_j Y^j_u f_j$, and $c_u = \sum_k Z^k_u g_k$ where the X^i_u, Y^j_u, Z^k_u are constants to be determined. We want $\sum_{u=1}^r a_u \otimes b_u \otimes c_u = T$, i.e.,

$$\sum_{u=1}^{r} X_u^i Y_u^j Z_u^k = t^{ijk} \tag{4.4.1}$$

for all i, j, k. If we restrict ourselves to real coefficients, we want

$$\text{objfn}_1 := \sum_{i,j,k} (\sum_{u=1}^{r} X_u^i Y_u^j Z_u^k - t^{ijk})^2, \tag{4.4.2}$$

called the *objective function*, to be zero. (One can obtain a similar equation for complex coefficients by splitting all complex numbers into their real and imaginary parts. I stick to the real presentation just for simplicity of exposition.) Now (4.4.2) is a degree six polynomial, but it is quadratic in each of the unknown quantities. To solve in practice, one begins with an initial "guess" of the X_u^i, Y_u^j, Z_u^k, e.g., chosen at random. Then one tries to minimize (4.4.2) e.g., as a function of the X_u^i while holding the Y_u^j, Z_u^k fixed. This is a linear **quadratic??** problem. Once one obtains a solution, one starts again, holding the X_u^i and Z_u^k fixed and solving for the Y_u^j. Then one repeats, minimizing for the Z_u^k, and then cycling around again and again until the result converges (or fails to, in which case one can start again with different initial points). This algorithm was first written down in [**Bre70**].

Now if $\underline{\mathbf{R}}(T) < \mathbf{R}(T)$ (as is expected to be the case with matrix multiplication), this procedure could "attempt" to find a border rank solution, that is, the coefficients could go off to infinity. If one wants a rank decomposition, one can add a penalty term to (4.4.2), instead minimizing

$$\text{objfn}_2 := \sum_{i,j,k} (\sum_{u=1}^{r} X_u^i Y_u^j Z_u^k - t^{ijk})^2 + \epsilon(\sum_{u,i,j,k}^{r} (X_u^i)^2 + (Y_u^j)^2 + (Z_u^k)^2) \tag{4.4.3}$$

for some ϵ that in practice is found by trial and error.

In the literature (e.g. [**Lad76, JM86, Smi13, ?**]) they prefer coefficient values to be from a small list of numbers, ideally confined to something like $0, \pm 1$ or $0, \pm 1, \pm \frac{1}{2}$. If the tensor in question has a large symmetry group (as does matrix multiplication), one can use the group action to fix some of the coefficients to these desired values.

According to Smirnov, in [**Smi13**], for $T = M_{\langle \mathbf{n} \rangle}$ (but not rectangular matrix multiplication) the critical points of objfn_1 are integers in practice, although he does not give an explanation why one would expect this to be the case. Thus, by these heuristics, if one can obtain a decomposition with $\text{objfn}_1 < 1$, then it will converge to zero by the ALS process, producing either a decomposition or limiting to a border rank decomposition.

Smirnov allows ϵ in (4.4.3) to gradually increase while imposing restrictions from the previous ALS round on the values the coefficients are allowed to take.

4.5. Decomposition of $A^{\otimes 3}$ under $\mathbb{Z}_3$

In order to search for cyclic $\mathbb{Z}_3$ decompositions of $M_{\langle \mathbf{n} \rangle}$ we need to understand the $GL(A)$-decomposition of $A^{\otimes 3}$.

Exercise 4.5.0.1: (1!) Verify that the cyclic $\mathbb{Z}_3$ acts trivially on both S^3A and $\Lambda^3 A$.

We have seen that $A^{\otimes 2} = S^2A \oplus \Lambda^2 A$ as a $GL(A)$-module and that this decomposition is into irreducible submodules. If we consider $A^{\otimes 3}$, we know it contains the irreducible submodules S^3A and Λ^3A, but a simple dimension count shows that these two modules do not span $A^{\otimes 3}$.

We have also seen that symmetrization and skew-symmetrization commute with the action of $GL(A)$. So the following skew-symmetrization map is a $GL(A)$-module map:

$$\Lambda^2 A \otimes A \to \Lambda^3 A.$$

Thus its kernel (a linear subspace of $A^{\otimes 3}$) is a $GL(A)$-submodule and it is distinct from S^3A and Λ^3A (either by dimension counting or in the first case observing the skew-symmetry in the first two factors and in the second, the lack of skew symmetry between the second and third). Similarly, the kernel of the symmetrization map

$$S^2 A \otimes A \to S^3 A$$

is a $GL(A)$-submodule.

Call these kernels K_Λ and K_S. We have a decomposition

$$A^{\otimes 3} = S^3A \oplus \Lambda^3 A \oplus K_\Lambda \oplus K_S.$$

This decomposition is $GL(A)$-invariant by Schur's lemma, since both K_Λ, K_S are kernels of $GL(A)$-module maps, but it is not canonical. In fact, K_Λ and K_S are isomorphic as $GL(A)$-modules. Their isomorphism class is denoted $S_{21}A$, and there is a canonical decomposition

$$A^{\otimes 3} = S^dA \oplus (S_{21}A)^{\oplus 2} \oplus \Lambda^3 A. \tag{4.5.1}$$

as a $GL(A)$-module. For the complete story see §8.7.1.

It is easy to see that the cyclic $\mathbb{Z}_3$ acts non-trivially on $K_\Lambda \oplus K_S$. It is slightly harder to see that in fact there is no subspace of $K_\Lambda \oplus K_S$ that is acted on trivially, see Exercise 8.7.2.4.

In summary

Proposition 4.5.0.2. *Let $\mathbb{Z}_3 \subset \mathfrak{S}_3$ act on $A^{\otimes 3}$ by cyclically permuting factors. Then*

$$(A^{\otimes 3})^{\mathbb{Z}_3} = S^3 A \oplus \Lambda^3 A.$$

Thus if we are searching for cyclic $\mathbb{Z}_3$-invariant decompositions for $M_{\langle \mathbf{n} \rangle}$, the size of our search space is cut down from $\mathbf{n}^6$ dimensions to $\frac{\mathbf{n}^6 + 2\mathbf{n}^2}{3}$ dimensions.

It is easy to write down the decomposition of $M_{\langle \mathbf{n} \rangle} \in S^3 A \oplus \Lambda^3 A$ into its symmetric and skew-symmetric components: $\mathrm{trace}(XYZ) = \frac{1}{2}[\mathrm{trace}(XYZ) + \mathrm{trace}(YXZ)] + \frac{1}{2}[\mathrm{trace}(XYZ) - \mathrm{trace}(YXZ)]$.

Exercise 4.5.0.3: (1) Verify that the first term in brackets lives in $S^3 A$ and second lives in $\Lambda^3 A$.

4.6. Invariants associated to a decomposition of $M_{\langle \mathbf{n} \rangle}$

Given two decompositions of $M_{\langle \mathbf{n} \rangle}$, how can we determine if they are in the same family? Given one, how can we determine its symmetry group? These questions are related, as a necessary condition for two decompositions to be in the same family is that they have isomorphic symmetry groups. We have already seen the invariants $\mathcal{S}_{s,t,u}$. I describe further invariants associated to a decomposition via graphs. I then discuss the points of $\mathcal{S}_U, \mathcal{S}_U^*$ in more detail: it turns out that the collection of points themselves has geometry that is also useful for distinguishing decompositions and determining symmetry groups.

4.6.1. Two graphs. Define a bipartite graph $\mathcal{IG}_{\mathcal{S}}$, the *incidence graph* where the top vertex set is given by elements in $\mathcal{S}_{U^*}$ and the bottom vertex set by elements in $\mathcal{S}_U$. Draw an edge between elements $[\mu]$ and $[v]$ if they are *incident*, i.e. $\mu(v) = 0$. Geometrically, $[v]$ belongs to the hyperplane determined by $[\mu]$ (and vice-versa). One can weight the vertices of this graph in several ways, the simplest (and in practice this has been enough) is just by the number of times the element appears in the decomposition. Let $\Gamma_{\mathcal{IG}_{\mathcal{S}}}$ denote the automorphism group of $\mathcal{IG}_{\mathcal{S}}$. If $\Gamma_{\mathcal{S}}$ is determined by its action on $\mathcal{S}_{1,1,1}$, we have $\Gamma_{\mathcal{S}} \subseteq \Gamma_{\mathcal{IG}_{\mathcal{S}}}$.

If a decomposition is $\mathbb{Z}_3$ invariant, or more precisely, if the three spaces have been identified by some automorphism of the decomposition, the incidence graphs form V, V^* and from W, W^* are isomorphic, and otherwise they give additional information.

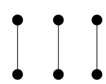
Given a $\mathbb{Z}_3$-invariant decomposition, a necessary condition for it to also have a transpose-like $\mathbb{Z}_2$ symmetry is that there is an isomorphism of the bipartite graph swapping the sets of (weighted) vertices.

In practice (see the examples below) the incidence graph has been enough to determine the symmetry group $\Gamma_{\mathcal{S}}$, in the sense that it cuts the possible size of the group down and it becomes straight-forward to verify that everything that can be in the group after this cut actually is there.

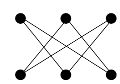
If we are only interested in automorphisms in $\Gamma'_{\mathcal{S}}$ that come from a diagonal $PGL_{\mathbf{n}} \subset PGL(U) \times PGL(V) \times PGL(W)$, we may define a second bipartite graph $\mathcal{PG}_{\mathcal{S}}$, the *pairing graph*, where the upper vertices are the points of $\mathcal{S}_{U^*}$ and the lower the points of $\mathcal{S}_V$, and one draws an edge between $[\mu]$ and $[v]$ if $\mu \otimes v$ appears in the decomposition. One can weight the edge by the number of times it appears.

As is clear from this discussion, one can continue labeling and coloring to get more and more refined information about the decomposition.

For Strassen's decomposition, these graphs are not so interesting:



incidence



pairing

4.6.2. Configurations of points in projective space. In practice, perhaps because of the numerical methods used, the sets $\mathcal{S}_U$, and $\mathcal{S}_{U^*}$ have been relatively small. It is not surprising that they each are spanning sets. Usually they have come from *configurations* in a sense I now describe. For $\mathbb{P}^1$, a configuration is simply a triple of points and the triple of points they determine in the dual vector space. For example Strassen's decomposition is built from a configuration. The higher dimensional analog of such pairs of triples is more complicated.

I emphasize that the decompositions of [**BILR**] were found by numerical searches, without distinguishing any configurations. However in most cases, we were able to give a simple description of the vectors appearing in the decomposition in terms of a configuration. This bodes well for future work.

I restrict the discussion to $\mathbb{P}^2$. The group PGL_3 acts simply transitively on the set of 4-ples of points in general linear position (i.e. such that any three of them span $\mathbb{P}^2$).

Start with any 4-ple of points in general linear position. In the decomposition, actual vectors will appear. Even in the decomposition, since what will appear are vectors tensored with each other, there is only a "global scale" for each term. Take the simplest (to write down) 4-ple, choosing the fourth vector in order to have the linear relation $u_1 + u_1 + u_3 + u_4 = 0$. I'll call this the *default configuration.* That is, the default configuration starts with

$$u_1 = \begin{pmatrix}1\\0\\0\end{pmatrix},\ u_2 = \begin{pmatrix}0\\1\\0\end{pmatrix},\ u_3 = \begin{pmatrix}0\\0\\1\end{pmatrix},\ u_4 = \begin{pmatrix}-1\\-1\\-1\end{pmatrix}.$$

The $\{[u_j]\}$ determine points in the dual space by taking pairwise intersections of the lines (hyperplanes) that they determine in $\mathbb{P}U^*$.

$$v_{12} = (0,0,1),\quad v_{13} = (0,1,0),\quad v_{14} = (0,1,-1),$$
$$v_{23} = (-1,0,0),\quad v_{24} = (-1,0,1),\quad v_{34} = (1,-1,0).$$

Here $[v_{ij}]$ is the line in $\mathbb{P}^2$ (considered as a point in the dual space $\mathbb{P}^{2*}$) through the points $[u_i]$ and $[u_j]$ in $\mathbb{P}^2$ (or dually, the point of intersection of the two lines $[u_i]$, $[u_j]$ in $\mathbb{P}^{2*}$). Note that here choices of representatives are being made. I have made choices that will be useful for the decomposition $\mathcal{S}_{BILR,\mathbb{Z}_4\times\mathbb{Z}_3}$ of §4.7.1 below.

The $v_{i,j}$ in turn determine their new points of intersection:

$$u_{12,34} = \begin{pmatrix}1\\1\\0\end{pmatrix},\ u_{13,24} = \begin{pmatrix}1\\0\\1\end{pmatrix},\ u_{14,23} = \begin{pmatrix}0\\1\\1\end{pmatrix}.$$

which determine new points

$$v_{(12,34),(13,24)} = (-1,1,1),\ v_{(12,34),(14,23)} = (1,-1,1),\ v_{(13,24),(14,23)} = (1,1,-1),$$

which determine

$$u_{34,(13,24|14,23)} = \begin{pmatrix}1\\1\\2\end{pmatrix},\ u_{24,(12,34|14,23)} = \begin{pmatrix}1\\2\\1\end{pmatrix},\ u_{23,(12,34|13,24)} = \begin{pmatrix}2\\1\\1\end{pmatrix},$$

$$u_{12,(12,34|13,24)} = \begin{pmatrix}1\\1\\0\end{pmatrix},\ u_{13,(12,34|13,24)} = \begin{pmatrix}1\\0\\1\end{pmatrix},\ u_{14,(12,34|13,24)} = \begin{pmatrix}0\\1\\1\end{pmatrix}.$$

This process continues, but in practice only vectors from the first few rounds appeared in decompositions.

Of course any other choices of initial points leads to an equally good configuration. The decompositions we found initially produced different configurations and we converted them to the standard ones for convenience.

Sometimes there was more than one way to label the points in terms of a configuration. I remark on this more below.

4.7. Cyclic $\mathbb{Z}_3$-invariant rank 23 decompositions of $M_{\langle 3\rangle}$

The following examples are from [**BILR**].

4.7.1. A rank 23 decomposition of $M_{\langle 3\rangle}$ with $\mathbb{Z}_4\times\mathbb{Z}_3$ symmetry. Take a configuration and let $a_0 : U \to V$ send u_j to v_{j+1}. In the default configuration

$$a_0 = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$$

corresponds to the generator of $\mathbb{Z}_4$ that cyclically permutes indices.

$$a_0^{\otimes 3} \tag{4.7.1}$$
$$(u_{24}v_{12|34})^{\otimes 3} \tag{4.7.2}$$
$$(u_{13}v_{14|23})^{\otimes 3} \tag{4.7.3}$$
$$(u_{34}v_1)^{\otimes 3} \tag{4.7.4}$$
$$(u_{14}v_2)^{\otimes 3} \tag{4.7.5}$$
$$(u_{12}v_3)^{\otimes 3} \tag{4.7.6}$$
$$(u_{23}v_4)^{\otimes 3} \tag{4.7.7}$$
$$-[u_{24}v_4 + u_{12}v_3]^{\otimes 3} \tag{4.7.8}$$
$$-[u_{13}v_3 + u_{14}v_2]^{\otimes 3} \tag{4.7.9}$$
$$-[u_{24}v_2 + u_{34}v_1]^{\otimes 3} \tag{4.7.10}$$
$$-[u_{13}v_1 + u_{23}v_4]^{\otimes 3} \tag{4.7.11}$$
$$\langle (u_{23}v_2)\otimes(u_{34}v_4)\otimes(u_{13}v_1)\rangle_{\mathbb{Z}_3} \tag{4.7.12}$$
$$\langle (u_{34}v_3)\otimes(u_{14}v_1)\otimes(u_{24}v_2)\rangle_{\mathbb{Z}_3} \tag{4.7.13}$$
$$\langle (u_{14}v_4)\otimes(u_{12}v_2)\otimes(u_{13}v_3)\rangle_{\mathbb{Z}_3} \tag{4.7.14}$$
$$\langle (u_{12}v_1)\otimes(u_{23}v_3)\otimes(u_{24}v_4)\rangle_{\mathbb{Z}_3}. \tag{4.7.15}$$

Exercise 4.7.1.1: (2) Verify that this is indeed a decomposition of $M_{\langle 3\rangle}$.

The $\mathbb{Z}_4$-invariance of the decomposition may be seen as follows: Since $a_0(a_0){a_0}^{-1} = a_0$, the first term is $\mathbb{Z}_4$-invariant. Then conjugation by a_0 swaps (4.7.2) and (4.7.3), it cyclically permutes ((4.7.4), (4.7.7),(4.7.6),(4.7.5)), as well as ((4.7.8), (4.7.11),(4.7.10),(4.7.9)) and ((4.7.12), (4.7.13),(4.7.14),(4.7.15)). All this is transparent from the presentation.

It is better to write the decomposition as

$$a_0^{\otimes 3} \tag{4.7.16}$$

$$\langle (u_{24}v_{12|34})^{\otimes 3}\rangle_{\mathbb{Z}_2\subset\mathbb{Z}_4} \tag{4.7.17}$$

$$\langle -[u_{24}v_4+u_{12}v_3]^{\otimes 3}\rangle_{\mathbb{Z}_4} \tag{4.7.18}$$

$$\langle (u_{12}v_3)^{\otimes 3}\rangle_{\mathbb{Z}_4} \tag{4.7.19}$$

$$\langle (u_{12}v_1)\otimes(u_{23}v_3)\otimes(u_{24}v_4)\rangle_{\mathbb{Z}_4\times\mathbb{Z}_3}. \tag{4.7.20}$$

Note also that since the $\mathbb{Z}_3$ is purely external and the $\mathbb{Z}_4$ purely internal, the symmetry group transparently contains $\mathbb{Z}_4\times\mathbb{Z}_3$.

Given the distribution of the frequencies of the points: $(4,4,4,4,1,1)$ in V, $(3,3,3,3,3,3)$ in U^*, a transpose-like symmetry is not possible. Moreover, it is clear one cannot upgrade the $\mathbb{Z}_4$ to $\mathfrak{S}_4$ since only two of the three $v_{ij|kl}$ appear in the decomposition: $v_{12|34}, v_{14|23}$ ($v_{13|24}$ is omitted). So, e.g. the transposition $(2,3)$ takes $\mathcal{S}_{BILR,\mathbb{Z}_4\times\mathbb{Z}_3}$ to a different decomposition in the family.

Thus the symmetry group of $\mathcal{S}_{BILR,\mathbb{Z}_4\times\mathbb{Z}_3}$ is indeed $\mathbb{Z}_4\times\mathbb{Z}_3$ **is this a complete proof??**

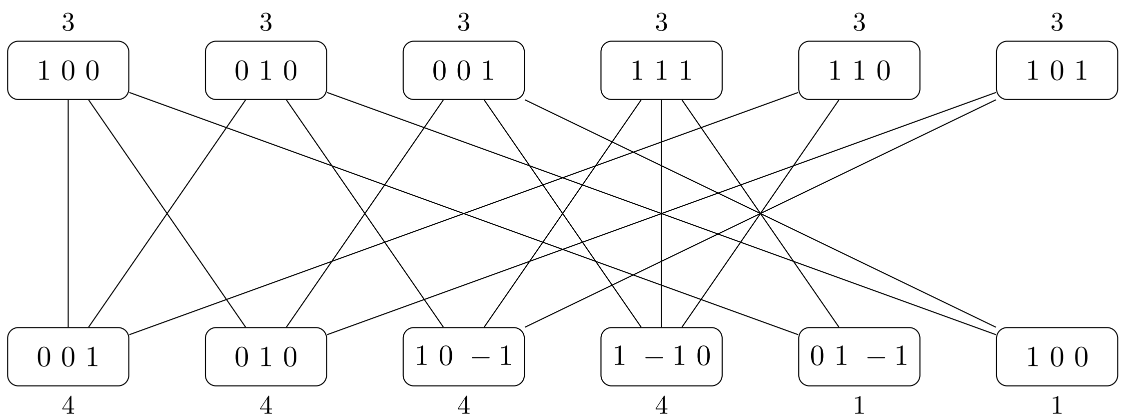
This default configuration has the added benefit that when we have a $\mathbb{Z}_4$-invariant decomposition, the realization of $\mathbb{Z}_4\subset GL_3$ will be the standard one. The choice of scale was made so that $v_{i,i+1}(u_{i+2})=1$, $v_{i,i+1}(u_{i+3})=-1$ (indices considered mod four). This has the advantage of $v_{i,i+1}=\tau v_{12}$ where $\tau\in\mathbb{Z}_4$ is the generator of the standard $\mathbb{Z}_4$. For v_{13} there was no obvious choice of sign, but then we chose $v_{24}=\tau(v_{13})$.

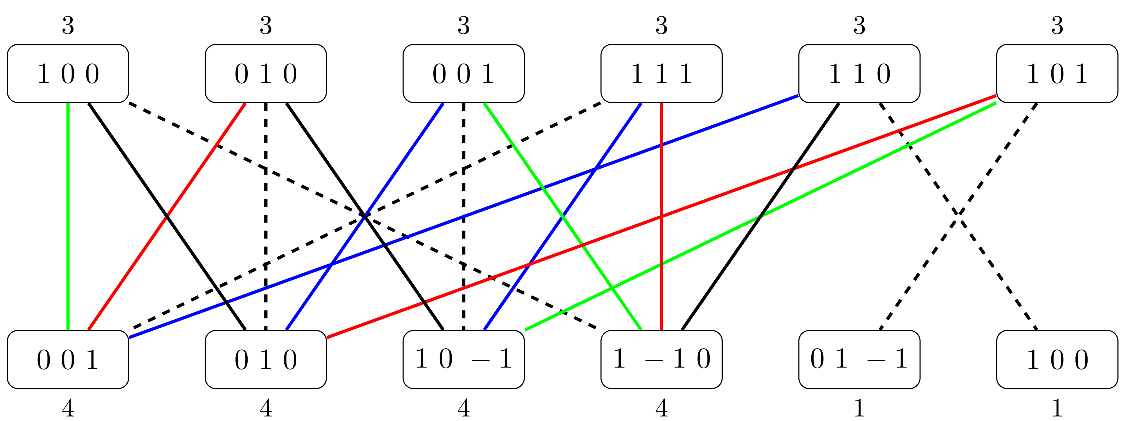
For standard $\mathbb{Z}_4$-invariant decompositions, notice that these vectors split into two $\mathbb{Z}_4$-orbits: the $v_{i,i+1}$'s which consist of four vectors, and the $v_{i,i+2}$'s of which there are two.

Let $v_1,\dots,v_4\in U$ be a configuration, determining $u_{ij}\in U^*$, determining further $v_{ij|kl}\in U$ where $[u_{ij}]=v_i{}^{\perp}\cap v_j{}^{\perp}$ and $[v_{ij|kl}]=u_{ij}{}^{\perp}\cap u_{jl}{}^{\perp}$. It turns out only two of the $v_{ij|kl}$ appear in the decomposition, say $v_{12|34}, v_{14|23}$, so $v_{13|24}$ is omitted. Each v_i appears in exactly 4 rank one terms, each u_{ij} appears in exactly three, and $v_{12|34}, v_{14|23}$ appear in one each. Because the frequencies are different, we cannot have a transpose-like symmetry so $\Gamma\subset(GL(U)\times GL(V)\times GL(W))\rtimes\mathbb{Z}_3$ because we cannot swap the u's and v's. Further, because of the incidence relations, letting $\Gamma'=\Gamma\cap GL(U)\times GL(V)\times GL(W)$, Γ' is determined by its action on the points in $\mathbb{P}U$, and in fact on the points in the initial configuration, which says $\Gamma'\subset\mathfrak{S}_4$. But to have the intersections preserved Γ' must also preserve the pair $\{v_{12|34}, v_{14|23}\}$, i.e., $\Gamma'\subseteq\mathbb{Z}_4$, generated by the cycle (1234). (Note that (1234) swaps $v_{12|34}$ and $v_{14|23}$, while $(1234)^2=(12)(34)$ preserves each of them.)

Thus $\Gamma \subseteq \mathbb{Z}_4 \rtimes \mathbb{Z}_3$.

Note that in the presentation, that (4.7.12)-(4.7.15) are nilpotent.





4.7.2. Laderman's decomposition. I now discuss Laderman's rank 23 decomposition of $M_{\langle 3\rangle}$, which I denote $\mathcal{L}ad$ According to Burichenko [**Bur15**], one has a $\mathbb{Z}_2 \times \mathbb{Z}_2 \subset SL(U) \times SL(V) \times SL(W)$ contained in $\Gamma_{\mathcal{L}ad}$ and the full cyclic permutation and transpose $\mathbb{Z}_3 \rtimes \mathbb{Z}_2$ also in $\Gamma_{\mathcal{L}ad}$, acting in a twisted way. Thus in the family generated by the decomposition, there is a standard $\mathbb{Z}_3$ invariant decomposition. Thanks to the transpose symmetry, it is better to label points in the dual space by their image under transpose rather than annihilators, to make the transpose-like symmetry more transparent. Here it is:

Points:

$$u_1 = \begin{pmatrix}1\\0\\0\end{pmatrix},\ u_2 = \begin{pmatrix}0\\1\\0\end{pmatrix},\ u_3 = \begin{pmatrix}0\\0\\1\end{pmatrix},\ u_{12} = \begin{pmatrix}1\\-1\\0\end{pmatrix},\ u_{23} = \begin{pmatrix}0\\1\\-1\end{pmatrix}.$$

$$v_1 = (1,0,0),\ v_2 = (0,1,0),\ v_3 = (0,0,1),$$
$$v_{12} = (1,1,0),\ v_{23} = (0,1,1).$$

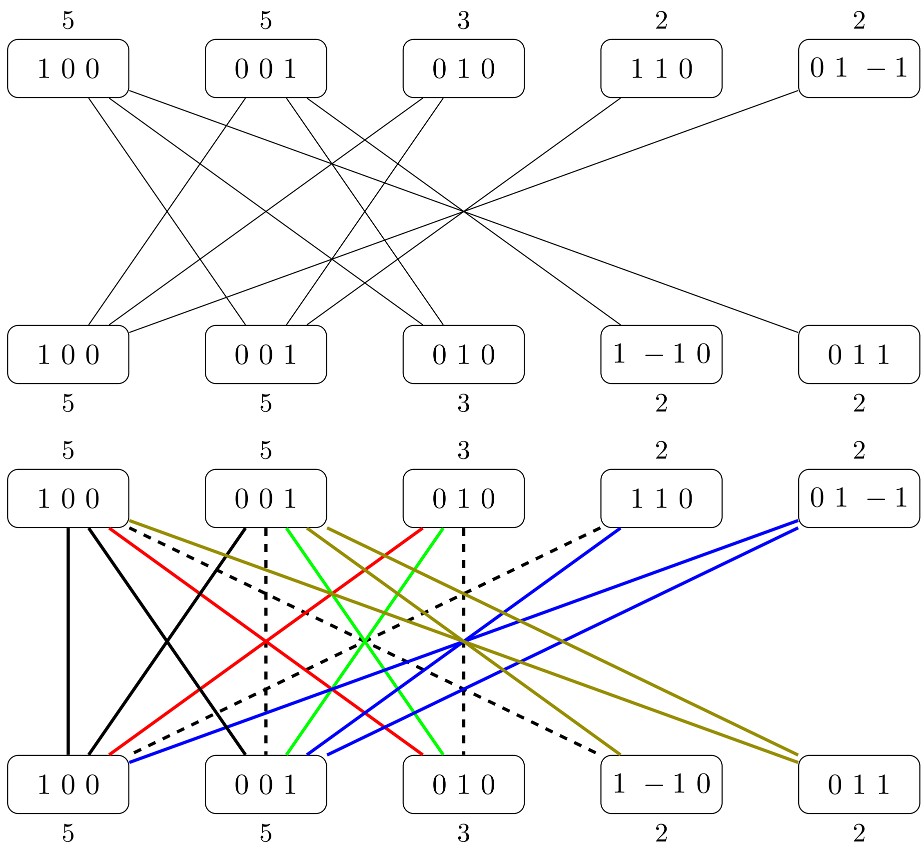
$$(u_2v_2)^{\otimes 3} \tag{4.7.21}$$
$$(u_3v_3)^{\otimes 3} \tag{4.7.22}$$
$$(u_{12}v_1)^{\otimes 3} \tag{4.7.23}$$
$$(u_1v_{12})^{\otimes 3} \tag{4.7.24}$$
$$(u_2v_1 - u_1v_{12})^{\otimes 3} \tag{4.7.25}$$
$$\langle (u_1v_3)\otimes(u_3v_1)\otimes(u_1v_1)\rangle_{\mathbb{Z}_3} \tag{4.7.26}$$
$$\langle (u_{23}v_1)\otimes(u_{12}v_3)\otimes(u_{23}v_3)\rangle_{\mathbb{Z}_3} \tag{4.7.27}$$
$$\langle (u_3v_{12})\otimes(u_1v_{23})\otimes(u_3v_{23})\rangle_{\mathbb{Z}_3} \tag{4.7.28}$$
$$\langle (u_2v_3 - u_{23}v_1)\otimes(u_1v_2 - u_{12}v_3)\otimes(u_3v_2 - u_{23}v_3)\rangle_{\mathbb{Z}_3} \tag{4.7.29}$$
$$\langle (u_{23}v_{12} + u_2v_3 - u_1v_{23})\otimes(u_2v_3)\otimes(u_3v_2)\rangle_{\mathbb{Z}_3} \tag{4.7.30}$$
$$\langle (u_{12}v_{12} + u_2v_3 - u_3v_2)\otimes(u_2v_1)\otimes(u_1v_2)\rangle_{\mathbb{Z}_3} \tag{4.7.31}$$

Exercise 4.7.2.1: (2) Verify that this indeed is a decomposition of $M_{\langle 3\rangle}$.

The transpose-like $\mathbb{Z}_2$ is $x\otimes y\otimes z \mapsto (\epsilon_2 y\epsilon_2)^T\otimes(\epsilon_2 x\epsilon_2)^T\otimes(\epsilon_2 z\epsilon_2)^T$, where $\epsilon_2 = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1\end{pmatrix}$. (Note the similarities with Strassen's decomposition.) In other words send $u_1 \leftrightarrow v_1$, $u_2 \leftrightarrow -v_2$, $u_3 \leftrightarrow v_3$ and then switch the first two factors in $A\otimes B\otimes C$. This action fixes all terms except it performs the exchanges (4.7.23) $\leftrightarrow$ (4.7.24) and (4.7.27) $\leftrightarrow$ (4.7.28).

Now the $\mathbb{Z}_2\times\mathbb{Z}_2$ unfortunately is not in the diagonal GL_9, as one must act differently on each of A, B, C. I.e, instead of being in GL_9 as I had hoped, it is in $GL_9\times GL_9\times GL_9$. In other words, it does not respect the above structure: For example, his Φ_1 converts (4.7.23) to the second summand in (4.7.27), and that second summand becomes the $\mathbb{Z}_3$-invariant term (4.7.23). Thus the decomposition as a whole is unchanged, but $\mathbb{Z}_3$-invariant and non-$\mathbb{Z}_3$-invariant terms are mixed. This explains why Nick found more than one solution: there are (at least) four distinct ways to convert Laderman to a standard $\mathbb{Z}_3$-invariant decomposition, and each of the four leads to the same standard $\mathbb{Z}_3$-invariant decomposition.

Exercise 4.7.2.2: (2) Verify the asserted automorphisms for Laderman's decomposition



4.8. Secant varieties and additional geometric language

At this point, it will be useful to introduce the additional geometric language of *secant varieties* that will enable us to discuss rank decompositions in a larger context and analyze border rank decompositions.

Secant varieties will also arise naturally in the study of Valiant's conjecture and its variants, so even as far as complexity theory it is worth discussing border rank from the larger perspective of secant varieties.

4.8.1. Secant Varieties. In order to better study σ_r, which governs the complexity of $M_{\langle \mathbf{n} \rangle}$, it will be useful to place the study in the broader context of *secant varieties*, an extensively studied class of varieties.

Given a variety $X \subset \mathbb{P}V$, define the *X-rank* of $[p] \in \mathbb{P}V$, $\mathbf{R}_X([p])$, to be the smallest r such that there exist $x_1, \ldots, x_r \in \hat{X}$ such that p is in the span of $x_1, \ldots, x_r$, and the *X-border rank* $\underline{\mathbf{R}}_X([p])$ is defined to be the smallest r such that there exist curves $x_1(t), \ldots, x_r(t) \in \hat{X}$ such that p is in the span of the limiting plane $\lim_{t\to 0}\langle x_1(t), \ldots, x_r(t)\rangle$, where

$\langle x_1(t), \ldots, x_r(t)\rangle \subset G(r, V)$ is viewed as a curve the Grassmannian. Here and in what follows, I am assuming that for $t \neq 0$, $x_1(t), \ldots, x_r(t)$ are linearly independent (otherwise we are really dealing with a decomposition of lower border rank).

Let $\sigma_r(X) \subset \mathbb{P}V$ denote the set of points of X-border rank at most r, called the *r-th secant variety of X*. (Theorem 3.1.6.1 assures us that $\sigma_r(X)$ is indeed a variety.) In other words

$$\sigma_r(X) = \overline{\bigcup_{x_1,\ldots,x_r \in X} \langle x_1, \ldots, x_r\rangle}$$

where $\langle x_1, \ldots, x_r\rangle$ denotes the linear span in projective space. The notation is such that $\sigma_1(X) = X$. When $X = Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)$ is the set of rank one tensors, $\sigma_r(X) = \sigma_r$.

Let $X \subset \mathbb{P}V$ be a smooth variety, and let $p \in \sigma_2(X)$. If p is not a point of X, nor a point on an honest secant line, then p must line on some tangent line to X, where here I take the naïve definition of tangent line, namely a point on a limit of secant lines.

Terracini's lemma (see, e.g., [**Lan12**, §5.3]) generalizes our caculation of $\hat{T}_{[a_1\otimes b_1\otimes c_1 + a_2\otimes b_2\otimes c_2]} Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ of §3.1.3: if $z = [x_1 + \cdots + x_r]$ with $[x_j] \in X$ general points, then $\hat{T}_z\sigma_r(X) = \sum_{j=1}^r \hat{T}_{[x_j]}X$. In particular $\dim \sigma_r(X) \leq r \dim X + r - 1$.

Thus $\dim \sigma_r(X) \leq \min\{r \dim X + r - 1, \mathbf{v} - 1\}$, and when equality holds we will say $\sigma_r(X)$ is of the *expected dimension*. The expected dimension is indeed what occurs "most" of the time. For example, $\dim \sigma_r(\mathbb{P}^N \times \mathbb{P}^N \times \mathbb{P}^N)$ is the expected dimension for all (r, N) except $(r, N) = (4, 2)$ [**Lic85**].

4.8.2. Homogeneous varieties, orbit closures, and G-varieties. The Segre, Veronese and Grassmannian are examples of *homogeneous varieties*:

Definition 4.8.2.1. A subvariety $X \subset \mathbb{P}V$, is *homogeneous* if it is a closed orbit of some point $x \in \mathbb{P}V$ under the action of some group $G \subset GL(V)$. If $P \subset G$ is the subgroup fixing x, we write $X = G/P$.

A variety $X \subset \mathbb{P}V$ is called a *G-variety* for a group $G \subset GL(V)$, if for all $g \in G$ and $x \in X$, $g \cdot x \in X$.

Orbit closures (see §3.3.1) and homogeneous varieties are G-varieties.

Exercise 4.8.2.2: (1) What are the points in $\overline{GL_n \cdot (x_1 \cdots x_n)}$ that are not in $GL_n \cdot (x_1 \cdots x_n)$?

4.8.3. The abstract secant variety. I now construct a variety that will facilitate the study of decompositions of a tensor. I make the construction in the more general context of secant varieties.

Let $X \subset \mathbb{P}V$ be a variety. Consider the set

$$S_r(X)^0 := \{(x_1, \dots, x_r, z) \in X^{\times r} \times \mathbb{P}V \mid z \in \operatorname{span}\{x_1, \dots, x_r\}\} \subset Seg(X^{\times r} \times \mathbb{P}V) \subset \mathbb{P}V^{\otimes r+1}$$

and let $S_r(X) := \overline{S_r(X)^0}$ denote its Zariski closure. (For those familiar with quotients, it would be more convenient to deal with $X^{(\times r)} := X^{\times r}/\mathfrak{S}_r$.) We have a map $\pi^0 : S_r(X)^0 \to \mathbb{P}V$, extending to a map $\pi : S_r(X) \to \mathbb{P}V$, given by projection onto the last factor and the image is $\sigma_r^0(X)$ (resp. $\sigma_r(X)$). We will call $S_r(X)$ the *abstract r-th secant variety of X*. As long as $r < \mathbf{v}$ and X is not contained in a linear subspace of $\mathbb{P}V$, $\dim S_r(X) = r \dim X + r - 1$ because $\dim X^{\times r} = r \dim X$ and a general set of r points on X will span a $\mathbb{P}^{r-1}$.

If $\sigma_r(X)$ is of the expected dimension, so its dimension equals that of $S_r(X)$, then for general points $z \in \sigma_r(X)^0$, $(\pi^0)^{-1}(z)$ will consist of a finite number of points and each point will correspond to a decomposition $\overline{z} = \overline{x_1} + \cdots + \overline{x_r}$ for $\overline{x_j} \in \hat{x}_j$, $\overline{z} \in \hat{z}$. In summary:

Proposition 4.8.3.1. *If $X^n \subset \mathbb{P}^N$ and $\sigma_r(X)$ is of (the expected) dimension $rn + r - 1 < N$, then a Zariski open subset of points on $\sigma_r(X)$ have a finite number of decompositions into a sum of r elements of X.*

If the fiber of π^0 over $z \in \sigma_r^0(X)$ is k-dimensional, then there is a k-parameter family of decompositions of z as a sum of r rank one tensors. This occurs, for example if $z \in \sigma_{r-1}(X)$, but it can also occur for points in $\sigma_r(X) \backslash \sigma_{r-1}(X)$. We have seen that this is indeed the case for $M_{\langle 2,2,2\rangle} \in \sigma_7(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$.

If X is a G-variety, then $\sigma_r(X)$ is also a G-variety, and if $z \in \sigma_r^0(X)$ is fixed by $G_z \subset G$, then G_z will act (possibly trivially) on $(\pi^0)^{-1}(z)$, and every distinct (up to re-ordering if one is not working with $X^{(\times r)}$) point in its orbit will correspond to a distinct decomposition of z. Let $q \in (\pi^0)^{-1}(x)$. If $\dim(G_z \cdot q) = d_z$, then there is at least a d_z parameter family of decompositions of z as a sum of r elements of X. We have seen that in the case of $X = Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$, if z is concise, then $d_z = \dim G_z$.

Remark 4.8.3.2. Note that $\operatorname{codim}(S_{r-1}(X), S_r(X)) \leq \dim X - 1$, where the inclusion is just by adding a point to a border rank $r-1$ decomposition. In particular, in the case of the Segre relevant for matrix multiplication, this codimension is at most $3(\mathbf{n}^2 - 1)$. On the other hand Image $G_{M_{\langle \mathbf{n}\rangle}} = 3(\mathbf{n}^2 - 1)$, so by a dimension count, one might "expect" $\pi_r^{-1}(M_{\langle \mathbf{n}\rangle})$ to intersect $S_{r-1}(X)$, meaning that we could keep reducing the border rank of $M_{\langle \mathbf{n}\rangle}$ all the way down to one. Of course since $S_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is not a projective space, no intersection is implied, but this dimension count just illustrates the pathology of the tensor $M_{\langle \mathbf{n}\rangle}$.

4.8.4. What is a border rank decomposition? Usually an X-border rank decomposition of some $v \in V$ is presented as $v = \lim_{t\to 0}(x_1(t) + \cdots + x_r(t))$ where $[x_j(t)]$ are curves in X. In order to discuss border rank decompositions geometrically, it will be useful to study the corresponding curve in the Grassmannian $\langle x_1(t), \ldots, x_r(t)\rangle \subset G(r, V)$. The limiting r plane that contains v will have several geometric aspects, in particular the geometry of its intersection with X.

To better understand this geometry, consider

$$\tilde{S}^0_r(X) := \{([v], ([x_1], \ldots, [x_r]), E) \mid v \in \langle x_1, \ldots, x_r\rangle \subseteq E\} \subset \mathbb{P}V \times X^{\times r} \times G(r, V)$$

and $\tilde{S}_r(X) := \overline{\tilde{S}^0_r(X)}$.

We can stratify $\sigma_r(X)$ and $\tilde{S}_r(X)$ by the h's of the intermediate ranks $\mathbf{R}_h$ of §3.2.1. The case $h = 0$ is rank. The next case $h = 1$ has a straight-forward geometry.

To understand the $h = 1$ case, first consider the case $r = 2$, so $v = \lim_{t\to 0} \frac{1}{t}(x_1(t)+x_2(t))$ for curves $[x_j(t)] \subset X$. Then we must have $\lim_{t\to 0}[x_1(t)] = \lim_{t\to 0}[x_2(t)]$ and if this limiting point is $[x]$, we obtain an element of $\hat{T}_x X$. In the case of $\sigma_r(X)$, one needs r curves such that the points are linearly independent for $t \neq 0$ and such that they become dependent when $t = 0$. This is most interesting when no subset of $r - 1$ points becomes linearly dependent. Then it is not hard to see (see [**Lan12**, §10.8.1], that one may obtain an arbitrary point of $\hat{T}_{x_1} X + \cdots + \hat{T}_{x_r} X$. For some varieties there may not exist r distinct points on them that are linearly dependent (e.g., $v_d(\mathbb{P}^1)$ when $d > r$). An easy way for such sets of points to exist is if there is a $\mathbb{P}^{r-1}$ on the variety. The decompositions for $M^{red}_{\langle \mathbf{m},2,2\rangle}$ I discuss in the next section are not quite from such simple configurations, but nearly are. Because of this I next discuss the geometry of linear spaces on the Segre.

4.8.5. Lines on Segre varieties. There are three types of lines on $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$: α-lines, which are of the form $\mathbb{P}(\langle a_1, a_2\rangle \otimes b \otimes c)$ for some $a_j \in A$, $b \in B$, $c \in C$, and the other two types are defined similarly and called β and γ lines.

Exercise 4.8.5.1: (2) Show that all lines on $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ are one of these types.

Given two lines $L_\beta, L_\gamma \subset Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ respectively of type β, γ, if they do not intersect, then $\langle L_\beta, L_\gamma\rangle = \mathbb{P}^3$ and if the lines are general, furthermore $\langle L_\beta, L_\gamma\rangle \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) = L_\beta \sqcup L_\gamma$.

However if $L_\beta = \mathbb{P}(a \otimes \langle b_1, b_2\rangle \otimes c)$ and $L_\gamma = \mathbb{P}(a' \otimes b \otimes \langle c_1, c_2\rangle)$ with $b \in \langle b_1, b_2\rangle$ and $c \in \langle c_1, c_2\rangle$, then they still span a $\mathbb{P}^3$ but $\langle L_\beta, L_\gamma\rangle \cap Seg(\mathbb{P}A \times$

$\mathbb{P}B \times \mathbb{P}C) = L_\beta \sqcup L_\gamma \sqcup L_\alpha$, where $L_\alpha = \mathbb{P}(\langle a, a'\rangle \otimes b \otimes c)$, and L_α intersects both L_β and L_γ.

Let $x, y, z \in Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ be distinct points that all lie on a line $L \subset Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$. Then

$$\hat{T}_x Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) \subset \langle \hat{T}_y Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C), \hat{T}_z Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)\rangle. \tag{4.8.1}$$

In fact, the analogous statement is true for lines on any cominuscule variety, see [**BL14**, Lemma 3.3]. Because of this, it will be more geometrical to refer to $\hat{T}_L Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) := \langle \hat{T}_y Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C), \hat{T}_z Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)\rangle$, as the choice of $y, z \in L$ is irrelevant.

Exercise 4.8.5.2: (1) Verify (4.8.1).

The matrix multiplication tensor $M_{\langle U,V,W\rangle}$ endows A, B, C with additional structure, e.g., $B = V^* \otimes W$, so there are two types of distinguished β-lines (corresponding to lines of rank one matrices), call them (β, ν^*)-lines and (β, ω)-lines, where, e.g., a ν^*-line is of the form $\mathbb{P}(a \otimes (\langle v^1, v^2\rangle \otimes w) \otimes c)$, and among such lines there are further distinguished ones where moreover both a and c also have rank one. Call such further distinguished lines *special* (β, ν^*)-lines.

4.9. Border rank decompositions

4.9.1. $M^{red}_{\langle 2\rangle}$. Here $A \subset U^* \otimes V$ has dimension three.

What follows is a slight modification of the decomposition of $M^{red}_{\langle 2\rangle}$ from [**BCRL79**] that appeared in [**LR0**]. Call it the $BCLR$-decomposition. I label the points such that x^1_1 is set equal to zero. The main difference is that in the original all five points moved, but here one is stationary.

$$\begin{aligned}
p_1(t) &= x^1_2 \otimes (y^2_2 + y^2_1) \otimes (z^2_2 + t z^1_1)\\
p_2(t) &= -(x^1_2 - t x^2_2) \otimes y^2_2 \otimes (z^2_2 + t(z^1_1 + z^2_1))\\
p_3(t) &= x^2_1 \otimes (y^2_1 + t y^1_2) \otimes (z^2_2 + z^1_2)\\
p_4(t) &= (x^2_1 - t x^2_2) \otimes (-y^2_1 + t(y^1_1 - y^1_2)) \otimes z^1_2\\
p_5(t) &= -(x^2_1 + x^1_2) \otimes y^2_1 \otimes z^2_2
\end{aligned}$$

and

$$M^{red}_{\langle 2\rangle} = \lim_{t\to 0} \frac{1}{t}[p_1(t) + \cdots + p_5(t)]. \tag{4.9.1}$$

Use the notation $x^i_j = u^i \otimes v_j$, $y^j_k = v^j \otimes w_k$ and $z^k_i = w^k \otimes u_i$.

Theorem 4.9.1.1. [**LR0**] *Let* $E^{BCLR} = \lim_{t\to 0}\langle p_1(t),\dots,p_5(t)\rangle \in G(5, A\otimes B\otimes C)$*. Then* $E^{BCLR}\cap Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C)$ *is the union of three lines:*

$$
\begin{aligned}
L_{12,(\beta,\omega)} &= x^1_2\otimes(v^2\otimes W)\otimes z^1_2\\
L_{21,(\gamma,\omega^*)} &= x^2_1\otimes y^2_2\otimes(W^*\otimes u_2)\\
L_\alpha &= \langle x^2_1, x^1_2\rangle\otimes y^2_2\otimes z^1_2.
\end{aligned}
$$

Here $L_{12,(\beta,\omega)}$ *is a special* (β,ω)*-line,* $L_{21,(\gamma,\omega^*)}$*, is a special* (γ,ω^*)*-line, and* L_α*, is an* α*-line with rank one* B *and* C *points. Moreover, the* C*-point of* $L_{12,(\beta,\omega)}$ *lies in the* ω^**-line of* $L_{21,(\gamma,\omega^*)}$*, the* B*-point of* $L_{21,(\gamma,\omega^*)}$ *lies in the* ω*-line of* $L_{12,(\beta,\omega)}$ *and* L_α *is the unique line on the Segre intersecting* $L_{12,(\beta,\omega)}$ *and* $L_{21,(\gamma,\omega^*)}$ *(and thus it is contained in their span).*

Furthermore, $E^{BCLR} = \langle M^{red}_{\langle 2\rangle}, L_{12,(\beta,\omega)}, L_{21,(\gamma,\omega^*)}\rangle$ *and*

$$
M^{red}_{\langle 2\rangle}\in\langle \hat{T}_{L_{12,(\beta,\omega)}}Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C), \hat{T}_{L_{21,(\gamma,\omega^*)}}Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C)\rangle.
$$

Proof. Write $p_j = p_j(0)$. Then (up to sign, which is irrelevant for geometric considerations)

$$
\begin{aligned}
p_1 &= x^1_2\otimes(y^2_2+y^2_1)\otimes z^2_2\\
p_2 &= x^1_2\otimes y^2_2\otimes z^2_2\\
p_3 &= x^2_1\otimes y^2_1\otimes(z^2_2+z^1_2)\\
p_4 &= x^2_1\otimes y^2_1\otimes z^1_2\\
p_5 &= (x^2_1+x^1_2)\otimes y^2_1\otimes z^2_2
\end{aligned}
$$

The configuration of lines is as follows:

$$
\begin{aligned}
L_{12,(\beta,\omega)} &= \langle p_1,p_2\rangle = x^1_2\otimes(v^2\otimes W)\otimes z^2_2\\
L_{21,(\gamma,\omega^*)} &= \langle p_3,p_4\rangle = x^2_1\otimes y^2_1\otimes(W^*\otimes u_2)\\
p_5\in L_\alpha &= \langle x^1_2, x^2_1\rangle\otimes y^2_1\otimes z^2_2.
\end{aligned}
$$

To see there are no other points in $E^{BCLR}\cap Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C)$, first note that any such point would have to lie on $Seg(\mathbb{P}\langle x^1_2,x^2_1\rangle\times\mathbb{P}\langle y^2_1,y^2_2\rangle\times\mathbb{P}\langle z^1_2,z^2_2\rangle)$ because there is no way to eliminate the rank two $x^2_2\otimes(y^2_1\otimes z^1_2+y^2_2\otimes z^2_2)$ term in $M^{red}_{\langle 2\rangle}$ with a linear combination of $p_1,\dots,p_4$. Let $[(sx^1_2+tx^2_1)\otimes(uy^2_2+vy^2_1)\otimes(pz^2_2+qz^1_2)]$ be an arbitrary point on this variety. To have it be in the span of $p_1,\dots,p_4$ it must satisfy the equations $suq=0$, $svq=0$, $tuq=0$, $tup=0$. Keeping in mind that one cannot have $(s,t)=(0,0)$, $(u,v)=(0,0)$, or $(p,q)=(0,0)$, we conclude the only solutions are the three lines already exhibited.

We have

$$\begin{aligned}
p_1(0)' &= x_2^1 \otimes (y_2^2 + y_1^2) \otimes z_1^1 \\
p_2(0)' &= x_2^2 \otimes y_2^2 \otimes z_2^2 - x_2^1 \otimes y_2^2 \otimes (-z_1^2 + z_1^1) \\
p_3(0)' &= x_1^2 \otimes y_2^1 \otimes (z_2^2 + z_2^1) \\
p_4(0)' &= x_2^2 \otimes y_1^2 \otimes z_2^1 + x_1^2 \otimes (y_1^1 - y_2^1) \otimes z_2^1 \\
p_5(0)' &= 0.
\end{aligned}$$

Then $M^{red}_{\langle 2\rangle} = (p_1' + p_2') + (p_3' + p_4')$ where $p_1' + p_2' \in T_{L_{12,(\beta,\omega)}} Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ and $p_3' + p_4' \in T_{L_{21,(\gamma,\omega^*)}} Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$. $\square$

Remark 4.9.1.2. By removing x_1^1 from our tensor, we lose the cyclic $\mathbb{Z}_3$-symmetry but retain a standard transpose action $x \otimes y \otimes z \mapsto x^T \otimes z^T \otimes y^T$. Similarly we lose our $GL(U) \times GL(V)$ symmetry but retain our $GL(W)$ action. By composing our standard transpose symmetry with another $\mathbb{Z}_2$ action which switches the basis vectors of W, the action swaps $p_1(t) + p_2(t)$ with $p_3(t) + p_4(t)$ and $L_{12,(\beta,\omega)}$ with $L_{21,(\gamma,\omega^*)}$. This action fixes p_5.

Remark 4.9.1.3. Note that it is important that p_5 lies neither on $L_{12,(\beta,\omega)}$ nor on $L_{21,(\gamma,\omega^*)}$, so that no subset of the five points lies in a linearly degenerate position to enable us to have tangent vectors coming from all five points, but I emphasize that any point on the line L_α not on the original lines would have worked equally well, so the geometric object is this configuration of lines.

4.9.2. $M^{red}_{\langle 3,2,2\rangle}$. Here is the decomposition in [**AS13**, Thm. 2] due to Alexeev and Smirnov, only changing the element set to zero in their decomposition to x_1^1. Note that the decomposition is order two and the nonzero coefficients appearing are $\pm 1, \pm\frac{1}{2}$.

$$\begin{aligned}
p_1(t) &= (\frac{-1}{2}t^2x_2^3 - \frac{1}{2}tx_1^2 + x_1^2)\otimes(-y_1^2 + y_2^2 + ty_1^1)\otimes(z_3^1 + tz_2^1)\\
p_2(t) &= (x_1^2 + \frac{1}{2}x_2^1)\otimes(y_1^2 - y_2^2)\otimes(z_3^1 + z_3^2 + tz_2^1 + tz_2^2)\\
p_3(t) &= (t^2x_2^3 + tx_1^3 - \frac{1}{2}tx_2^2 - x_1^2)\otimes(y_1^2 + y_2^2 + ty_2^1)\otimes z_3^2\\
p_4(t) &= (\frac{1}{2}t^2x_2^3 - tx_1^3 - \frac{1}{2}tx_2^2 + x_1^2)\otimes(y_1^2 + y_2^2 - ty_1^1)\otimes z_3^1\\
p_5(t) &= (-t^2x_2^3 + tx_2^2 - x_2^1)\otimes y_1^2\otimes(z_3^2 + \frac{1}{2}tz_2^1 + \frac{1}{2}tz_2^2 - t^2z_1^1)\\
p_6(t) &= (\frac{1}{2}tx_2^2 + x_1^2)\otimes(-y_1^2 + y_2^2 + ty_2^1)\otimes(z_3^2 + tz_2^2)\\
p_7(t) &= (-tx_1^3 + x_1^2 + \frac{1}{2}x_2^1)\otimes(y_1^2 + y_2^2)\otimes(-z_3^1 + z_3^2)\\
p_8(t) &= (tx_2^2 + x_2^1)\otimes y_2^2\otimes(z_3^1 + \frac{1}{2}tz_2^1 + \frac{1}{2}tz_2^2 + t^2z_1^2).
\end{aligned}$$

Then

$$M^{red}_{\langle 3,2,2\rangle} = \frac{1}{t^2}[p_1(t) + \cdots + p_8(t)].$$

Theorem 4.9.2.1. [**LR0**] *Let $E^{AS,3} = \lim_{t\to 0}\langle p_1(t), \ldots, p_8(t)\rangle \in G(8, A\otimes B\otimes C)$. Then $E^{AS,3} \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is the union of two irreducible algebraic surfaces, both abstractly isomorphic to $\mathbb{P}^1 \times \mathbb{P}^1$: The first is a sub-Segre variety:*

$$Seg_{21,(\beta,\omega),(\gamma,\omega^*)} := [x_1^2] \times \mathbb{P}(v^2\otimes W) \times \mathbb{P}(W^*\otimes u_3),$$

The second, $\mathbb{L}_\alpha$ is a one-parameter family of lines passing through a parametrized curve in $Seg_{21,(\beta,\omega),(\gamma,\omega^*)}$ and the plane conic curve (which has the same parametrization):

$$C_{12,(\beta,\omega),(\gamma,\omega^*)} := \mathbb{P}(\cup_{[s,t]\in\mathbb{P}^1} x_2^1\otimes(sy_1^2 - ty_2^2)\otimes(sz_3^2 + tz_3^1)).$$

The three varieties $C_{12,(\beta,\omega),(\gamma,\omega^*)}$, $Seg_{21,(\beta,\omega),(\gamma,\omega^*)}$, and $\mathbb{L}_\alpha$ respectively play roles analogous to the lines $L_{12,(\beta,\omega)}$, $L_{21,(\gamma,\omega^*)}$, and L_α, as described below.

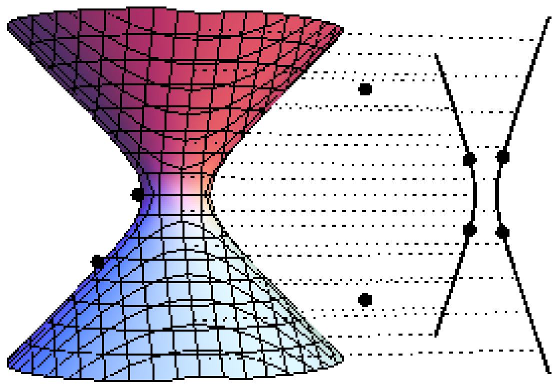


Figure 4.9.1. The curve $C_{12,(\beta,\omega),(\gamma,\omega^*)}$ with its four points, the surface $Seg_{21,(\beta,\omega),(\gamma,\omega^*)}$, with its four points (only two of which are visible), and the surface $\mathbb{L}_\alpha$ with its two points which don't lie on either the curve or surface $Seg_{21,(\beta,\omega),(\gamma,\omega^*)}$.

Proof. The limit points are (up to sign):

$$\begin{aligned}
p_1 =& x_1^2 \otimes (y_1^2 - y_2^2) \otimes z_3^1 \\
p_3 =& x_1^2 \otimes (y_1^2 + y_2^2) \otimes z_3^2 \\
p_4 =& x_1^2 \otimes (y_1^2 + y_2^2) \otimes z_3^1 \\
p_6 =& x_1^2 \otimes (y_1^2 - y_2^2) \otimes z_3^2
\end{aligned}$$

$$\begin{aligned}
p_5 =& x_2^1 \otimes y_1^2 \otimes z_3^2 \\
p_8 =& x_2^1 \otimes y_2^2 \otimes z_3^1
\end{aligned}$$

$$\begin{aligned}
p_2 =& (x_1^2 + \frac{1}{2} x_2^1) \otimes (y_1^2 - y_2^2) \otimes (z_3^1 + z_3^2) \\
p_7 =& (x_1^2 + \frac{1}{2} x_2^1) \otimes (y_1^2 + y_2^2) \otimes (z_3^1 - z_3^2)
\end{aligned}$$

Just as with $M_{\langle 2 \rangle}^{red}$, the limit points all lie on a $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$, in fact the "same" $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$. Pictorially the Segres are:

$$\begin{pmatrix} 0 & * \\ * & \end{pmatrix} \times \begin{pmatrix} & \\ * & * \end{pmatrix} \times \begin{pmatrix} & * \\ & * \end{pmatrix}$$

for $M^{red}_{\langle 2,2,2\rangle}$ and

$$\begin{pmatrix} 0 & * \\ * & \\ & \end{pmatrix} \times \begin{pmatrix} & \\ * & * \end{pmatrix} \times \begin{pmatrix} & & * \\ & & * \end{pmatrix}$$

for $M^{red}_{\langle 3,2,2\rangle}$. Here $E^{AS,3} \cap Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is the union of a one-parameter family of lines $\mathbb{L}_\alpha$ passing through a plane conic and a special $\mathbb{P}^1 \times \mathbb{P}^1$: $Seg_{21,(\beta,\omega),(\gamma,\omega^*)} := [x^2_1] \times \mathbb{P}(v^2 \otimes W) \times \mathbb{P}(W^* \otimes u_3)$ (which contains p_1, p_3, p_4, p_6). To define the family and make the similarity with the BCLR case clearer, first define the plane conic curve

$$C_{12,(\beta,\omega),(\gamma,\omega^*)} := \mathbb{P}(\cup_{[s,t]\in\mathbb{P}^1} x^1_2 \otimes (sy^2_1 - ty^2_2) \otimes (sz^2_3 + tz^1_3)).$$

The points p_5, p_8 lie on this conic (respectively the values $(s,t) = (1,0)$ and $(s,t) = (0,1)$). Then define the variety

$$\mathbb{L}_\alpha := \mathbb{P}(\cup_{[\sigma,\tau]\in\mathbb{P}^1} \cup_{[s,t]\in\mathbb{P}^1} (\sigma x^1_2 + \tau x^2_1) \otimes (sy^2_1 - ty^2_2) \otimes (sz^2_3 + tz^1_3)),$$

which is a one-parameter family of lines intersecting the conic and the special $\mathbb{P}^1 \times \mathbb{P}^1$. The points p_2, p_7 lie on $\mathbb{L}_\alpha$ but not on the conic. Explicitly p_2 (resp. p_7) is the point corresponding to the values $(\sigma,\tau) = (1, \frac{1}{2})$ and $(s,t) = (1,1)$ (resp. $(s,t) = (1,-1)$).

The analog of L_α in the $M^{red}_{\langle 2\rangle}$ decomposition is $\mathbb{L}_\alpha$, and $C_{12,(\beta,\omega),(\gamma,\omega^*)}$ and $Seg_{21,(\beta,\omega),(\gamma,\omega^*)}$ are the analogs of the lines $L_{12,(\beta,\omega)}, L_{21,(\gamma,\omega^*)}$. (A difference here is that $C_{12,(\beta,\omega),(\gamma,\omega^*)} \subset \mathbb{L}_\alpha$.)

The span of the configuration is the span of a $\mathbb{P}^2$ (the span of the conic) and a $\mathbb{P}^3$ (the span of the $\mathbb{P}^1 \times \mathbb{P}^1$), i.e., a $\mathbb{P}^6$.

The proof that these are the only points in the intersection is similar to the BCLR case. □

More decompositions are described geometrically in [**LR0**].

It would be reasonable to expect that the BCLR and Alekseev-Smirnov decompositions generalize to all $\mathbf{m}$, so that $\underline{\mathbf{R}}(M^{red}_{\langle \mathbf{m},2,2\rangle}) \leq 3\mathbf{m} - 1$, which would imply that $\underline{\mathbf{R}}(M_{\langle \mathbf{n},2,2\rangle}) \leq 3\mathbf{n} + 1$ for all $\mathbf{n}$.

Chapter 5

The complexity of Matrix multiplication IV: The complexity of tensors and more lower bounds

In Chapter 2 we developed equations to test the border rank of tensors. The first non-classical such were Strassen's equations for tensors $T \in A{\otimes}B{\otimes}C$. Strassen's equations, as originally presented, were via a study of the geometry of $T(A^*) \subset B{\otimes}C$. In this chapter I explain further techniques for proving lower bounds for border rank and rank of tensors, some of which use Strassen's original perspective. I also discuss geometric properties that could be useful for future investigations.

I begin, in §5.1 by making explicit the dictionary between (1_A-generic) tensors in $\mathbb{C}^{\mathbf{a}}{\otimes}\mathbb{C}^{\mathbf{m}}{\otimes}\mathbb{C}^{\mathbf{m}}$ and linear subspaces of $\mathrm{End}(\mathbb{C}^{\mathbf{m}})$. This enables one to both find further equations for tensors and to use knowledge of tensors to make further progress on classical questions in linear algebra. Classical linear algebra can also be used to show that in certain situations one can conclude upper bounds on border rank that match the lower ones.

While up until now I have emphasized using explicit polynomials to test membership in varieties, sometimes varieties satisfy Zariski closed conditions that are easy to describe but difficult to write as polynomials. Some such are already discussed in §5.1. Two more such conditions are discussed in

§5.2. One particularly useful such technique, the *border substitution method* is discussed in detail in §5.4. In particular, it enables the $2\mathbf{n}^2 - \log_2(2\mathbf{n})$ lower bound for $M_{\langle \mathbf{n} \rangle}$ presented in §5.4.3.

Regarding tensor rank, the only general method for proving tensor rank lower bounds I am aware of is the *substitution* method discussed in §5.3.

The best upper bounds for ω were obtained with $T_{STR}, T_{cw,q}, T_{CW,q}$. What makes these tensors special? It is clear they have nice combinatorial properties, but do they have distinguishing geometric features? I discuss several such geometric properties in §5.5. If such features could be identified, one could in principle look for other tensors with the same properties with which to apply the laser method, as was proposed in [**AFLG15**].

Several tensors that have been studied arise naturally as structure tensors of algebras. I discuss rank and border rank lower bounds for structure tensors in §5.6. In particular, I present Zuiddam's sequence of tensors with rank to border rank ratio approaching three.

5.1. Tensors and classical linear algebra

5.1.1. 1-genericity. How good are Strassen's equations? We have seen that unless there exists $\alpha \in A^*$ with $T(\alpha) \subset B \otimes C$ of maximal rank (or $\beta \in B^*$, resp. $\gamma \in C^*$ with $T(\beta)$, resp. $T(\gamma)$, of maximal rank), they are essentially useless. The following definition names the class of tensors they are useful for.

Definition 5.1.1.1. A tensor $T \in A \otimes B \otimes C$ is 1_A*-generic* if there exists $\alpha \in A^*$ with $T(\alpha) \subset B \otimes C$ of maximal rank, and T is 1-generic if it is $1_A, 1_B$ and 1_C-generic.

Fortunately $M_{\langle \mathbf{n} \rangle}$ and all tensors used to study the exponent of matrix multiplication are 1-generic.

The 1-genericity of $M_{\langle \mathbf{n} \rangle}$ has the consequence that for the purpose of proving upper bounds on $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle})$, one only needs set-theoretic equations for σ_R union the set of non-1-generic tensors. In other words, it would be sufficient to find a collection of polynomials such that their common zero set simply contains σ_R as an irreducible component, as long as all other components of the zero set are contained in the set of non-1-generic tensors.

Say a tensor T is 1_A-generic, $\mathbf{b} = \mathbf{c}$ and Strassen's commutators are identically zero– can we conclude $\underline{\mathbf{R}}(T) = \mathbf{b}$?

I address these questions in this section and the next. I first show that the properties of tensor rank and border rank of tensors in $A \otimes B \otimes C$ can be studied as properties of $\mathbf{a}$-dimensional linear subspaces of $B \otimes C$.

5.1.2. The dictionary. The following standard result shows that when studying the rank and border rank of a tensor $T \in A\otimes B\otimes C$, there is no loss of information in restricting attention to $T(A^*) \subset B\otimes C$. I present a version of it from [**LM15**].

Proposition 5.1.2.1. *Let $T \in A\otimes B\otimes C$.*

(1) *$\mathbf{R}(T)$ equals the minimal number of rank one elements of $B\otimes C$ needed to span (a space containing) $T(A^*)$, and similarly for the permuted statements.*

Say $\dim T(A^*) = k$. Let $Z_r \subset G(k, B\otimes C)$ denote the set of k-planes in $B\otimes C$ that are contained in the span of r rank one elements, so $\mathbf{R}(T) \leq r$ if and only if $T(A^*) \in Z_r$.

(2) *$\underline{\mathbf{R}}(T) \leq r$ if and only if $T(A^*) \in \overline{Z_r}$.*

Proof. Let T have rank r so there is an expression $T = \sum_{i=1}^r a_i\otimes b_i\otimes c_i$. (The vectors a_i need not be linearly independent, and similarly for the b_i and c_i.) Then $T(A^*) \subseteq \langle b_1\otimes c_1, \ldots, b_r\otimes c_r\rangle$ shows that the number of rank one matrices needed to span $T(A^*) \subset B\otimes C$ is at most $\mathbf{R}(T)$.

For the other inequality, say $T(A^*)$ is contained in the span of rank one elements $b_1\otimes c_1, \ldots, b_r\otimes c_r$. Let $\alpha^1, \ldots, \alpha^{\mathbf{a}}$ be a basis of A^*, with dual basis $e_1, \ldots, e_{\mathbf{a}}$ of A. Then $T(\alpha^i) = \sum_{s=1}^r x_s^i b_s\otimes c_s$ for some constants x_s^i. But then $T = \sum_{s,i} e_i\otimes(x_s^i b_s\otimes c_s) = \sum_{s=1}^r (\sum_i x_s^i e_i)\otimes b_s\otimes c_s$ proving $\mathbf{R}(T)$ is at most the number of rank one matrices needed to span $T(A^*) \subset B\otimes C$.

Exercise 5.1.2.2: (1) Prove the border rank assertion.

$\square$

5.1.3. Equations via linear algebra. This section follows [**LM15**]. All the equations we have seen so far arise as Koszul flattenings, which all vanish if Strassen's equations for minimal border rank are zero, as can be seen by the coordinate expressions (2.2.1) and the discussion in §2.6.4. Thus we have robust equations only if T is $1_A, 1_B$ or 1_C-generic, because otherwise the presence of $T(\alpha)^{\wedge \mathbf{a}-1}$ in the expressions make them likely to vanish. When T is 1_A-generic, the Koszul flattenings $T_A^{\wedge p} : \Lambda^p A\otimes B^* \to \Lambda^{p+1} A\otimes C$ provide measures of the failure of $T(A^*)T(\alpha)^{-1} \subset \mathrm{End}(B)$ to be an abelian subspace.

A first concern is that perhaps the choice of $\alpha \in A^*$ effects this failure. The following lemma addresses that concern, at least in the case of minimal border rank:

Lemma 5.1.3.1. [**LM15**] *Let $T \in A\otimes B\otimes C = \mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}$ be 1_A-generic and assume $\mathrm{rank}(T(\alpha_0)) = \mathbf{a}$. If $T(A^*)T(\alpha_0)^{-1}$ is abelian then $T(A^*)T(\alpha_0')^{-1}$ is abelian for any $\alpha_0' \in A^*$ such that $\mathrm{rank}(T(\alpha_0')) = \mathbf{a}$.*

Proof. Say $T(A^*)T(\alpha_0)^{-1}$ is abelian, and set $X_i = T(\alpha_i)T(\alpha_0)^{-1}$, so $[X_1, X_2] = 0$. Set $X_i' = T(\alpha_i)T(\alpha_0)^{-1}$ and $X' = T(\alpha_0')T(\alpha_0)^{-1}$, so $[X_i, X'] = 0$ as well, which implies $[X_i, (X')^{-1}] = 0$. We want to show $[X_1', X_2'] = 0$. But $X_j' = X_j(X')^{-1}$, so

$$\begin{aligned} X_1'X_2' - X_2'X_1' &= X_1(X')^{-1}X_2(X')^{-1} - X_2(X')^{-1}X_1(X')^{-1} \\ &= [X_1, X_2](X')^{-1}(X')^{-1} \\ &= 0. \end{aligned}$$

□

Definition 5.1.3.2. Let $\mathbf{a} = \mathbf{b} = \mathbf{c}$ and let $\mathrm{Abel}_A \subset A \otimes B \otimes C$ denote the set of concise, 1_A-generic tensors such that for some (and hence any) $\alpha \in A^*$ with $T(\alpha)$ of maximal rank, $T(A^*)T(\alpha)^{-1} \subset \mathrm{End}(B)$ is abelian. Note that Abel_A is not Zariski closed in general.

Let $\mathrm{Diag}^0_{\mathrm{End}(B)} \subset G(\mathbf{b}, \mathrm{End}(B))$ denote the set of $\mathbf{b}$-dimensional subspaces that are simultaneously diagonalizable under the action of $GL(B)$ and let $\mathrm{Diag}_{\mathrm{End}(B)} = \overline{\mathrm{Diag}^0_{\mathrm{End}(B)}}$ denote its Zariski closure. Let $\alpha \in A^*$ be such that $T(\alpha)$ is of maximal rank (by Lemma 5.1.3.1, it does not matter which α we take), and let

$$\mathrm{Diag}_A := \overline{\{T \in \mathrm{Abel}_A \mid T(A^*)T(\alpha)^{-1} \in \mathrm{Diag}_{\mathrm{End}(B)}\}} \cap \mathrm{Abel}_A \,.$$

By definition, $\mathrm{Diag}_A \subseteq \mathrm{Abel}_A$. We now study to what extent equality holds. The following proposition gives a necessary algebraic condition to be in Diag_A:

Proposition 5.1.3.3. [**Ger61**] *The set*

$$\{U \in G(\mathbf{a}, \mathrm{End}(B)) \mid U \text{ is closed under composition}\}$$

is Zariski closed.

In particular, if $T \in A \otimes B \otimes C = \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{a}}$ is 1_A-generic with $\underline{\mathbf{R}}(T) = \mathbf{a}$, then for all $\alpha \in A^*$ with $T(\alpha)$ invertible, $T(A^*)T(\alpha)^{-1}$ is closed under composition.

Proof. If $u_1, \ldots, u_{\mathbf{a}}$ is a basis of U, then U is closed under composition if and only if for all $u \in U$,

$$(uu_j) \wedge u_1 \wedge \cdots \wedge u_{\mathbf{a}} = 0 \ \forall 1 \le j \le \mathbf{a}.$$

Let $(Abel_A \times A^*)^0 = \{(T, \alpha) \mid \mathrm{rank}(T(\alpha)) = \mathbf{b}\}$, and note that the map $(Abel_A \times A^*)^0 \to G(\mathbf{a}, \mathrm{End}(B))$, given by $(T, \alpha) \mapsto T(A^*)T(\alpha)^{-1}$ is continuous (it is a regular map of quasi-projective varieties). The "in particular" assertion follows from this continuity because if $U \in \mathrm{Diag}^0_{\mathrm{End}(B)}$, then U is closed under composition. □

Exercise 5.1.3.4: (2) Show that if $T(\alpha), T(\alpha')$ are invertible and $T(A^*)T(\alpha)^{-1}$ is closed under composition, then $T(A^*)T(\alpha')^{-1}$ is closed under composition.

Let $\operatorname{End}\operatorname{Abel}_A \subseteq \operatorname{Abel}_A$ denote the subset of tensors with $T(A)T(\alpha)^{-1}$ closed under composition for some (and hence all) $\alpha \in A^*$ with $T(\alpha)$ invertible. We have

$$\operatorname{Diag}_A \subseteq \operatorname{End}\operatorname{Abel}_A \subseteq \operatorname{Abel}_A, \tag{5.1.1}$$

where the first inclusion is Proposition 5.1.3.3 and the second is by definition. Are these containments strict?

A classical theorem states that when $\mathbf{a} = 3$ all three are equal. Moreover:

Theorem 5.1.3.5. [**IM05**] *When* $\mathbf{a} \leq 4$, $\operatorname{Diag}_A = \operatorname{End}\operatorname{Abel}_A = \operatorname{Abel}_A$.

See [**IM05**] for the proof, which has numerous cases.

What happens when $\mathbf{a} = 5$?

Proposition 5.1.3.6. [**Lei16**] *Let* $T_{Leit,5} = a_1{\otimes}(b_1{\otimes}c_1 + b_2{\otimes}c_2 + b_3{\otimes}c_3 + b_4{\otimes}c_4 + b_5{\otimes}c_5) + a_2{\otimes}(b_1{\otimes}c_3 + b_3{\otimes}c_5) + a_3{\otimes}b_1{\otimes}c_4 + a_4{\otimes}b_2{\otimes}c_4 + a_5{\otimes}b_2{\otimes}c_5$, *which gives rise to the linear space*

$$T_{Leit,5}(A^*) = \begin{pmatrix} x_1 & & & & \\ & x_1 & & & \\ x_2 & & x_1 & & \\ x_3 & x_4 & & x_1 & \\ & x_5 & x_2 & & x_1 \end{pmatrix}. \tag{5.1.2}$$

Then $T_{Leit,5}(A^*)T(\alpha^1)^{-1}$ *is an abelian Lie algebra, but not* End*-closed. I.e.,* $T_{Leit,5} \in \operatorname{Abel}_A$ *but* $T_{Leit,5} \notin \operatorname{End}\operatorname{Abel}_A$.

Throughout this chapter, an expression of the form (5.1.2) is to be read as $T(x_1\alpha^1 + \cdots x_{\mathbf{a}}\alpha^{\mathbf{a}})$ where $\alpha^1, \dots, \alpha^{\mathbf{a}}$ is a basis of A^*.

Exercise 5.1.3.7: (1) Verify that $T_{Leit,5}(A^*)T(\alpha^1)^{-1}$ is not closed under composition.

Thus when $\mathbf{a} \geq 5$, $\operatorname{End}\operatorname{Abel}_A \subsetneq \operatorname{Abel}_A$. The following proposition shows that the first containment in (5.1.1) is also strict when $\mathbf{a} \geq 7$:

Proposition 5.1.3.8. [**LM15**] *The tensor corresponding to*

$$T_{end,7}(A^*) = \begin{pmatrix} x_1 & & & & & & \\ & x_1 & & & & & \\ & & x_1 & & & & \\ & & & x_1 & & & \\ & x_2 + x_7 & x_3 & x_4 & x_1 & & \\ x_2 & x_3 & x_5 & x_6 & & x_1 & \\ x_4 & x_5 & x_6 & x_7 & & & x_1 \end{pmatrix}$$

is in $\operatorname{End}\operatorname{Abel}_A$*, but has border rank at least* 8.

The proof is given in §5.2.1.

We have seen that set-theoretic equations for $\operatorname{End}\operatorname{Abel}_A$ are easy, whereas set-theoretic equations for Diag_A are not known. One might hope that if $T \in \operatorname{End}\operatorname{Abel}_A$, that at least $\underline{\mathbf{R}}(T)$ should be *close* to $\mathbf{a}$. This hope fails miserably:

Proposition 5.1.3.9. [**LM15**] *There exist* 1_A*-generic tensors in* $\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}$ *in* $\operatorname{End}\operatorname{Abel}_A$ *of border rank* $\Omega(\frac{\mathbf{a}^2}{8})$.

In particular, a 1_A-generic tensor satisfying Strassen's equations can have very high border rank.

Proof. Consider T such that

$$T(A^*) \subset \begin{pmatrix} x_1 & & & & & \\ & \ddots & & & & \\ & & x_1 & & & \\ * & \cdots & * & x_1 & & \\ \vdots & \vdots & \vdots & & \ddots & \\ * & \cdots & * & & & x_1 \end{pmatrix}. \tag{5.1.3}$$

and set $x_1 = 0$. We obtain a generic tensor in $\mathbb{C}^{\mathbf{a}-1}\otimes\mathbb{C}^{\lfloor\frac{\mathbf{a}}{2}\rfloor}\otimes\mathbb{C}^{\lceil\frac{\mathbf{a}}{2}\rceil}$, which will have border greater than $\frac{\mathbf{a}^2}{8}$. Conclude by applying Exercise 2.1.6.2. $\square$

Tensors of the form (5.1.3) expose a weakness of Strassen's equations, even under 1-genericity. (Variants of the tensors of the form (5.1.3) are 1-generic and still exhibit the same behavior.)

5.1.4. Sufficient conditions for a concise tensor to be of minimal border rank. A classical result in linear algebra **ref??** says a subspace $U \subset \operatorname{End}(B)$ is diagonalizable if and only if U is abelian and every $x \in U$ (or equivalently for each x_j in a basis of U), x is diagonalizable. This implies:

Proposition 5.1.4.1. *A necessary and sufficient condition for a concise* 1_A*-generic tensor* $T \in A\otimes B\otimes C$ *with* $\mathbf{a} = \mathbf{b} = \mathbf{c}$ *to be of minimal rank* $\mathbf{a}$ *is that for some basis* $\alpha_1, \ldots, a_{\mathbf{a}}$ *of* A^* *with* $\operatorname{rank}(T(\alpha_1)) = \mathbf{b}$*, the space* $T(A)T(\alpha_1)^{-1} \subset \operatorname{End}(B)$ *is abelian and each* $T(\alpha_j)T(\alpha_1)^{-1}$ *is diagonalizable.*

Although we have seen several necessary conditions to be of minimal border rank, a computable necessary and sufficient condition to be of minimal border rank is not known. Below is a sufficient condition to be of minimal border rank.

For $x \in \operatorname{End}(B)$, define the *centralizer* of x, denoted $C(x)$, by

$$C(x) := \{y \in \operatorname{End}(B) \mid [y, x] = 0\}.$$

Definition 5.1.4.2. An element $x \in \operatorname{End}(B)$ is *regular* if $\dim C(x) = \mathbf{b}$, and it is *regular semi-simple* if x is diagonalizable with distinct eigenvalues.

Exercise 5.1.4.3: (2) An $\mathbf{m} \times \mathbf{m}$ matrix is *regular nilpotent* if it is zero except for the super diagonal where the entries are all 1's. Show that a regular nilpotent element is indeed regular, in fact that its centralizer is the space of upper-triangular matrices where the entries on each (upper) diagonal are the same, e.g., when $\mathbf{m} = 3$ the centralizer is

$$\left\{ \begin{pmatrix} x & y & z \\ & x & y \\ & & x \end{pmatrix} \mid x, y, z \in \mathbb{C} \right\}.$$

Exercise 5.1.4.4: (2) Show that $\dim C(x) \geq \mathbf{b}$, with equality if and only if the minimal polynomial of x equals the characteristic polynomial. ⊚

Note that x is regular semi-simple if and only if $C(x) \subset \operatorname{End}(B)$ is a diagonalizable subspace. In this case the eigenvalues of x are distinct.

Proposition 5.1.4.5. *(L. Manivel,* [**LM15**]*) Let* $U \subset \operatorname{End}(B)$ *be an abelian subspace of dimension* $\mathbf{b}$ *such that there exists* $x \in U$ *that is regular. Then* $U \in \operatorname{Diag}_{\operatorname{End}(B)} \subset G(\mathbf{b}, \operatorname{End}(B))$.

Proof. Since the Zariski closure of the regular semi-simple elements is all of $\operatorname{End}(B)$, for any $x \in \operatorname{End}(B)$, there exists a curve x_t of regular semi-simple elements with $\lim_{t\to 0} x_t = x$. Consider the induced curve in the Grassmannian $C(x_t) \subset G(\mathbf{b}, \operatorname{End}(B))$. Then $C_0 := \lim_{t\to 0} C(x_t)$ exists and is contained in $C(x) \subset \operatorname{End}(B)$ and since U is abelian, we also have $U \subseteq C(x)$. But if x is regular, then $\dim C(x) = \dim(U) = \mathbf{b}$, so $\lim_{t\to 0} C(x_t)$, C_0 and U must all be equal and thus U is a limit of diagonalizable subspaces. $\square$

Proposition 5.1.4.5 applied to $T(A)T(\alpha)^{-1}$ provides a sufficient condition for a concise 1_A-generic tensor $T \in A \otimes B \otimes C$ to be of minimal border rank. The condition is not necessary, even for 1-generic tensors, e.g., the Coppersmith-Winograd tensor $T_{q,CW}$ of (3.4.5), is 1-generic of minimal border rank but $T_{q,CW}(A^*)T_{q,CW}(\alpha)^{-1}$ does not contain a regular element for any $\alpha \in A^*$.

Exercise 5.1.4.6: (2) Show that the centralizer of $M_{\mathbb{C}[\mathbb{Z}_m]}(x_1)$ from Example 3.5.1.2 is $M_{\mathbb{C}[\mathbb{Z}_m]}(\mathbb{C}[\mathbb{Z}_m])$ to obtain a second proof that $\underline{\mathbf{R}}(M_{\mathbb{C}[\mathbb{Z}_m]}) = m$.

Problem 5.1.4.7. Determine a criterion for $U \in G(\mathbf{b}, \operatorname{End}(B))$ to be in the closure of the diagonalizable $\mathbf{b}$-planes, when U does not contain a regular element.

Proposition 5.1.4.8. [**LM15**] *Let $T \in A\otimes B\otimes C = \mathbb{C}^m\otimes\mathbb{C}^m\otimes\mathbb{C}^m$ be 1_A and 1_B generic and satisfy the A-Strassen equations. Then, after a suitable choice of identification of A with B via bases, T is isomorphic to a tensor in $S^2A\otimes C$.*

In particular:

(1) *After making choices of general $\alpha \in A^*$ and $\beta \in B^*$, $T(A^*)$ and $T(B^*)$ are GL_m-isomorphic subspaces of $\mathrm{End}(\mathbb{C}^m)$.*
(2) *If T is 1-generic, then T is isomorphic to a tensor in $S^3\mathbb{C}^m$.*

Proof. Let $\{a_i\}, \{b_j\}, \{c_k\}$ respectively be bases of A, B, C, with dual bases $\{\alpha_j\}, \{\beta_j\}, \{\gamma_k\}$. Write $T = \sum t^{ijk} a_i\otimes b_j\otimes c_k$. After a change of basis in A so that $\mathrm{rank}(T(\alpha_1)) = m$ and in B, C, so that it is the identity matrix, we may assume $t^{1jk} = \delta_{jk}$ and after a change of basis B so that $T(\beta_1)$ is of full rank and further changes of bases in A, B, C, we may assume $t^{i1k} = \delta_{ik}$ as well. (To obtain $t^{i1k} = \delta_{ik}$ only requires changes of bases in A, C, but a further change in B may be needed to preserve $t^{1jk} = \delta_{jk}$.) Take $\{\alpha^i\}$ the dual basis to $\{a_j\}$ and identify $T(A^*) \subset \mathrm{End}(\mathbb{C}^m)$ via α^1. Strassen's A-equations then say

$$0 = [T(\alpha^{i_1}), T(\alpha^{i_2})]_{(j,k)} = \sum_l t^{i_1jl}t^{i_2lk} - t^{i_2jl}t^{i_1lk} \ \forall i_1, i_2, j, k.$$

Consider when $j = 1$:

$$0 = \sum_l t^{i_11l}t^{i_2lk} - t^{i_21l}t^{i_1lk} = t^{i_2i_1k} - t^{i_1i_2k} \ \forall i_1, i_2, k,$$

because $t^{i_11l} = \delta_{i_1,l}$ and $t^{i_21l} = \delta_{i_2,l}$. But this says $T \in S^2\mathbb{C}^m\otimes\mathbb{C}^m$.

For the last assertion, say $L_B : B \to A$ is such that $Id_A\otimes L_B\otimes Id_C(T) \in S^2A\otimes C$ and $L_C : C \to A$ is such that $Id_A\otimes Id_B\otimes L_C \in S^2A\otimes B$. Then $Id_A\otimes L_B\otimes L_C(T)$ is in $A^{\otimes 3}$, symmetric in the first and second factors as well as the first and third. But $\mathfrak{S}_3$ is generated by two transpositions, so $Id_A\otimes L_B\otimes L_C(T) \in S^3A$. □

Thus the A, B, C-Strassen's equations, despite being very different modules, when restricted to 1-generic tensors, all have the same zero sets. Strassen's equations in the case of partially symmetric tensors were essentially known to Emil Toeplitz [**Toe77**], and in the symmetric case to Aronhold [**Aro58**].

5.2. Indirectly defined equations

This section and §5.4.1 discuss Zariski closed conditions that in principle give rise to equations, but they are difficult to write down explicitly- to do so systematically one would need to use elimination theory which is impossible to implement in practice other than in very small cases. Nonetheless, for certain tensors these conditions can be used to prove lower bounds on border rank, e.g., the lower bound on $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle})$ via Griesser's equations in §5.2.2 and the state of the art lower bound on $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle})$ of Theorem 5.4.3.1.

5.2.1. Intersection properties.

Exercise 5.2.1.1: (2) [**BCS97**, Ex. 15.14] Given $T \in \mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}} = A\otimes B\otimes C$ that is concise, show that $\mathbb{P}T(A^*) \cap Seg(\mathbb{P}B \times \mathbb{P}C) = \emptyset$ implies $\underline{\mathbf{R}}(T) > \mathbf{a}$. ⊚

Proof of Proposition 5.1.3.8. The fact that $T_{end,7}(A^*)$ is End-closed follows by inspection. The tensor has border rank at least 8 by Exercise 5.2.1.1 as $T_{end,7}(A^*)$ does not intersect the Segre. Indeed, if it intersected Segre, the vanishing of size two minors implies $x_1 = x_4 = 0$, $(x_2 + x_7)x_2 = 0$ and $(x_2 + x_7)x_7 = 0$. If $x_2 + x_7 = 0$ then $x_3 = 0$, and $x_7^2 = (x_2 + x_7)x_7 = 0$ and hence $x_2 = 0$ as well and we are done. If $x_2 = 0$ analogously we obtain $x_7 = 0$ and $x_3 = x_5 = x_6 = 0$. □

A *complete flag* in a vector space V is a sequence of subspaces $0 \subset V_1 \subset V_2 \subset \cdots \subset V_{\mathbf{v}}$ with $\dim V_j = j$.

Proposition 5.2.1.2. [**Lei16, LM15**] *Let $T \in \mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}} = A\otimes B\otimes C$ be concise. If $\underline{\mathbf{R}}(T) = \mathbf{a}$, then there exists a complete flag $A_1 \subset \cdots \subset A_{\mathbf{a}-1} \subset A_{\mathbf{a}} = A^*$, with $\dim A_j = j$, such that $\mathbb{P}T(A_j) \subset \sigma_j(Seg(\mathbb{P}B \times \mathbb{P}C))$.*

Proof. Write $T = \lim_{t\to 0} \sum_{j=1}^{r} a_j(t)\otimes X_j(t)$ where $X_j(t) \in B\otimes C$ have rank one. Since T is concise, we may assume (possibly after re-ordering) without loss of generality that $a_1(t), \ldots, a_{\mathbf{a}}(t)$ is a basis of A for $t \neq 0$. Let $\alpha^1(t), \ldots, \alpha^{\mathbf{a}}(t) \in A^*$ be the dual basis. Then take $A_k(t) = \operatorname{span}\{\alpha^1(t), \ldots, \alpha^k(t)\} \in G(k, A^*)$ and $A_k = \lim_{t\to 0} A_k(t)$. Since $\mathbb{P}T^*(A_k(t)) \subset \sigma_k(Seg(\mathbb{P}B \times \mathbb{P}C))$ the same must be true in the limit. □

One can say even more. For example:

Proposition 5.2.1.3. [**LM15**] *Let $T \in \mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}} = A\otimes B\otimes C$. If $\underline{\mathbf{R}}(T) = \mathbf{a}$ and $T(A^*) \cap Seg(\mathbb{P}B \times \mathbb{P}C) = [X_0]$ is a single point, then $\mathbb{P}(T(A^*) \cap \hat{T}_{[X_0]}Seg(\mathbb{P}B \times \mathbb{P}C))$ must contain a $\mathbb{P}^1$.*

Proof. Say $T(A^*)$ were the limit of $\operatorname{span}\{X_1(t), \ldots, X_{\mathbf{a}}(t)\}$ with each $X_j(t)$ of rank one. Then since $\mathbb{P}T(A^*) \cap Seg(\mathbb{P}B \times \mathbb{P}C) = [X_0]$, we must have each $X_j(t)$ limiting to X_0. But then $\lim_{t\to 0} \operatorname{span}\{X_1(t), X_2(t)\}$, which must be two-dimensional, must be contained in $\hat{T}_{[X_0]}Seg(\mathbb{P}B \times \mathbb{P}C)$ and $T(A^*)$. □

5.2.2. Griesser's equations. The following theorem describes potential equations for $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ in the range $\mathbf{b} < r \leq 2\mathbf{b} - 1$.

Theorem 5.2.2.1. [**Gri86**] *Let* $\mathbf{b} = \mathbf{c}$. *Given a* 1_A*-generic tensor* $T \in A \otimes B \otimes C$ *with* $\underline{\mathbf{R}}(T) \leq r$, *let* $\alpha_0 \in A^*$ *be such that* $T(\alpha_0)$ *is invertible. For* $\alpha' \in A^*$, *let* $X(\alpha') = T(\alpha')T(\alpha_0)^{-1} \in \mathrm{End}(B)$. *Fix* $\alpha_1 \in A^*$. *Consider the space of endomorphisms* $U := \{[X(\alpha_1), X(\alpha')] : B \to B \mid \alpha' \in A^*\} \subset \mathfrak{sl}(B)$. *Then there exists* $E \in G(2\mathbf{b} - r, B)$ *such that* $\dim(U.E) \leq r - \mathbf{b}$.

Remark 5.2.2.2. Compared with the minors of $T_A^{\wedge p}$, here one is just examining the first block column of the matrix appearing in the expression $Q\tilde{Q}$ in (2.6.6), but one is apparently extracting more refined information from it.

Proof. For the moment assume $\mathbf{R}(T) = r$ and $T = \sum_{j=1}^r a_j \otimes b_j \otimes c_j$. Let $\hat{B} = \mathbb{C}^r$ be equipped with basis $e_1, \ldots, e_r$. Define $\pi : \hat{B} \to B$ by $\pi(e_j) = b_j$. Let $i : B \to \hat{B}$ be such that $\pi \circ i = \mathrm{Id}_B$. Choose $B' \subset \hat{B}$ of dimension $r - \mathbf{b}$ such that $\hat{B} = i(B) \oplus B'$, and denote the inclusion and projection respectively $i' : B' \to \hat{B}$ and $\pi' : \hat{B} \to B'$. Pictorially:

$$\begin{array}{ccccc} & & \hat{B} & & \\ & i \nearrow\swarrow \pi & & \pi' \searrow\nwarrow i' & \\ B & & & & B' \end{array}$$

Let $\alpha_0, \alpha_1, \ldots, \alpha_{\mathbf{a}-1}$ be a basis of A^*. Let $\hat{T} = \sum_{j=1}^r a_j \otimes e_j \otimes e_j^* \in A \otimes \hat{B} \otimes \hat{B}^*$ and let $\hat{X}_j := \hat{T}(\alpha_j)\hat{T}(\alpha_0)^{\wedge r-1}$. (Recall that the matrix of $\hat{T}(\alpha_0)^{\wedge r-1}$ is the cofactor matrix of $\hat{T}(\alpha_0)$.) Now in $\mathrm{End}(\hat{B})$ all the commutators $[\hat{X}_i, \hat{X}_j]$ are zero because $\mathbf{R}(\hat{T}) = r$. For all $2 \leq s \leq \mathbf{a} - 1$, $[\hat{X}_1, \hat{X}_s] = 0$ implies

$$\begin{aligned} 0 &= \pi[\hat{X}_1, \hat{X}_s]i \\ &= [X_1, X_s] + (\pi\hat{X}_1 i')(\pi'\hat{X}_s i) - (\pi\hat{X}_s i')(\pi'\hat{X}_1 i) \end{aligned} \tag{5.2.1}$$

Now take $E \subseteq \ker \pi'\hat{X}_1 i \subset B$ of dimension $2\mathbf{b} - r$. Then for all s, $[X_1, X_s] \cdot E \subset \mathrm{Image}\, \pi\hat{X}_1 i'$, which has dimension at most $r - \mathbf{b}$ because $\pi\hat{X}_1 i' : B' \to B$ and $\dim B' = r - \mathbf{b}$. The general case follows by taking limits. $\square$

Proof of Theorem 2.2.2.1. Here there is just one commutator $[X_1, X_2]$ and its rank is at most the sum of the ranks of the other two terms in (5.2.1). But each of the other two terms is a composition of linear maps including i' which can have rank at most $r - \mathbf{b}$, so their sum can have rank at most $2(r - \mathbf{b})$. $\square$

Remark 5.2.2.3. It is not known to what extent Griesser's equations are non-trivial. Proving non-triviality of equations, even when the equations can be written down explicitly, is often more difficult than finding the equations! For example, it took several years after Koszul-flattenings were discovered to prove they were non-trivial to almost the full extent possible. Regarding

Griesser's equations, it is known they are non-trivial up to $r \leq \frac{3}{2}\mathbf{m} + \frac{\sqrt{\mathbf{m}}}{2} - 2$ when $\mathbf{m}$ is odd, and a similar, slightly smaller bound when $\mathbf{m}$ is even by Proposition 5.2.2.5 below. On the other hand the equations are trivial when $r = 2\mathbf{b} - 1$ and all $\mathbf{a}$, and when $r = 2\mathbf{b} - 2$, and $\mathbf{a} \leq \frac{\mathbf{b}}{2} + 2$, in particular $\mathbf{a} = \mathbf{b} = 4$ by [**Lan15**]. I do not know whether or not the equations are trivial for $r = 2\mathbf{b} - 2$, $\mathbf{a} = \mathbf{b}$ and $\mathbf{b} > 4$.

Griesser's equations are most robust when $T(\alpha_1)T(\alpha_0)^{-1}$ is a generic endomorphism, which motivates the following definition:

Definition 5.2.2.4. For a 1_A-generic tensor $T \in A \otimes B \otimes C$, define T to be 2_A-*generic* if there exist $\alpha \in A^*$ such that $T(\alpha) : C^* \to B$ is of maximal rank and $\alpha' \in A^*$ such that $T(\alpha')T(\alpha)^{-1} : B \to B$ is regular semi-simple.

Proposition 5.1.4.5 implies that when $T \in \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$ is concise, 2_A-generic and satisfies Strassen's equations, then $\underline{\mathbf{R}}(T) = \mathbf{m}$.

Unfortunately for proving lower bounds, $M_{\langle \mathbf{n} \rangle}$ is not 2_A-generic. The equations coming from Koszul flattenings, and even more so Griesser's equations, are less robust for tensors that fail to be 2_A-generic. This partially explains why $M_{\langle \mathbf{n} \rangle}$ satisfies some of the Koszul flattening equations and Griesser's equations below. Thus an important problem is to identify modules of equations for σ_r that are robust for non-2-generic tensors.

Proposition 5.2.2.5. [**Lan15**] *Matrix multiplication $M_{\langle \mathbf{n} \rangle}$ fails to satisfy Griesser's equations for $r \leq \frac{3}{2}\mathbf{n}^2 - 1$ when $\mathbf{n}$ is even and $r \leq \frac{3}{2}\mathbf{n}^2 + \frac{\mathbf{n}}{2} - 2$ when $\mathbf{n}$ is odd, and satisfies the equations for all larger r.*

Proof. Consider matrix multiplication $M_{\langle \mathbf{n} \rangle} \in \mathbb{C}^{\mathbf{n}^2} \otimes \mathbb{C}^{\mathbf{n}^2} \otimes \mathbb{C}^{\mathbf{n}^2} = A \otimes B \otimes C$. Recall from Exercise 2.1.7.4 that with a judicious ordering of bases, $M_{\langle \mathbf{n} \rangle}(A^*)$ is block diagonal

$$\begin{pmatrix} x & & \\ & \ddots & \\ & & x \end{pmatrix} \tag{5.2.2}$$

where $x = (x^i_j)$ is $\mathbf{n} \times \mathbf{n}$. In particular, the image is closed under brackets. Choose $X_0 \in M_{\langle \mathbf{n} \rangle}(A^*)$ to be the identity. It is not possible to have $X_1 \in M_{\langle \mathbf{n} \rangle}(A^*)$ diagonal with distinct entries on the diagonal, the most generic choice for X_1 is to be block diagonal with each block having the same $\mathbf{n}$ distinct entries. For a subspace E of dimension $2\mathbf{b} - r = d\mathbf{n} + e$ (recall $\mathbf{b} = \mathbf{n}^2$) with $0 \leq e \leq \mathbf{n} - 1$, the image of a generic choice of $[X_1, X_2], \dots, [X_1, X_{\mathbf{n}^2 - 1}]$ applied to E is of dimension at least $(d+1)\mathbf{n}$ if $e \geq 2$, at least $(d+1)\mathbf{n} - 1$ if $e = 1$ and $d\mathbf{n}$ if $e = 0$, and equality will hold if we choose E to be, e.g., the span of the first $2\mathbf{b} - r$ basis vectors of B. (This is because the $[X_1, X_s]$ will span the entries of type (5.2.2) with zeros on the diagonal.) If $\mathbf{n}$ is even, taking $2\mathbf{b}-r=\frac{\mathbf{n}^2}{2}+1$, so $r=\frac{3\mathbf{n}^2}{2}-1$, the image occupies a space of dimension $\frac{\mathbf{n}^2}{2}+\mathbf{n}-1>\frac{\mathbf{n}^2}{2}-1=r-\mathbf{b}$. If one takes $2\mathbf{b}-r=\frac{\mathbf{n}^2}{2}$, so $r=\frac{3\mathbf{n}^2}{2}$, the image occupies a space of dimension $\frac{\mathbf{n}^2}{2}=r-\mathbf{b}$, showing Griesser's equations cannot do better for $\mathbf{n}$ even. If $\mathbf{n}$ is odd, taking $2\mathbf{b}-r=\frac{\mathbf{n}^2}{2}-\frac{\mathbf{n}}{2}+2$, so $r=\frac{3\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}-2$, the image will have dimension $\frac{\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}>r-\mathbf{b}=\frac{\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}-1$, and taking $2\mathbf{b}-r=\frac{\mathbf{n}^2}{2}-\frac{\mathbf{n}}{2}+1$ the image can have dimension $\frac{\mathbf{n}^2}{2}-\frac{\mathbf{n}}{2}+(\mathbf{n}-1)=r-\mathbf{b}$, so the equations vanish for this and all larger r. Thus Griesser's equations for $\mathbf{n}$ odd give Lickteig's bound $\underline{\mathbf{R}}(M_{\langle\mathbf{n}\rangle})\geq\frac{3\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}-1$. □

5.3. The substitution method

The following method has a long history dating back to [**Pan66**], see [**BCS97**, Chap. 6] and [**Blä14**, Chapter 6] for a history and many applications. It is the only general technique available for proving lower bounds on tensor rank that I am aware of. However, the method cannot be used to prove tensor rank lower bounds of $3\mathbf{m}$ in $\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$. (In §**??** I will describe a powerful method for proving lower bounds on symmetric rank.)

5.3.1. Lower bounds on tensor rank via the substitution method.

Proposition 5.3.1.1. [**AFT11**, Appendix B] *Let $T\in A\otimes B\otimes C$. Fix a basis $a_1,\dots,a_{\mathbf{a}}$ of A, with dual basis $\alpha^1,\dots,\alpha^{\mathbf{a}}$. Write $T=\sum_{i=1}^{\mathbf{a}}a_i\otimes M_i$, where $M_i\in B\otimes C$. Let $\mathbf{R}(T)=r$ and $M_1\neq 0$. Then there exist constants $\lambda_2,\dots,\lambda_{\mathbf{a}}$, such that the tensor*

$$\tilde{T}:=\sum_{j=2}^{\mathbf{a}}a_j\otimes(M_j-\lambda_jM_1)\in\operatorname{span}\{a_2,\dots,a_{\mathbf{a}}\}\otimes B\otimes C,$$

has rank at most $r-1$. Moreover, if $\operatorname{rank}(M_1)=1$ then for any choice of λ_j, $\mathbf{R}(\tilde{T})$ is either r or $r-1$.

The same assertions hold exchanging the role of A with that of B or C.

Proof. (Following [**LM15**].) By Proposition 5.1.2.1 there exist $X_1,\dots,X_r\in\hat{S}eg(\mathbb{P}B\times\mathbb{P}C)$ and scalars d_j^i such that:

$$M_j=\sum_{i=1}^{r}d_j^iX_i.$$

Since $M_1\neq 0$ we may assume $d_1^1\neq 0$ and define $\lambda_j=\frac{d_j^1}{d_1^1}$. Then the subspace $\tilde{T}(\langle\alpha^2,\dots,\alpha^{\mathbf{a}}\rangle)$ is spanned by $X_2,\dots,X_r$ so Proposition 5.1.2.1 implies $\mathbf{R}(\tilde{T})\leq r-1$. The last assertion holds because if $\operatorname{rank}(M_1)=1$

then we may assume $X_1 = M_1$, so we cannot lower the rank by more than one. $\square$

In practice, the method is used iteratively, to reduce T to a smaller and smaller tensor, at each step gaining one in the lower bound for the rank of T.

Example 5.3.1.2. [**AFT11**] Let $T_{aft,3} \in A \otimes B \otimes C$ have an expression in bases such that, letting the columns of the following matrix correspond to B-basis vectors and the rows to C basis vectors,

$$T_{aft,3}(A^*) = \begin{pmatrix} x_1 & & & & & & & \\ & x_1 & & & & & & \\ & & x_1 & & & & & \\ & & & x_1 & & & & \\ x_2 & & & & x_1 & & & \\ & x_2 & & & & x_1 & & \\ x_3 & & x_2 & & & & x_1 & \\ x_4 & x_3 & & x_2 & & & & x_1 \end{pmatrix}.$$

For the first iteration of the substitution method, start with $b_8 \in B$ in the role of a_1 in the Proposition, so write

$$\begin{aligned} T_{aft,3} = & b_1 \otimes (a_1 \otimes c_1 + a_2 \otimes c_5 + a_3 \otimes c_7 + a_4 \otimes c_8) + b_2 \otimes (a_1 \otimes c_2 + a_2 \otimes c_6 + a_3 \otimes c_8) \\ & + b_3 \otimes (a_1 \otimes c_3 + a_2 \otimes c_7) + b_4 \otimes (a_1 \otimes c_4 + a_2 \otimes c_8) \\ & + b_5 \otimes a_1 \otimes c_5 + b_6 \otimes a_1 \otimes c_6 + b_6 \otimes a_1 \otimes c_6 + b_7 \otimes a_1 \otimes c_7 + b_8 \otimes a_1 \otimes c_8. \end{aligned}$$

Then there exist $\lambda_1, \dots, \lambda_7$ and a new tensor $T' \in A \otimes \mathbb{C}^7 \otimes C$ with $\mathbf{R}(T) \geq \mathbf{R}(T') + 1$ where

$$T'(A^*) = \begin{pmatrix} x_1 & & & & & & \\ & x_1 & & & & & \\ & & x_1 & & & & \\ & & & x_1 & & & \\ x_2 & & & & x_1 & & \\ & x_2 & & & & x_1 & \\ x_3 & & x_2 & & & & x_1 \\ x_4 & x_3 & & x_2 & & & \end{pmatrix} + \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ \lambda_1 x_1 & \lambda_2 x_1 & \cdots & \lambda_7 x_1 \end{pmatrix}.$$

Continue removing the last three columns until we get a tensor $T'' \in A\otimes\mathbb{C}^4\otimes C$ with

$$T''(A^*) = \begin{pmatrix} x_1 & & & \\ & x_1 & & \\ & & x_1 & \\ & & & x_1 \\ x_2 & & & \\ & x_2 & & \\ x_3 & & x_2 & \\ x_4 & x_3 & & x_2 \end{pmatrix} + \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \\ \mu_{1,1}x_1 & \mu_{2,1}x_1 & \mu_{3,1}x_1 & \mu_{4,1}x_1 \\ \mu_{1,2}x_1 & \mu_{2,2}x_1 & \mu_{3,2}x_1 & \mu_{4,2}x_1 \\ \mu_{1,3}x_1 & \mu_{2,3}x_1 & \mu_{3,3}x_1 & \mu_{4,3}x_1 \\ \mu_{1,4}x_1 & \mu_{2,4}x_1 & \mu_{3,4}x_1 & \mu_{4,4}x_1 \end{pmatrix}.$$

Now apply the method successively to $c_1, \dots, c_4$ to obtain a tensor T''' with $T'''(A^*) \in \mathbb{C}^4\otimes\mathbb{C}^4$ such that $\mathbf{R}(T_{aft,3}) \geq 8 + \mathbf{R}(T''')$. Now project T''' to the space given by $x_1 = 0$, so in particular all the unknown constants disappear. The new tensor cannot have rank or border rank greater than that of T'''. Iterate the method with the projection of T''' (which is isomorphic to $T_{aft,2}$) until one arrives at $\tilde{T}(A^*) \in \mathbb{C}^1\otimes\mathbb{C}^1$ and the bound $\mathbf{R}(T_{aft,3}) \geq 8 + 4 + 2 + 1 = 15$. In fact $\mathbf{R}(T_{aft,3}) = 15$: observe that $T_{aft,3}(A^*)T_{aft,3}(\alpha^1)^{-1}$ is a projection of the centralizer of a regular nilpotent element as in Exercise 5.3.1.8, which implies $\mathbf{R}(T_{aft,3}) \leq 15$.

On the other hand $\underline{\mathbf{R}}(T_{aft,3}) = 8$, again because $T_{aft,3}(A^*)T_{aft,3}(\alpha^1)^{-1}$ is a projection of the centralizer of a regular nilpotent element so Proposition 5.1.4.5 applies.

This example generalizes to $T_{aft,k} \in \mathbb{C}^{k+1}\otimes\mathbb{C}^{2^k}\otimes\mathbb{C}^{2^k}$ of rank $2 \cdot 2^k - 1$ and border rank 2^k.

Example 5.3.1.3. [**AFT11**] Let $T_{AFT,3} = a_1\otimes(b_1\otimes c_1 + \cdots + b_8\otimes c_8) + a_2\otimes(b_1\otimes c_5 + b_2\otimes c_6 + b_3\otimes c_7 + b_4\otimes c_8) + a_3\otimes(b_1\otimes c_7 + b_2\otimes c_8) + a_4\otimes b_1\otimes c_8 + a_5\otimes b_8\otimes c_1 + a_6\otimes b_8\otimes c_2 + a_7\otimes b_8\otimes c_3 + a_8\otimes b_8\otimes c_4$, so

$$T_{AFT,3}(A^*) = \begin{pmatrix} x_1 & & & & & & & x_5 \\ & x_1 & & & & & & x_6 \\ & & x_1 & & & & & x_7 \\ & & & x_1 & & & & x_8 \\ x_2 & & & & x_1 & & & \\ & x_2 & & & & x_1 & & \\ x_3 & & x_2 & & & & x_1 & \\ x_4 & x_3 & & x_2 & & & & x_1 \end{pmatrix}.$$

Begin the substitution method by distinguishing the spaces A and B, projecting respectively to $\alpha^{8\perp}, \dots, \alpha^{5\perp}$ to obtain a tensor $\tilde{T}$ represented by the matrix

$$\begin{pmatrix} x_1 & & & & & & \\ & x_1 & & & & & \\ & & x_1 & & & & \\ & & & x_1 & & & \\ x_2 & & & & x_1 & & \\ & x_2 & & & & x_1 & \\ x_3 & & x_2 & & & & x_1 \\ x_4 & x_3 & & x_2 & & & \end{pmatrix},$$

and $\mathbf{R}(T_{AFT,3}) \geq 4+\mathbf{R}(\tilde{T})$. The substitution method then gives $\mathbf{R}(\tilde{T}) \geq 14$ by Example 5.3.1.2 and thus $\mathbf{R}(T_{AFT,3}) \geq 18$. This example generalizes to $T_{AFT,k} \in \mathbb{C}^{2^k+1} \otimes \mathbb{C}^{2^k} \otimes \mathbb{C}^{2^k+1}$ of rank at least $3(2^k+1)-k-4$. In fact, equality holds: in the example above, it is enough to consider 17 matrices with just one nonzero entry corresponding to all nonzero entries of $T_{AFT,3}(A^*)$, apart from the top left and bottom right corner and one matrix with 1 at each corner and all other entries equal to 0. Moreover, as observed in [**Lan15**], for these tensors $(2^k+1)+1 \leq \underline{\mathbf{R}}(T_{AFT,k}) \leq 2^{k+1}-k$.

Exercise 5.3.1.4: (2) Prove $(2^k+1)+1 \leq \underline{\mathbf{R}}(T_{AFT,k}) \leq 2^{k+1}-k$. ⊚

In summary:

Proposition 5.3.1.5. *The tensors* $T_{AFT,k} \in \mathbb{C}^{2^k+1} \otimes \mathbb{C}^{2^k} \otimes \mathbb{C}^{2^k+1}$ *of* [**AFT11**] *satisfy* $(2^k+1)+1 \leq \underline{\mathbf{R}}(T_{AFT,k}) \leq 2(2^k+1)-2-k < 3(2^k+1)-k-4 = \mathbf{R}(T_{AFT,k})$.

Exercise 5.3.1.6: (2) Show that for all $\mathbf{m}, \mathbf{n}, N$, $\mathbf{R}(M_{\langle 1,\mathbf{m},\mathbf{n}\rangle} \oplus M_{\langle N,1,1\rangle}) = \mathbf{mn} + N$.

Exercise 5.3.1.7: (2) Show that Strassen's tensor from §5.6, $T_{STR,q} = \sum_{j=1}^{q}(a_0 \otimes b_j \otimes c_j + a_j \otimes b_0 \otimes c_j) \in \mathbb{C}^{q+1} \otimes \mathbb{C}^{q+1} \otimes \mathbb{C}^q$ satisfies $\mathbf{R}(T_{STR,q}) = 2q$.

Exercise 5.3.1.8: (3) Show that a tensor $T \in \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$ corresponding to the centralizer of a regular nilpotent element satisfies $\mathbf{R}(T) = 2\mathbf{m}-1$. ⊚

The limit of this method would be to prove a $3\mathbf{m}-1$ rank lower bound for tensor in $\mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$.

To date, $T_{AFT,k}$ and its cousins are the only known examples of explicit tensors $T \in \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$ satisfying $\mathbf{R}(T) \geq 3\mathbf{m} - O(\log(\mathbf{m}))$. There are several known to satisfy $\mathbf{R}(T) \geq 3\mathbf{m} - O(\mathbf{m})$, e.g., $M_{\langle \mathbf{n} \rangle}$, as was shown in §2.7, and $T_{WState}^{\otimes n} \in \mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n} \otimes \mathbb{C}^{2^n}$ discussed in §5.6.

Problem 5.3.1.9. [**Blä14**] Find an explicit tensor $T \in \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$ satisfying $\mathbf{R}(T) \geq (3+\epsilon)\mathbf{m}$ for any $\epsilon > 0$.

Remark 5.3.1.10. Proposition 5.3.1.1 holds with any choice of basis, so we get to pick $[\alpha^1] \in \mathbb{P}A^*$, as long as $M_1 \neq 0$ (which is automatic if T is

A-concise). On the other hand, there is no choice of the λ_j, so when dealing with $\tilde{T}$, one has to assume the λ_j are as bad as possible for proving lower bounds. For this reason, it is easier to implement this method on tensors with simple combinatorial structure or tensors that are sparse in some basis.

From a geometric perspective, we are restricting T, considered as a trilinear form $A^* \times B^* \times C^* \to \mathbb{C}$, to the hyperplane $A' \subset A^*$ defined by $\alpha^1 + \sum_{j=2}^{\mathbf{a}} \lambda_j \alpha^j = 0$ and our condition is that $\mathbf{R}(T|_{A' \otimes B^* \otimes C^*}) \leq \mathbf{R}(T) - 1$. Our freedom is the choice of $\langle a_2, \ldots, a_{\mathbf{a}} \rangle \subset A$, and then A' is any hyperplane with $A' \cap \langle a_2, \ldots, a_{\mathbf{a}} \rangle^{\perp} = 0$.

5.3.2. Strassen's additivity conjecture. Given $T_1 \in A_1 \otimes B_1 \otimes C_1$ and $T_2 \in A_2 \otimes B_2 \otimes C_2$, if one considers $T_1 + T_2 \in (A_1 \oplus A_2) \otimes (B_1 \oplus B_2) \otimes (C_1 \oplus C_2)$, where each $A_j \otimes B_j \otimes C_j$ is naturally included in $(A_1 \oplus A_2) \otimes (B_1 \oplus B_2) \otimes (C_1 \oplus C_2)$, we saw that $\mathbf{R}(T_1 + T_2) \leq \mathbf{R}(T_1) + \mathbf{R}(T_2)$. Also recall Schönhage's example §3.3.2 that $\underline{\mathbf{R}}(M_{\langle 1,\mathbf{m},\mathbf{n}\rangle} \oplus M_{\langle (\mathbf{n}-1)(\mathbf{m}-1),1,1\rangle}) = \mathbf{mn} + 1 < 2\mathbf{mn} - \mathbf{m} - \mathbf{n} + 1 = \underline{\mathbf{R}}(M_{\langle 1,\mathbf{m},\mathbf{n}\rangle}) + \underline{\mathbf{R}}(M_{\langle (\mathbf{n}-1)(\mathbf{m}-1),1,1\rangle})$. Before this example was known, Strassen made the following conjecture:

Conjecture 5.3.2.1. [**Str73**] *With the above notation,* $\mathbf{R}(T_1 + T_2) = \mathbf{R}(T_1) + \mathbf{R}(T_2)$.

Exercise 5.3.1.6 shows that despite the failure of a border rank analog of the conjecture for $M_{\langle 1,\mathbf{m},\mathbf{n}\rangle} \oplus M_{\langle (\mathbf{n}-1)(\mathbf{m}-1),1,1\rangle}$, the rank version does hold in this case.

While this conjecture has been studied from several different perspectives, e.g. [**FW84, JT86, Bsh98, CCC15b, BGL13**], very little is known about it, and experts are divided as to whether it should be true or false.

In many cases of low rank the substitution method provides the correct rank. In light of this, the following theorem indicates why providing a counter-example to Strassen's conjecture will need new techniques for proving rank lower bounds.

Theorem 5.3.2.2. [**LM15**] *Let $T_1 \in A_1 \otimes B_1 \otimes C_1$ and $T_2 \in A_2 \otimes B_2 \otimes C_2$ be such that that $\mathbf{R}(T_1)$ can be determined by the substitution method applied to two of A_1, B_1, C_1. Then Strassen's additivity conjecture holds for $T_1 \oplus T_2$, i.e., $\mathbf{R}(T_1 \oplus T_2) = \mathbf{R}(T_1) + \mathbf{R}(T_2)$.*

Proof. With each application of the substitution method to elements of A_1, B_1, and C_1, T_1 is modified to a tensor of lower rank living in a smaller space and T_2 is unchanged. After all applications, T_1 has been modified to zero and T_2 is still unchanged. $\square$

The rank of any tensor in $\mathbb{C}^2 \otimes B \otimes C$ can be computed using the substitution method as follows: by dimension count, we can always find either

$\beta \in B^*$ or $\gamma \in C^*$, such that $T(\beta)$ or $T(\gamma)$ is a rank one matrix. In particular, Theorem 5.3.2.2 provides an easy proof of Strassen's additivity conjecture if the dimension of any of A_1, B_1 or C_1 equals 2. This was first shown in [**JT86**] by other methods.

5.4. The border substitution method

What follows are indirectly defined equations for border rank, in other words, algebraic varieties that contain $\sigma_r(Seg(\mathbb{P}A\times\mathbb{P}B\times\mathbb{P}C))$. While we don't have equations for these varieties, sometimes one can prove membership or non-membership by direct arguments. The method is primarily useful for tensors with symmetry, as there border rank decompositions come in families, and we can choose a convenient member of the family to work with.

5.4.1. The border substitution method. The substitution method may be restated as follows:

Proposition 5.4.1.1. *Let $T \in A\otimes B\otimes C$ be A-concise. Fix $\mathbf{a}' < \mathbf{a}$ and $\tilde{A} \subset A$ of dimension $\mathbf{a} - \mathbf{a}'$. Then*

$$\underline{\mathbf{R}}(T) \geq \min_{\{A'\in G(\mathbf{a}',A^*)\mid A'\cap\tilde{A}^{\perp}\neq 0\}} \underline{\mathbf{R}}(T|_{A'\otimes B^*\otimes C^*}) + (\mathbf{a}-\mathbf{a}').$$

Here $\tilde{A}$ in the case $\mathbf{a}' = \mathbf{a}-1$ plays the role of $\langle a_1\rangle$ in Proposition 5.3.1.1. Recall that $T|_{A'\otimes B^*\otimes C^*} \in (A/(A')^{\perp})\otimes B\otimes C$.

More generally,

Proposition 5.4.1.2. *Let $T \in A\otimes B\otimes C$ be concise. Fix $\mathbf{a}' < \mathbf{a}$, $\tilde{A} \subset A$, $\tilde{B} \subset B$ and $\tilde{C} \subset C$ respectively of dimensions $\mathbf{a} - \mathbf{a}'$, $\mathbf{b}' < \mathbf{b}$, and $\mathbf{c} - \mathbf{c}'$. Then*

$$\begin{aligned}\underline{\mathbf{R}}(T) \geq &(\mathbf{a}-\mathbf{a}') + (\mathbf{b}-\mathbf{b}') + (\mathbf{c}-\mathbf{c}')\\ &+ \min_{\left\{\begin{matrix} A'\in G(\mathbf{a}',A^*)\mid A'\cap\tilde{A}^{\perp}\neq 0\\ B'\in G(\mathbf{b}',B^*)\mid B'\cap\tilde{B}^{\perp}\neq 0\\ C'\in G(\mathbf{c}',C^*)\mid A'\cap\tilde{C}^{\perp}\neq 0\end{matrix}\right\}} \underline{\mathbf{R}}(T|_{A'\otimes B^*\otimes C^*}).\end{aligned}$$

A border rank version is as follows:

Proposition 5.4.1.3. [**BL16, LM16b**] *Let $T \in A\otimes B\otimes C$ be A-concise. Fix $\mathbf{a}' < \mathbf{a}$. Then*

$$\underline{\mathbf{R}}(T) \geq \min_{A'\in G(\mathbf{a}',A^*)} \underline{\mathbf{R}}(T|_{A'\otimes B^*\otimes C^*}) + (\mathbf{a}-\mathbf{a}').$$

Proof. Say $\underline{\mathbf{R}}(T) = r$, so $T = \lim_{t\to 0} T_t$, for some tensors $T_t = \sum_{j=1}^r a_j(t)\otimes b_j(t)\otimes c_j(t)$. Without loss of generality, we may assume $a_1(t),\dots,a_{\mathbf{a}}(t)$ form a basis of A. Let $A'_t = \langle a_{\mathbf{a}'+1},\dots,a_{\mathbf{a}}\rangle^{\perp} \subset A^*$. Then $\mathbf{R}(T_t\,|_{A'_t\otimes B^*\otimes C^*}) \leq r - (\mathbf{a}-\mathbf{a}')$ by Proposition 5.4.1.1. Let $A' = \lim_{t\to 0} A'_t \in G(\mathbf{a}',A^*)$. Then

$T|_{A'\otimes B^*\otimes C^*} = \lim_{t\to 0} T_t|_{A'_t\otimes B^*\otimes C^*}$ so $\underline{\mathbf{R}}(T|_{A'\otimes B^*\otimes C^*}) \leq r - (\mathbf{a}-\mathbf{a}')$, i.e., $r \geq \underline{\mathbf{R}}(T|_{A'\otimes B^*\otimes C^*}) + (\mathbf{a}-\mathbf{a}')$. □

Although our freedom in the substitution method was minor (a restriction to a Zariski open subset of the Grassmannian determined by $\tilde{A}^{\perp}$), it is still useful for tensors with simple combinatorial structure. With the border substitution method we have no freedom at all, but nevertheless it will be useful for tensors with symmetry, as the symmetry group will enable us to restrict to special A'.

Corollary 5.4.1.4. [**BL16**] *Let $T \in A\otimes B\otimes C$ be A-concise. Then $\underline{\mathbf{R}}(T) \geq \mathbf{a} - 1 + \min_{\alpha\in A^*\backslash\{0\}} \operatorname{rank}(T(\alpha))$.*

The Corollary follows because for matrices, rank equals border rank, and $\mathbb{C}^1\otimes B\otimes C = B\otimes C$.

As was the case for the substitution method, this procedure can be iterated: write $T_1 = T|_{A'\otimes B^*\otimes C^*}$. If T_1 is B-concise, apply the proposition again with B, if not, let $B_1 \subset B$ be maximal such that T_1 is B_1-concise and then apply the proposition. By successive iterations one finds:

Corollary 5.4.1.5. [**LM16a**] *If for all $A' \subset A^*, B' \subset B^*, C' \subset C^*$ respectively of dimensions $\mathbf{a}', \mathbf{b}', \mathbf{c}'$ one has $T|_{A'\otimes B'\otimes C'} \neq 0$, then $\underline{\mathbf{R}}(T) > \mathbf{a}+\mathbf{b}+\mathbf{c} - (\mathbf{a}'+\mathbf{b}'+\mathbf{c}')$.*

It is obvious this method cannot prove border rank bounds better than $\mathbf{a}+\mathbf{b}+\mathbf{c}-3$. The actual limit of the method is even less than this as I explain in §5.4.4.

5.4.2. How to exploit symmetry. As mentioned above, the border substitution method is particularly useful for tensors T with a large symmetry group G_T, as one can replace the unknown A' by representatives of the closed G_T-orbits in the Grassmannian. I explain the theory in this section and then illustrate it in the next with an improvement in the lower bound for $\underline{\mathbf{R}}(M_{\langle \mathbf{n}\rangle})$. One can also use these methods to limit one's searches for decompositions to certain normal forms. In order to discuss these methods, I first develop language to discuss the G_T orbit closures in the Grassmannian.

To simplify notation for the border substitution method, for a tensor $T \in A_1\otimes\dots\otimes A_k$, and $\tilde{A} \subset A_1$, write

$$T/\tilde{A} := T \mid_{\tilde{A}^{\perp}\otimes A_2^*\otimes\cdots\otimes A_k^*} \in (A_1/\tilde{A})\otimes A_2\otimes\dots\otimes A_k.$$

Define

$$B_{\rho,\mathbf{a}'}(T) := \{\tilde{A} \in G(\mathbf{a}', A_1) \mid \underline{\mathbf{R}}(T/\tilde{A}) \leq \rho\}.$$

Proposition 5.4.2.1. [**LM16a**] *The set $B_{\rho,\mathbf{a}'}(T)$ is Zariski closed.*

As preparation for the proof, I describe two tautological vector bundles over the Grassmannian $G(k,V)$ that are ubiquitous. First the *tautological subspace bundle* $\pi_{\mathcal{S}}:\mathcal{S}\to G(k,V)$ where ${\pi_{\mathcal{S}}}^{-1}(E)=E$. This is a vector subbundle of the trivial bundle with fiber V, which I denote $\underline{V}$. The *tautological quotient bundle* $\pi_{\mathcal{Q}}:\mathcal{Q}\to G(k,V)$ has fiber ${\pi_{\mathcal{Q}}}^{-1}(E)=V/E$, i.e., we have an exact sequence of vector bundles

$$0\to\mathcal{S}\to\underline{V}\to\mathcal{Q}\to 0.$$

All three bundles are $GL(V)$-homogeneous. See ***refs here*** for more details.

For any vector bundle over a projective variety, the corresponding bundle of projective spaces is a projective variety, and a sub-fiber bundle defined by homogeneous equations is also projective. **add ref***

Proof. Consider the bundle $\pi:\mathcal{Q}{\otimes}A_1{\otimes}\cdots\otimes A_k\to G(\mathbf{a}',A_1)$, where $\pi^{-1}(\tilde{A})=(A_1/\tilde{A}){\otimes}A_2{\otimes}\cdots\otimes A_k$. Given T, define a natural section $s_T:G(\mathbf{a}',A_1)\to\mathcal{Q}{\otimes}A_1{\otimes}\cdots\otimes A_k$ by $s_T(\tilde{A}):=T/\tilde{A}$. Let $X\subset\mathbb{P}(\mathcal{Q}{\otimes}A_2{\otimes}\cdots\otimes A_k)$ denote the subvariety (that is also a sub-fiber bundle) defined by $X\cap\mathbb{P}((A_1/\tilde{A}){\otimes}A_2{\otimes}\cdots\otimes A_k)=\sigma_\rho(Seg(\mathbb{P}((A_1/\tilde{A})\times\mathbb{P}A_2\times\cdots\times\mathbb{P}A_k))$. By the discussion above, X is realizable as a projective variety. Let $\tilde{\pi}:X\to G(\mathbf{a}',A_1)$ denote the projectivization of π restricted to X. Then $B_{\rho,\mathbf{a}'}(T)=\tilde{\pi}(X\cap\mathbb{P}s_T(G(\mathbf{a}',A_1)))$. Since the intersection of two projective varieties is a projective variety, as is the image of a projective variety under a regular map (see Theorem 3.1.4.7), we conclude. $\square$

Lemma 5.4.2.2. [**LM16a**] *Let $T\in A_1{\otimes}\dots{\otimes}A_k$ be a tensor, let $G_T\subset GL(A_1)\times\cdots\times GL(A_k)$ denote its stabilizer and let $G_1\subset GL(A_1)$ denote its projection to $GL(A_1)$. Then $B_{\rho,\mathbf{a}'}(T)$ is a G_1-variety.*

Proof. Let $g=(g_1,\dots,g_n)\in G_T$. Then $\underline{\mathbf{R}}(T/\tilde{A})=\underline{\mathbf{R}}(g\cdot T/g\cdot\tilde{A})=\underline{\mathbf{R}}(T/g_1\tilde{A})$. $\square$

Recall the definition of a homogeneous variety $X=G/P\subset\mathbb{P}V$ from Definition 4.8.2.1.

Lemma 5.4.2.3. [**BL14**, Lemma 2.1] *Let $X=G/P\subset\mathbb{P}V$ be a homogeneous variety and let $p\in\sigma_r(X)$. Then there exist a point $x_0\in\hat{X}$ and $r-1$ curves $z_j(t)\in\hat{X}$ such that $p\in\lim_{t\to0}\langle x_0,z_1(t),\dots,z_{r-1}(t)\rangle$.*

Proof. Since $p\in\sigma_r(X)$, there exist r curves $x(t),y_1(t),\dots,y_{r-1}(t)\in\hat{X}$ such that

$$p\in\lim_{t\to0}\mathbb{P}\langle x(t),y_1(t),\dots,y_{r-1}(t)\rangle.$$

Choose a curve $g_t \in G$, such that $g_t(x(t)) = x_0 = x(0)$ for all t and $g_0 = \mathrm{Id}$. We have

$$\begin{aligned}\langle x(t), y_1(t), \dots, y_{r-1}(t)\rangle &= {g_t}^{-1} \cdot \langle x_0, g_t \cdot y_1(t), \dots, g_t \cdot y_{r-1}(t)\rangle \text{ and}\\ \lim_{t\to 0}\langle x(t), y_1(t), \dots, y_{r-1}(t)\rangle &= \lim_{t\to 0}\big({g_t}^{-1} \cdot \langle x_0, g_t \cdot y_1(t), \dots, g_t \cdot y_{r-1}(t)\rangle\big)\\ &= \lim_{t\to 0}\langle x_0, g_t \cdot y_1(t), \dots, g_t \cdot y_{r-1}(t)\rangle.\end{aligned}$$

Set $z_j(t) = g_t \cdot y_j(t)$ to complete the proof. □

Exercise 5.4.2.4: (1) Show that if an algebraic group G acts algebraically on a variety X, that any orbit of minimal dimension must be Zariski closed.

The following Lemma applies both to $M_{\langle \mathbf{n}\rangle}$ and to the determinant polynomial:

Lemma 5.4.2.5 (Normal form lemma)**.** [**LM16b**] *Let $X = G/P \subset \mathbb{P}V$ be a homogeneous variety and let $v \in V$ be such that $G_v := \{g \in G \mid g[v] = [v]\}$ has a single closed orbit $\mathcal{O}_{min}$ in X. Then any border rank r decomposition of v may be modified using G_v to a a border rank r decomposition whose limit plane is $E = \lim_{t\to 0}\langle x_1(t), \dots, x_r(t)\rangle$ where there is a stationary point $x_1(t) \equiv x_1$ lying in $\mathcal{O}_{min}$.*

If moreover every orbit of $G_v \cap G_{x_1}$ contains x_1 in its closure, we may further assume that all other $x_j(t)$ limit to x_1.

Proof. I prove the second statement. By Lemma 5.4.2.3, it is sufficient to show that we can have all points limiting to the same point $x_1(0)$.

Work by induction. Say we have shown that $x_1(t), \dots, x_q(t)$ all limit to the same point $x_1 \in \mathcal{O}_{min}$. It remains to show that our curve can be modified so that the same holds for $x_1(t), \dots, x_{q+1}(t)$. Take a curve $g_\epsilon \in G_v \cap G_{x_1}$ such that $\lim_{\epsilon\to 0} g_\epsilon x_{q+1}(0) = x_1$. For each fixed ϵ, acting on the $x_j(t)$ by g_ϵ, we obtain a border rank decomposition for which $g_\epsilon x_i(t) \to g_\epsilon x_1(0) = x_1(0)$ for $i \le q$ and $g_\epsilon x_{q+1}(t) \to g_\epsilon x_{q+1}(0)$. Fix a sequence $\epsilon_n \to 0$. Claim: we may choose a sequence $t_n \to 0$ such that

- $\lim_{n\to\infty} g_{\epsilon_n} x_{q+1}(t_n) = x_1(0)$,
- $\lim_{n\to\infty} < g_{\epsilon_n} x_1(t_n), \dots, g_{\epsilon_n} x_r(t_n) >$ contains v and
- $\lim_{n\to\infty} g_{\epsilon_n} x_j(t_n) = x_1(0)$ for $j \le q$.

The first point holds as $\lim_{\epsilon\to 0} g_\epsilon x_{q+1}(0) = x_1$. The second follows as for each fixed ϵ_n, taking t_n sufficiently small we may assure that a ball of radius $1/n$ centered at v intersects $< g_{\epsilon_n} x_1(t_n), \dots, g_{\epsilon_n} x_r(t_n) >$. In the same way we may assure that the third point is satisfied. Considering the sequence $\tilde{x}_i(t_n) := g_{\epsilon_n} x_i(t_n)$ we obtain the desired border rank decomposition. □

Exercise 5.4.2.6: (1) Write out a proof of the first assertion in the normal form lemma.

Applying the normal form lemma to matrix multiplication, in order to prove $[M_{\langle \mathbf{n} \rangle}] \notin \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$, it is sufficient to prove it is not contained in a smaller variety. This variety, called the *pointed greater areole* is discussed in §**??**.

5.4.3. The border rank bound $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \geq 2\mathbf{n}^2 - \lceil \log_2(\mathbf{n}) \rceil - 1$.

Theorem 5.4.3.1. [**LM16a**] *Let* $0 < m < \mathbf{n}$. *Then*

$$\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{w} \rangle}) \geq 2\mathbf{n}\mathbf{w} - \mathbf{w} + m - \left\lfloor \frac{\mathbf{w}\binom{\mathbf{n}-1+m}{m-1}}{\binom{2\mathbf{n}-2}{\mathbf{n}-1}} \right\rfloor.$$

In particular, taking $\mathbf{w} = \mathbf{n}$ *and* $m = \mathbf{n} - \lceil \log_2(\mathbf{n}) \rceil - 1$,

$$\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \geq 2\mathbf{n}^2 - \lceil \log_2(\mathbf{n}) \rceil - 1.$$

Proof. First observe that the "In particular" assertion follows from the main assertion because, taking $m = \mathbf{n} - c$, we want c such that

$$\frac{\mathbf{n}\binom{2\mathbf{n}-1-c}{\mathbf{n}}}{\binom{2\mathbf{n}-2}{\mathbf{n}-1}} < 1.$$

This ratio is

$$\frac{(\mathbf{n}-1)\cdots(\mathbf{n}-c)}{(2\mathbf{n}-2)(2\mathbf{n}-3)\cdots(2\mathbf{n}-c)} = \frac{\mathbf{n}-c}{2^{c-1}}\frac{\mathbf{n}-1}{\mathbf{n}-\frac{2}{2}}\frac{\mathbf{n}-2}{\mathbf{n}-\frac{3}{2}}\frac{\mathbf{n}-3}{\mathbf{n}-\frac{4}{2}}\cdots\frac{\mathbf{n}-c+1}{\mathbf{n}-\frac{c}{2}}$$

so if $c - 1 \geq \log_2(\mathbf{n})$ it is less than one.

For the rest of the proof, introduce the following notation: a *Young diagram* associated to a partition $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ is a collection of left aligned boxes, with λ_j boxes in the j-th row. Label it with the upside-down convention as representing entries in the south-west corner of an $\mathbf{n} \times \mathbf{n}$ matrix. More precisely for $(i,j) \in \lambda$ we number the boxes of λ by pairs (row,column) however we number the rows starting from $\mathbf{n}$, i.e. $i = \mathbf{n}$ is the first row. For example

$$\begin{array}{|c|c|} \hline x & y \\ \hline z & \multicolumn{1}{c}{} \\ \cline{1-1} w & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array} \tag{5.4.1}$$

is labeled $x = (\mathbf{n}, 1), y = (\mathbf{n}, 2), z = (\mathbf{n}-1, 1), w = (\mathbf{n}-2, 1)$. Let $\tilde{A}_\lambda :=$ span$\{u^i \otimes v_j \mid (i,j) \in \lambda\}$ and write $M^\lambda_{\langle \mathbf{n},\mathbf{n},\mathbf{w} \rangle} := M_{\langle \mathbf{n},\mathbf{n},\mathbf{w} \rangle}/\tilde{A}_\lambda$.

The proof consists of two parts. The first is to show that for any $k < \mathbf{n}$ there exists a Young diagram λ with k boxes such that $\underline{\mathbf{R}}(M^\lambda_{\langle \mathbf{n},\mathbf{n},\mathbf{w} \rangle}) \leq \underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{w} \rangle}) - k$, and this is done by induction on k.

The second is to use Koszul flattenings to obtain a lower bound on $\underline{\mathbf{R}}(M^{\lambda}_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})$ for any λ.

Part 1) First consider the case $k=1$. By Proposition 5.4.1.3 there exists $[a]\in B_{\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})-1,\mathbf{n}^2-1}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})$ such that the reduced tensor drops border rank. The group $GL(U)\times GL(V)\times GL(W)$ stabilizes $M_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle}$. By Lemma 5.4.2.2 with $G_1=GL(U)\times GL(V)\subset GL(A)$, we may act on $[a]$ and even take limits. Since the $GL(U)\times GL(V)$-orbit closure of any $[a]\in\mathbb{P}A$ contains $[u^{\mathbf{n}}{\otimes}v_1]$, we may replace $[a]$ by $[u^{\mathbf{n}}{\otimes}v_1]$.

Now assume that $\underline{\mathbf{R}}(M^{\lambda'}_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})\leq\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})-k+1$, where λ' has $k-1$ boxes. Again by Proposition 5.4.1.3 there exists $[a']\in B_{\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})-k,\mathbf{n}^2-k}(M^{\lambda'}_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle})$ such that when we reduce by $[a']$ the border rank of the reduced tensor drops. We no longer have the full action of $GL(U)\times GL(V)$. However, the product of *parabolic subgroups* of $GL(U)\times GL(V)$, which by definition are the subgroups that stabilize the flags in U^* and V induced by λ', stabilizes $M^{\lambda'}_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle}$. In particular, all parabolic groups are contained in a *Borel* subgroup of upper-triangular matrices. By the diagonal (torus) action and Lemma 5.4.2.2 we may assume that a has just one nonzero entry outside of λ. Further, using the upper-triangular (Borel) action we can move the entry south-west to obtain the Young diagram λ.

For example, when the Young diagram is (5.4.1) with $\mathbf{n}=4$, and we want to move x^1_4 into the diagram, we may multiply it on the left and right respectively by

$$\begin{pmatrix}\epsilon&&&\\1&1&&\\&&1&\\&&&1\end{pmatrix}\text{ and }\begin{pmatrix}\epsilon&&&1\\&\epsilon&&\\&&1&\\&&&1\end{pmatrix}$$

where blank entries are zero. Then $x^1_4\mapsto\epsilon^2x^1_4+\epsilon(x^2_4+x^4_1)+x^2_1$ and we let $\epsilon\to 0$.

Part 2) Recall that for the matrix multiplication operator, the Koszul flattening of §2.6 factors as $M_{\langle \mathbf{n},\mathbf{n},\mathbf{w}\rangle}=M_{\langle \mathbf{n},\mathbf{n},1\rangle}{\otimes}\mathrm{Id}_W$, so it will suffice to apply the Koszul flattening to $M^{\lambda}_{\langle \mathbf{n},\mathbf{n},1\rangle}\in[(U^*{\otimes}V)/A_\lambda]{\otimes}V^*{\otimes}U$, where $\mathbf{u}=\mathbf{v}=\mathbf{n}$. We need to show that for all λ of size m,

$$\underline{\mathbf{R}}(M^{\lambda}_{\langle \mathbf{n},\mathbf{n},1\rangle})\geq 2\mathbf{n}-1-\frac{\binom{\mathbf{n}-1+m}{m-1}}{\binom{2\mathbf{n}-1}{\mathbf{n}-1}}.$$

This will be accomplished by restricting to a suitable $A' \subset [(U^*\otimes V)/A_\lambda]^*$ of dimension $2\mathbf{n}-1$, such that, setting $\hat{A} = (A')^*$,

$$\mathrm{rank}(M^\lambda_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A'\otimes V\otimes U^*})^{\wedge \mathbf{n}-1}_{\hat{A}}) \geq \binom{2\mathbf{n}-1}{\mathbf{n}-1}\mathbf{n} - \binom{\mathbf{n}-1+m}{m-1},$$

i.e.,

$$\dim\ker(M^\lambda_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A'\otimes V\otimes U^*})^{\wedge \mathbf{n}-1}_{\hat{A}}) \leq \binom{\mathbf{n}-1+m}{m-1},$$

and applying Proposition 2.6.1.1. Since we are working in bases, we may consider $M^\lambda_{\langle \mathbf{n},\mathbf{n},1\rangle} \in (A/A_\lambda)\otimes B\otimes C$ in $A\otimes B\otimes C$, with specific coordinates equal to 0.

Recall the map $\phi : A \to \mathbb{C}^{2\mathbf{n}-1} = \hat{A}$ given by $u^i\otimes v_j \mapsto e_{i+j-1}$ from (2.6.4) and the other notations from the proof of Theorem 2.6.3.6. The crucial part is to determine how many zeros are added to the diagonal when the entries of λ are set to zero. The map $(M^\lambda_{\langle \mathbf{n},\mathbf{n},1\rangle}|_{A'\otimes V\otimes U^*})^{\wedge \mathbf{n}-1}_{\hat{A}}$ is

$$(S,j) = e_{s_1}\wedge\cdots\wedge e_{s_{\mathbf{n}-1}}\otimes v_j \mapsto \sum_{\{k\in[\mathbf{n}]|(i,j)\notin\lambda\}} e_{j+i-1}\wedge e_{s_1}\wedge\cdots\wedge e_{s_{\mathbf{n}-1}}\otimes u^i.$$

Recall that when working with $M_{\langle \mathbf{n},\mathbf{n},1\rangle}$, the diagonal terms in the matrix were indexed by pairs $[(S,j) = (P\backslash p_l, 1+p_l-l), (P,l)]$, in other words that $(P\backslash p_l, 1+p_l-l)$ mapped to (P,l) plus terms that are lower in the order. So fix $(i,j)\in\lambda$, we need to count the number of terms (P,i) that will not appear anymore as a result of (i,j) being in λ. That is, fixing (i,j), we need to count the number of $(p_1,\dots,p_{i-1})$ with $p_1<\cdots<p_{i-1}<i+j-1$, of which there are $\binom{i+j-2}{i-1}$, and multiply this by the number of $(p_{i+1},\dots,p_{\mathbf{n}})$ with $i+j-1<p_{i+1}<\cdots<p_{\mathbf{n}}\leq 2\mathbf{n}-1$, of which there are $\binom{2\mathbf{n}-1-(i+j-1)}{\mathbf{n}-i}$. In summary, each $(i,j)\in\lambda$ kills $g(i,j) := \binom{i+j-1}{i-1}\binom{2\mathbf{n}-i-j}{\mathbf{n}-i}$ terms on the diagonal. Hence, it is enough to prove that $\sum_{(i,j)\in\lambda} g(i,j) \leq \binom{\mathbf{n}-1+m}{m-1}$.

Exercise 5.4.3.2: (1) Show that $\sum_{j=1}^m \binom{\mathbf{n}+j-2}{j-1} = \binom{m+\mathbf{n}-2}{m-1}$. ⊚

By Exercise 5.4.3.2 and a similar calculation, we see $\sum_{i=\mathbf{n}}^{\mathbf{n}-m+1} g(i,1) = \sum_{j=1}^m g(\mathbf{n},j) = \binom{\mathbf{n}-2+m}{m-1}$. So it remains to prove that the Young diagram that maximizes $f_\lambda := \sum_{(i,j)\in\lambda} g(i,j)$ has one row or column. Use induction on the size of λ, the case $|\lambda| = 1$ being trivial. Note that $g(\mathbf{n}-i,j) = g(\mathbf{n}-j,i)$. Moreover, $g(i,j+1)\geq g(i,j)$.

Now say that $\lambda = \lambda'\cup\{(i,j)\}$. By induction it is sufficient to show that:

$$g(\mathbf{n},ij) = \binom{\mathbf{n}+ij-1}{\mathbf{n}-1} \geq \binom{i+j-1}{i-1}\binom{2\mathbf{n}-i-j}{\mathbf{n}-i} = g(i,j), \tag{5.4.2}$$

where $\mathbf{n} > (\mathbf{n}-i)j$.

Exercise 5.4.3.3: (3) Prove the estimate. ⊚

$\square$

5.4.4. Limits of the border substitution method.

Definition 5.4.4.1. A tensor $T \in A\otimes B\otimes C$ is $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$-*compressible* if there exist subspaces $A' \subset A^*, B' \subset B^*, C' \subset C^*$ of respective dimensions $\mathbf{a}', \mathbf{b}', \mathbf{c}'$ such that $T|_{A'\otimes B'\otimes C'} = 0$, i.e., there exists $(A', B', C') \in G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*)$, such that $A'\otimes B'\otimes C' \subset T^\perp$, where $T^\perp \subset (A\otimes B\otimes C)^*$ is the hyperplane annihilating T. Otherwise one says T is $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$-*compression generic*.

Let $X(\mathbf{a}', \mathbf{b}', \mathbf{c}')$ be the set of all tensors that are $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$-compressible.

Proposition 5.4.1.3 may be rephrased as:

$$\sigma_{\mathbf{a}+\mathbf{b}+\mathbf{c}-(\mathbf{a}'+\mathbf{b}'+\mathbf{c}')} Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) \subset X(\mathbf{a}', \mathbf{b}', \mathbf{c}').$$

Proposition 5.4.4.2. [**LM16a**] *The set $X(\mathbf{a}', \mathbf{b}', \mathbf{c}') \subseteq \mathbb{P}(A\otimes B\otimes C)$ is Zariski closed of dimension at most*

$$\min\{\mathbf{abc} - 1, (\mathbf{abc} - \mathbf{a}'\mathbf{b}'\mathbf{c}' - 1) + (\mathbf{a} - \mathbf{a}')\mathbf{a}' + (\mathbf{b} - \mathbf{b}')\mathbf{b}' + (\mathbf{c} - \mathbf{c}')\mathbf{c}'\}.$$

In particular, if

$$\mathbf{aa}' + \mathbf{bb}' + \mathbf{cc}' < (\mathbf{a}')^2 + (\mathbf{b}')^2 + (\mathbf{c}')^2 + \mathbf{a}'\mathbf{b}'\mathbf{c}' \tag{5.4.3}$$

then $X(\mathbf{a}', \mathbf{b}', \mathbf{c}') \subsetneq \mathbb{P}(A\otimes B\otimes C)$, so in this range the substitution methods may be used to prove nontrivial lower bounds for border rank.

The proof and examples show that beyond this bound one expects $X(\mathbf{a}', \mathbf{b}', \mathbf{c}') = \mathbb{P}(A\otimes B\otimes C)$, so that the method cannot be used. Also note that tensors could be quite compressible and still have near maximal border rank, a weakness we already saw with the tensor of (5.1.3) (which also satisfies Strassen's equations).

The inequality in Proposition 5.4.4.2 may be sharp or nearly so. For tensors in $\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$ the limit of this method alone would be a border rank lower bound of $3(\mathbf{m} - \sqrt{3\mathbf{m} + \frac{9}{4}} + \frac{3}{2})$. However, it is unlikely the method alone could attain such a bound due to technical difficulties in proving an explicit tensor does not belong to $X(\mathbf{a}', \mathbf{b}', \mathbf{c}')$.

Proof of Proposition 5.4.4.2. The following is a standard construction in algebraic geometry called an *incidence correspondence* (see, e.g., [**Har95**, §6.12] for a discussion): Let

$$\mathcal{I} :=$$
$$\{((A', B', C'), [T]) \in [G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*)] \times \mathbb{P}(A\otimes B\otimes C) \mid A'\otimes B'\otimes C' \subset T^\perp\}$$

and note that the projection of $\mathcal{I}$ to $\mathbb{P}(A \otimes B \otimes C)$ has image $X(\mathbf{a}', \mathbf{b}', \mathbf{c}')$. A fiber of the other projection $\mathcal{I} \to G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*)$ is $\mathbb{P}((A' \otimes B' \otimes C')^{\perp})$, a projective space of dimension $\mathbf{abc} - \mathbf{a}'\mathbf{b}'\mathbf{c}' - 1$. Hence:

$$\dim \mathcal{I} := (\mathbf{abc} - \mathbf{a}'\mathbf{b}'\mathbf{c}' - 1) + (\mathbf{a} - \mathbf{a}')\mathbf{a}' + (\mathbf{b} - \mathbf{b}')\mathbf{b}' + (\mathbf{c} - \mathbf{c}')\mathbf{c}'.$$

Since the map $\mathcal{I} \to X$ is surjective, this proves the dimension assertion. Since the projection to $\mathbb{P}(A \otimes B \otimes C)$ is a regular map, the Zariski closed assertion also follows. □

The proof of Corollary 5.4.4.3 below uses elementary properties of Chern classes and can be skipped by readers unfamiliar with them. Let $\pi_A : G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*) \to G(\mathbf{a}', A^*)$ denote the projection and similarly for π_B, π_C. Let $\mathcal{E} = \mathcal{E}(\mathbf{a}', \mathbf{b}', \mathbf{c}') := \pi_A^*(\mathcal{S}_A) \otimes \pi_B^*(\mathcal{S}_B) \otimes \pi_C^*(\mathcal{S}_C)$ be the vector bundle that is the tensor product of the pullbacks of tautological subspace bundles $\mathcal{S}_A, \mathcal{S}_B, \mathcal{S}_C$ In each particular case it is possible to explicitly compute how many different $A' \otimes B' \otimes C'$ a generic hyperplane may contain as follows:

Corollary 5.4.4.3. [**LM16a**]

1. *If (5.4.3) holds then a generic tensor is* $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$*-compression generic.*
2. *If (5.4.3) does not hold then* $\operatorname{rank} \mathcal{E}^* \leq \dim\,(G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*))$*. If the top Chern class of* $\mathcal{E}^*$ *is nonzero, then no tensor is* $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$*-compression generic.*

Proof. The first assertion is a restatement of Proposition 5.4.4.2.

For the second, notice that T induces a section $\tilde{T}$ of the vector bundle $\mathcal{E}^* \to G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*)$ defined by $\tilde{T}(A' \otimes B' \otimes C') = T|_{A' \otimes B' \otimes C'}$. The zero locus of $\tilde{T}$ is $\{(A', B', C') \in G(\mathbf{a}', A^*) \times G(\mathbf{b}', B^*) \times G(\mathbf{c}', C^*) \mid A' \otimes B' \otimes C' \subset T^{\perp}\}$. In particular, $\tilde{T}$ is non-vanishing if and only if T is $(\mathbf{a}', \mathbf{b}', \mathbf{c}')$-compression generic. If the top Chern class is nonzero, there cannot exist a non-vanishing section. □

5.5. Geometry of the Coppersmith-Winograd tensors

As we saw in Chapter 3, in practice, only tensors of minimal, or near minimal border rank have been used to prove upper bounds on the exponent of matrix multiplication. Call a tensor that gives a "good" upper bound for the exponent via the methods of [**Str87, CW90**], of *high Coppersmith-Winograd value* or *high CW-value* for short. Ambainis, Filmus and LeGall [**AFLG15**] showed that taking higher powers of $T_{CW,q}$ when $q \geq 5$ cannot prove $\omega < 2.30$ by this method alone. They posed the problem of finding additional tensors of high value. The work in this section was motivated by their problem - to isolate geometric features of the Coppersmith-Winograd tensors and find other tensors with such features. However, it turned out that the features described here actually characterize them! The study is incomplete because the CW-value of a tensor also depends on its presentation, and in different bases a tensor can have quite different CW-values. Moreover, even determining the value in a given presentation still involves some "art" in the choice of a good decomposition, choosing the correct tensor power, estimating the value and probability of each block [**Wil**].

5.5.1. The Coppersmith-Winograd tensors. Recall the Coppersmith-Winograd tensors

$$T_{q,cw} := \sum_{j=1}^{q} a_0\otimes b_j\otimes c_j + a_j\otimes b_0\otimes c_j + a_j\otimes b_j\otimes c_0 \in \mathbb{C}^{q+1}\otimes\mathbb{C}^{q+1}\otimes\mathbb{C}^{q+1} \tag{5.5.1}$$

and

$$\begin{aligned} T_{q,CW} := &\sum_{j=1}^{q}(a_0\otimes b_j\otimes c_j + a_j\otimes b_0\otimes c_j + a_j\otimes b_j\otimes c_0) \\ &+ a_0\otimes b_0\otimes c_{q+1} + a_0\otimes b_{q+1}\otimes c_0 + a_{q+1}\otimes b_0\otimes c_0 \in \mathbb{C}^{q+2}\otimes\mathbb{C}^{q+2}\otimes\mathbb{C}^{q+2} \end{aligned} \tag{5.5.2}$$

both of which have border rank $q+2$.

In terms of matrices,

$$T_{q,cw}(A^*) = \begin{pmatrix} 0 & x_1 & \cdots & & x_q \\ x_1 & x_0 & 0 & \cdots & \\ x_2 & 0 & x_0 & & \\ \vdots & \vdots & & \ddots & \\ x_q & 0 & \cdots & 0 & x_0 \end{pmatrix}.$$

Proposition 5.5.1.1. [**LM15**] $\mathbf{R}(T_{q,cw}) = 2q+1$, $\mathbf{R}(T_{q,CW}) = 2q+3$.

Proof. I prove the lower bound for $T_{q,cw}$. Apply Proposition 5.3.1.1 to show that the rank of the tensor is at least $2q-2$ plus the rank of

$$\begin{pmatrix} 0 & x_1 \\ x_1 & x_0 \end{pmatrix},$$

which has rank 3. An analogous estimate provides the lower bound for $\mathbf{R}(T_{q,CW})$. To show that $\mathbf{R}(T_{q,cw}) \leq 2q+1$ consider the following rank 1 matrices, whose span contains $T(A^*)$:

1) $q+1$ matrices with all entries equal to 0 apart from one entry on the diagonal equal to 1,

2) q matrices indexed by $1 \leq j \leq q$, with all entries equal to zero apart from the four entries $(0,0), (0,j), (j,0), (j,j)$ equal to 1. $\square$

Exercise 5.5.1.2: (2) Using the lower bound for $T_{q,cw}$, prove the lower bound for $T_{q,CW}$.

In §5.6 we saw that $\underline{\mathbf{R}}(T_{STR,q}) = q+1$, and by Exercise 5.3.1.7, $\mathbf{R}(T_{STR,q}) = 2q$. Strassen's tensor has rank nearly twice the border rank, like the Coppersmith-Winograd tensors. So one potential source of high CW-value tensors are tensors with a large gap between rank and border rank.

5.5.2. Extremal tensors. Let $A, B, C = \mathbb{C}^{\mathbf{a}}$. There are normal forms for curves in $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ up to order $\mathbf{a}-1$, namely

$$T_t = (a_1+ta_2+\cdots+t^{\mathbf{a}-1}a_{\mathbf{a}}+O(t^{\mathbf{a}}))\otimes(b_1+tb_2+\cdots+t^{\mathbf{a}-1}b_{\mathbf{a}}+O(t^{\mathbf{a}}))\otimes(c_1+tc_2+\cdots+t^{\mathbf{a}-1}c_{\mathbf{a}}+O(t^{\mathbf{a}}))$$

and if the a_j, b_j, c_j are each linearly independent sets of vectors, we will call the curve *general* to order $\mathbf{a}-1$.

Proposition 5.5.2.1. [**LM15**] *Let $T \in A\otimes B\otimes C = \mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}$. If*

$$T(A^*) = \frac{d^{\mathbf{a}-1}T_t(A^*)}{(dt)^{\mathbf{a}-1}}|_{t=0},$$

with T_t a curve that is general to order $\mathbf{a}$, then, for suitably chosen $\alpha \in A^*$ and bases, $T(A^*)T(\alpha)^{-1}$ is the centralizer of a regular nilpotent element.

Proof. Note that $\frac{d^qT_t}{(dt)^q}|_{t=0} = q!\sum_{i+j+k=q-3} a_i\otimes b_j\otimes c_k$, i.e.,

$$\frac{d^qT_t(A^*)}{(dt)^q}|_{t=0} = \begin{pmatrix} x_{q-2} & x_{q-3} & \cdots & \cdots & x_1 & 0 & \cdots \\ x_{q-3} & x_{q-4} & \cdots & x_1 & 0 & \cdots & \cdots \\ \vdots & & & & & & \\ \vdots & \iddots & & & & & \\ x_1 & 0 & \cdots & & & & \\ 0 & 0 & \cdots & & & & \\ \vdots & \vdots & & & & & \\ 0 & 0 & \cdots & & & & \end{pmatrix}.$$

In particular, each space contains the previous ones, and the last equals

$$\begin{pmatrix} x_{\mathbf{a}} & x_{\mathbf{a}-1} & \cdots & & x_1 \\ x_{\mathbf{a}-1} & x_{\mathbf{a}-2} & \cdots & x_1 & 0 \\ \vdots & \vdots & \ddots & & \\ \vdots & x_1 & & & \\ x_1 & 0 & & & \end{pmatrix}$$

which is isomorphic to the centralizer of a regular nilpotent element. □

This provides another, explicit proof that the centralizer of a regular nilpotent element belongs to the closure of diagonalizable algebras.

Note that the Coppersmith-Winograd tensor $T_{\mathbf{a}-2,CW}$ satisfies $\mathbb{P}T(A^*)\cap Seg(\mathbb{P}B\times\mathbb{P}C)=[X]$ is a single point, and $\mathbb{P}\hat{T}_{[X]}Seg(\mathbb{P}B\times\mathbb{P}C)\cap\mathbb{P}T(A^*)$ is a $\mathbb{P}^{\mathbf{a}-2}$. It turns out these properties characterize it among 1_A-generic tensors:

Theorem 5.5.2.2. [**LM15**] *Let $T\in A\otimes B\otimes C=\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}$ be of border rank $\mathbf{a}>2$. Assume $\mathbb{P}T(A^*)\cap Seg(\mathbb{P}B\times\mathbb{P}C)=[X]$ is a single point, and $\mathbb{P}\hat{T}_{[X]}Seg(\mathbb{P}B\times\mathbb{P}C)\supset\mathbb{P}T(A^*)$. Then T is not 1_A-generic.*

If

(i) $\mathbb{P}T(A^*)\cap Seg(\mathbb{P}B\times\mathbb{P}C)=[X]$ is a single point,

(ii) $\mathbb{P}\hat{T}_{[X]}Seg(\mathbb{P}B\times\mathbb{P}C)\cap\mathbb{P}T(A^*)$ is a $\mathbb{P}^{\mathbf{a}-2}$, and

(iii) T is 1_A-generic,

then T is isomorphic to the Coppersmith-Winograd tensor $T_{\mathbf{a}-2,CW}$.

Proof. For the first assertion, no element of $\mathbb{P}\hat{T}_{[X]}Seg(\mathbb{P}B\times\mathbb{P}C)$ has rank greater than two.

For the second, we first show that T is 1-generic. If we choose bases such that $X=b_1\otimes c_1$, then, after changing bases, the $\mathbb{P}^{\mathbf{a}-2}$ must be the projectivization of

$$E:=\begin{pmatrix} x_1 & x_2 & \cdots & x_{\mathbf{a}-1} & 0\\ x_2 & & & & \\ \vdots & & & & \\ x_{\mathbf{a}-1} & & & & \\ 0 & & & & \end{pmatrix}. \tag{5.5.3}$$

(Rank one tangent vectors cannot appear by property (i).)

Write $T(A^*)=\operatorname{span}\{E,M\}$ for some matrix M. As T is 1_A-generic we can assume that M is invertible. In particular, the last row of M must contain a nonzero entry. In the basis order where M corresponds to $T(\alpha^{\mathbf{a}})$, the space of matrices $T(B^*)$ has triangular form and contains matrices with nonzero diagonal entries. The proof for $T(C^*)$ is analogous, hence T is 1-generic.

By Proposition 5.1.4.8 we may assume that $T(A^*)$ is contained in the space of symmetric matrices. Hence, we may assume that E is as above and M is a symmetric matrix. By further changing the basis we may assume that M has:

(1) the first row and column equal to zero, apart from their last entries that are nonzero (we may assume they are equal to 1),

(2) the last row and column equal to zero apart from their first entries.

Hence the matrix M is determined by a submatrix M' of rows and columns 2 to $\mathbf{a}-1$. As $T(A^*)$ contains a matrix of maximal rank, the matrix M' must have rank $\mathbf{a}-2$. We can change the basis $\alpha^2, \dots, \alpha^{\mathbf{a}-1}$ in such a way that the quadric corresponding to M' equals $x_2^2+\cdots+x_{\mathbf{a}-1}^2$. This will also change the other matrices, which correspond to quadrics x_1x_i for $1 \leq i \leq \mathbf{a}-1$, but will not change the space that they span. We obtain the tensor $T_{\mathbf{a}-2,CW}$.

$\square$

5.5.3. Compression extremality. In this subsection I discuss tensors for which the border substitution method fails miserably. In particular, although the usual substitution method correctly determines the rank of the Coppersmith-Winograd tensors, the tensors are special in that they are nearly characterized by the failure of the border substitution method to give lower border rank bounds.

Definition 5.5.3.1. A 1-generic, tensor $T \in A \otimes B \otimes C$ is said to be *maximally compressible* if there exists hyperplanes $H_A \subset A^*$, $H_B \subset B^*$, $H_C \subset C^*$ such that $T\mid_{H_A \times H_B \times H_C} = 0$.

If $T \in S^3A \subset A \otimes A \otimes A$, T is *maximally symmetric compressible* if there exists a hyperplane $H_A \subset A^*$ such that $T\mid_{H_A \times H_A \times H_A} = 0$.

Recall from Proposition 5.1.4.8 that a tensor $T \in \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{a}}$ that is 1-generic and satisfies Strassen's equations, with suitable choices of bases becomes a tensor in $S^3\mathbb{C}^{\mathbf{a}}$.

Theorem 5.5.3.2. [**LM15**] *Let $T \in S^3\mathbb{C}^{\mathbf{a}}$ be 1-generic and maximally symmetric compressible. Then T is one of:*

(1) $T_{\mathbf{a}-1,cw}$
(2) $T_{\mathbf{a}-2,CW}$
(3) $T = a_1(a_1^2 + \cdots a_m^2)$.

In particular, the only 1-generic, maximally symmetric compressible, minimal border rank tensor in $\mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{a}}$ is isomorphic to $T_{\mathbf{a}-2,CW}$.

Proof. Let a_1 be a basis of the line $H_A{}^{\perp} \subset \mathbb{C}^{\mathbf{a}}$. Then $T = a_1 Q$ for some $Q \in S^2\mathbb{C}^{\mathbf{a}}$. By 1-genericity, the rank of Q is either $\mathbf{a}$ or $\mathbf{a}-1$. If the rank is $\mathbf{a}$, there are two cases, either the hyperplane H_A is tangent to Q, or it intersects it transversely. The second is case (3). The first has a normal form $a_1(a_1a_{\mathbf{a}} + a_2^2 + \cdots + a_{\mathbf{a}-1}^2)$, which, when written as a tensor, is $T_{\mathbf{a}-2,CW}$. If Q has rank $\mathbf{a}-1$, by 1-genericity, $\ker(Q_{1,1})$ must be in H_A and thus we may choose coordinates such that $Q = (a_2^2 + \cdots + a_{\mathbf{a}}^2)$, but then T, written as a tensor is $T_{\mathbf{a}-1,cw}$. $\square$

Proposition 5.5.3.3. *The Coppersmith-Winograd tensor T_{CW} is the unique up to isomorphism 1-generic tensor in $\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}\otimes\mathbb{C}^{\mathbf{a}}$ that is maximally compressible and satisfies any of the following:*

(1) *satisfies Strassen's equations*

(2) *is $\mathbb{Z}_3$-invariant*

(3) *is of border rank $\mathbf{a}$.*

Proof. Let $a_1, \ldots, a_m$ be a basis of A with $H_A = {a_1}^{\perp}$ and similarly for $H_B = {b_1}^{\perp}$ and $H_C = {c_1}^{\perp}$. Thus (allowing re-ordering of the factors A, B, C) $T = a_1\otimes X + b_1\otimes Y + c_1\otimes Z$ where $X \in B\otimes C$, $Y \in A\otimes C$, $Z \in A\otimes B$. Now no $\alpha \in H_A$ can be such that $T(\alpha)$ is of maximal rank, as for any $\beta_1, \beta_2 \in H_B$, $T(\alpha, \beta_j) \subset \mathbb{C}\{c_1\}$. So $T(a^1), T(b^1), T(c^1)$ are all of rank $\mathbf{a}$, where a^1 is the dual basis vector to a_1 etc. After a modification, we may assume X has rank $\mathbf{a}$.

Let $(g, h, k) \in GL(A) \times GL(B) \times GL(C)$. We may normalize $X = \mathrm{Id}$, which forces $g = h$. We may then rewrite X, Y, Z such that Y is full rank and normalize

$$X = Y = \begin{pmatrix} \frac{1}{3} & \\ & \mathrm{Id}_{\mathbf{a}-1} \end{pmatrix}.$$

which forces $h = k$ and uses up our normalizations.

Now we use any of the above three properties. The weakest is the second, but by $\mathbb{Z}_3$-invariance, if $X = Y$, we must have $Z = X = Y$ as well and T is the Coppersmith-Winograd tensor. The other two imply the second by Proposition 5.1.4.8. $\square$

5.6. Ranks and border ranks of Structure tensors of algebras

I now show how the substitution and border substitution methods can be applied to the structure tensors of algebras.

Let $\mathcal{A}$ be a finite dimensional associative algebra and let $T_{\mathcal{A}} \in \mathcal{A}^*\otimes\mathcal{A}^*\otimes\mathcal{A}$ denote its structure tensor as discussed in §3.5.1.

5.6.1. Structural tensors of abelian algebras are symmetric tensors. Let $\mathcal{I} \subset \mathbb{C}[x_1, \ldots, x_n]$ be an ideal with $\mathrm{Zeros}(\mathcal{I}) = \emptyset$, so that $\mathcal{A}_{\mathcal{I}} := \mathbb{C}[x_1, \ldots, x_n]/\mathcal{I}$ is a finite dimensional algebra. Let $\{p_I\}$ be a basis of $\mathcal{A}_{\mathcal{I}}$ with dual basis $\{p_I^*\}$ We can write the structural tensor of $\mathcal{A}_{\mathcal{I}}$ as

$$T_{\mathcal{A}_{\mathcal{I}}} = \sum_{p_I, p_J \in \mathcal{A}_{\mathcal{I}}} p_I^*\otimes p_J^*\otimes(p_I p_J \bmod \mathcal{I}).$$

This tensor is transparently in $S^2\mathcal{A}^*\otimes\mathcal{A}$.

Given an algebra $\mathcal{A} = \mathcal{A}_{\mathcal{I}} \in S^2\mathcal{A}^* \otimes \mathcal{A}$ defined by an ideal as above, note that since $T_{\mathcal{A}}(1, \cdot) \in \mathrm{End}(\mathcal{A})$ and $T_{\mathcal{A}}(\cdot, 1) \in \mathrm{End}(\mathcal{A})$ have full rank and the induced isomorphism $B^* \to C$ is just $(\mathcal{A}^*)^* \to \mathcal{A}$, and similarly for the isomorphism $A^* \to C$. Strassen's equations are thus satisfied, so by Proposition 5.1.4.8 there exists a choice of bases such that $T_{\mathcal{A}} \in S^3\mathcal{A}$.

Question: is there a general recipe for this choice of bases??

Example 5.6.1.1. [**Zui15**] Consider $\mathcal{A} = \mathbb{C}[x]/(x^2)$, with basis $1, x$, so

$$T_{\mathcal{A}} = 1^* \otimes 1^* \otimes 1 + x^* \otimes 1^* \otimes x + 1^* \otimes x^* \otimes x$$

writing $e_0 = 1^*$, $e_1 = x^*$ in the first two factors and $e_0 = x$, $e_1 = 1$ in the third, we see

$$T_{\mathcal{A}} = e_0 \otimes e_0 \otimes e_1 + e_1 \otimes e_0 \otimes e_0 + e_0 \otimes e_1 \otimes e_0$$

That is, $T_{\mathcal{A}} = T_{WState}$ is a general tangent vector to $Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$.

More generally, consider $\mathcal{A} = \mathbb{C}[x_1, \ldots, x_n]/(x_1^2, \ldots, x_n^2)$, with basis $x_I = x_{i_1} \cdots x_{i_{|I|}}$, where $1 \leq i_1 < \cdots < i_{|I|} \leq n$, and by convention $x_\emptyset = 1$. Then

$$T_{\mathcal{A}} = \sum_{I,J \subset [n] | I \cap J = \emptyset} x_I^* \otimes x_J^* \otimes x_{I \cup J}.$$

Similar to above, let $e_I = x_I^*$ in the first two factors and $e_I = x_{[n] \backslash I}$ in the third, we obtain

$$T_{\mathcal{A}} = \sum_{\{I,J,K | \substack{I \cup J \cup K = [n], \\ |I|+|J|+|K|=n}\}} e_I \otimes e_J \otimes e_K$$

so we explicitly see $T_{\mathcal{A}} \in S^3\mathbb{C}^{2^n}$.

Exercise 5.6.1.2: (2) Show that for $\mathcal{A} = \mathbb{C}[x_1, \ldots, x_n]/(x_1^2, \ldots, x_n^2)$, $T_{\mathcal{A}} \simeq T_{WState}^{\otimes n}$, where for $T \in A \otimes B \otimes C$, consider $T^{\otimes n} \in (A^{\otimes n}) \otimes (B^{\otimes n}) \otimes (C^{\otimes n})$ as a three-way tensor.

Exercise 5.6.1.3: (2) Let $\mathcal{A} = \mathbb{C}[x]/(x^n)$. Show that $T_{\mathcal{A}}(\mathcal{A}) T_{\mathcal{A}}(1)^{-1} \subset \mathrm{End}(\mathcal{A})$ corresponds to the centralizer of a regular nilpotent element, so in particular $\underline{\mathbf{R}}(T_{\mathcal{A}}) = n$ and $\mathbf{R}(T_{\mathcal{A}}) = 2n - 1$ by Exercise 5.3.1.8 and Proposition 5.1.4.5.

Exercise 5.6.1.4: (2) Fix natural numbers $a_1, \ldots, a_n$. Let $\mathcal{A} = \mathbb{C}[x_1, \ldots, x_n]/(x_1^{a_1}, \ldots, x_n^{a_n})$. Find an explicit identification $\mathcal{A}^* \to \mathcal{A}$ that renders $T_{\mathcal{A}} \in S^3\mathcal{A}$.

Example 5.6.1.5. [**Zui15**] Consider the tensor

$$T_{WState,k} = a_{1,0} \otimes \cdots \otimes a_{k-1,0} \otimes a_{k,1} + a_{1,0} \otimes \cdots \otimes a_{k-2,0} \otimes a_{k-1,1} \otimes a_{k,0} + \cdots + a_{1,1} \otimes a_{2,0} \otimes \cdots \otimes a_{k,0}$$

that corresponds to a general tangent vector to $Seg(\mathbb{P}^1 \times \cdots \times \mathbb{P}^1) \in \mathbb{P}((\mathbb{C}^2)^{\otimes k})$. (Note that $T_{WState} = T_{WState,3}$.) This tensor is called the *generalized W-state* by physicists . Let $\mathcal{A}_{d,N} = (\mathbb{C}[x]/(x^d))^{\otimes N} \simeq \mathbb{C}[x_1, \ldots, x_N]/(x_1^d, \ldots, x_N^d)$.

Exercise 5.6.1.6: (2) Show that $T_{\mathcal{A}_{d,N}} = (T_{WState,d})^{\otimes N}$.

Example 5.6.1.7 (The Coppersmith-Winograd tensor)**.** [**LM16b, BL16**] Consider the algebra

$$\mathcal{A}_{CW,q} = \mathbb{C}[x_1, \dots, x_q]/(x_i x_j, x_i^2 - x_j^2, x_i^3,\ i \neq j)$$

Let $\{1, x_i, [x_1^2]\}$ be a basis of $\mathcal{A}$, where $[x_1^2] = [x_j^2]$ for all j. Then

$$\begin{aligned} T_{\mathcal{A}_{CW,q}} =& 1^*\otimes 1^*\otimes 1 + \sum_{i=1}^{q} (1^*\otimes x_i^*\otimes x_i + x_i^*\otimes 1^*\otimes x_i) \\ & + x_i^*\otimes x_i^*\otimes [x_1^2] + 1^*\otimes [x_1^2]^*\otimes [x_1^2] + [x_1^2]^*\otimes 1^*\otimes [x_1^2]. \end{aligned}$$

Set $e_0 = 1^*$, $e_i = x_i^*$, $e_{q+1} = [x_1^2]^*$ in the first two factors and $e_0 = [x_1^2]$, $e_i = x_i$, $e_{q+1} = 1$ in the third to obtain

$$\begin{aligned} T_{\mathcal{A}_{CW,q}} =& T_{CW,q} = e_0\otimes e_0\otimes e_{q+1} + \sum_{i=1}^{q} (e_0\otimes e_i\otimes e_i + e_i\otimes e_0\otimes e_i + e_i\otimes e_i\otimes e_0) \\ & + e_0\otimes e_{q+1}\otimes e_0 + e_{q+1}\otimes e_0\otimes e_0 \end{aligned}$$

so we indeed obtain the Coppersmith-Winograd tensor.

When is the structure tensor of $\mathcal{A}_{\mathcal{I}}$ of minimal border rank? First of all, if $T \in \mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}\otimes\mathbb{C}^{\mathbf{m}}$ is the structure tensor of an algebra $\mathcal{A}$ that is a degeneration of $(\mathbb{C}[x]/(x))^{\oplus\mathbf{m}}$ (whose structure tensor is $M_{\langle 1\rangle}^{\oplus\mathbf{m}}$), then $\underline{\mathbf{R}}(T) = \mathbf{m}$. In §**??** we will see that a converse holds under the assumptions of 1_A and 1_B genericity.

5.6.2. The substitution method applied to structure tensors of algebras.

Let $\mathcal{A}$ be a finite dimensional associative algebra. The *radical* of $\mathcal{A}$ is the intersection of all maximal left ideals and denoted $\mathrm{Rad}(\mathcal{A})$. When $\mathcal{A}$ is abelian, the radical is often call the *nilradical*.

Exercise 5.6.2.1: (2) Show that every element of $\mathrm{Rad}(\mathcal{A})$ is nilpotent and that if $\mathcal{A}$ is abelian, $\mathrm{Rad}(\mathcal{A})$ consists exactly of the nilpotent elements of $\mathcal{A}$. (This exercise requires knowledge of standard notions from algebra.) ⊚

Theorem 5.6.2.2. [**Blä00**, Thm. 7.4] *For any integers $p, q \geq 1$,*

$$\mathbf{R}(T_{\mathcal{A}}) \geq \dim(Rad(\mathcal{A})^p) + \dim(Rad(\mathcal{A})^q) + \dim\mathcal{A} - \dim(Rad(\mathcal{A})^{p+q-1}).$$

For the proof we will need the following Lemma, whose proof I skip:

Lemma 5.6.2.3. [**Blä00**, Lem. 7.3] *Let $\mathcal{A}$ be a finite dimensional algebra, let $U, V \subseteq \mathcal{A}$ be vector subspaces such that $U + Rad(\mathcal{A})^p = \mathcal{A}$ and $V + Rad(\mathcal{A})^q = \mathcal{A}$. Then $\langle UV\rangle + Rad(\mathcal{A})^{p+q-1} = \mathcal{A}$.*

Proof of Theorem 5.6.2.2. Use Proposition 5.4.1.2 with

$$\tilde{A} = (\mathrm{Rad}(\mathcal{A})^p)^\perp \subset \mathcal{A}^*,$$
$$\tilde{B} = (\mathrm{Rad}(\mathcal{A})^q)^\perp \subset \mathcal{A}^*, \text{ and}$$
$$\tilde{C} = \mathrm{Rad}(\mathcal{A})^{p+q-1} \subset \mathcal{A}.$$

Then observe that any $A' \subset \mathcal{A}\backslash\mathrm{Rad}(\mathcal{A})^p$, $B' \subset \mathcal{A}\backslash\mathrm{Rad}(\mathcal{A})^q$, can play the roles of U, V in the Lemma, so $T_{\mathcal{A}}(A', B') \not\subset \mathrm{Rad}(\mathcal{A})^{p+q-1}$. Since $C' \subset \mathcal{A}^*\backslash(\mathrm{Rad}(\mathcal{A})^{p+q-1})^\perp$, we conclude. □

Remark 5.6.2.4. Theorem 5.6.2.2 illustrates the power of the (rank) substitution method over the border substitution method. By merely prohibiting a certain Zariski closed set of degenerations, we can make $T_{\mathcal{A}}$ noncompressible. Without that prohibition, $T_{\mathcal{A}}$ can indeed be compressed in general.

Remark 5.6.2.5. Using similar (but easier) methods, one can show that if $\mathcal{A}$ is simple of dimension $\mathbf{a}$, then $\mathbf{R}(T_{\mathcal{A}}) \geq 2\mathbf{a} - 1$, see, e.g., [**BCS97**, Prop. 17.22]. However in the literature, this use of the substitution method is phrased with respect to the elements appearing in a decomposition, making its implementation more complicated.

Theorem 5.6.2.6. [**Zui15**] $\mathbf{R}(T_{WState}^{\otimes n}) = 3 \cdot 2^n - o(2^n)$.

Proof. We have $\mathcal{A} = \mathbb{C}[x_1, \ldots, x_n]/(x_1^2, \ldots, x_n^2)$, so the degree s component of $\mathcal{A}$ is $\mathcal{A}_s = \mathrm{span}\sqcup_{S\subset[n]}\{x_1 \cdots \hat{x}_{i_1} \cdots \hat{x}_{i_s} \cdots x_n\}$. In particular $\dim \mathcal{A}_s = \binom{n}{s}$.

Note that $\mathrm{Rad}(\mathcal{A})^m = \oplus_{j\geq m}\mathcal{A}_j$. Recall that $\sum_{j=0}^n \binom{n}{j} = 2^n$. Take $p = q$ in Theorem 5.6.2.2. We have

$$\begin{aligned}\mathbf{R}(T_{\mathcal{A}}) &\geq 2^n + 2\sum_{j=p}^{n}\binom{n}{j} - \sum_{k=2p-1}^{n}\binom{n}{k} \\ &= 3 \cdot 2^n - 2\sum_{j=0}^{p}\binom{n}{j} - \sum_{k=0}^{n-2p+1}\binom{n}{k}.\end{aligned}$$

Write $p = \epsilon n$, for some $0 < \epsilon < 1$. Since $\sum_{j=0}^{\epsilon n}\binom{n}{j} \leq 2^{H(\epsilon)n}$ **ref big numbers**, taking, e.g., $\epsilon = \frac{1}{3}$ gives the result. □

Corollary 5.6.2.7. [**Zui15**] $\frac{\mathbf{R}(T_{WState}^{\otimes n})}{\underline{\mathbf{R}}(T_{WState}^{\otimes n})} \geq 3 - o(1)$, *where the right hand side is viewed as a function of* n.

More generally, Zuiddam shows, for $T_{WState,k}^{\otimes n} \in (\mathbb{C}^n)^{\otimes k}$:

Theorem 5.6.2.8. [**Zui15**] $\mathbf{R}(T_{WState,k}^{\otimes n}) = k2^n - o(2^n)$.

Regarding the maximum possible ratio for rank to border rank, there is the following theorem applicable even to X-rank and X-border rank:

Theorem 5.6.2.9. [**BT15**] *Let $X \subset \mathbb{P}V$ be a complex projective variety not contained in a hyperplane. Let $\underline{\mathbf{R}}_{X,max}$ denote the maximum X-border rank of a point in $\mathbb{P}V$ and $\mathbf{R}_{X,max}$ the maximum possible X-rank. Then $\mathbf{R}_{X,max} \leq 2\underline{\mathbf{R}}_{X,max}$.*

Proof. Let $U \subset \mathbb{P}V$ be a Zariski dense open subset of points of rank exactly $\mathbf{R}_{X,max}$. Let $q \in \mathbb{P}V$ be any point and let p be any point in U. The line L through q and p intersects U at another point p (in fact, at infinitely many more points). Since p and p' span L, q is a linear combination of p and p', thus $\mathbf{R}_X(q) \leq \mathbf{R}_X(p) + \mathbf{R}_X(p')$ □

Theorem 5.6.2.9 implies that the maximal possible rank of any tensor in $\mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}} \otimes \mathbb{C}^{\mathbf{m}}$ is at most $2\lceil \frac{\mathbf{m}^3-1}{3\mathbf{m}-2} \rceil$, so for any concise tensor the maximal rank to border rank ratio is bounded above by approximately $\frac{2\mathbf{m}}{3}$.

5.6.3. The border substitution method and tensor powers of $T_{cw,2}$.

Lemma 5.6.3.1. [**BL16**] *For any tensor $T_1 \in A_1 \otimes B_1 \otimes C_1$, and any $q \geq 2$,*

$\min(\mathrm{rank}_{\alpha \in (A \otimes A_1)^* \backslash \{0\}} \mathbf{R}((T_{cw,q} \otimes T_1) |_{\alpha \otimes B^* \otimes C^*}) \geq 2 \min \mathrm{rank}_{\alpha_1 \in A_1 \backslash \{0\}} \mathbf{R}(T_1 |_{\alpha_1 \otimes B_1^* \otimes C_1^*})$.

Proof. Write $\alpha = 1 \otimes \alpha_0 + \sum_{j=1}^q e_j^* \otimes \alpha_j \in (A \otimes A_1)^*$ for some $\alpha_0, \alpha_j \in A_1^*$. If all the α_j are zero for $1 \leq j \leq q$, then $T_{cw,q}(e_0^* \otimes \alpha_0)$ is the reordering and grouping of

$$\sum_{i=1}^q (e_i \otimes e_i) \otimes T_1(\alpha_0)$$

which has rank (as a linear map) at least $q \cdot \mathrm{rank} T_1(\alpha_0)$. Otherwise without loss of generality, assume $\alpha_1 \neq 0$. Note that $T_{cw,q}(e_1^* \otimes \alpha_1)$ is the reordering and grouping of

$$e_1 \otimes e_0 \otimes T_1(\alpha_1) + e_0 \otimes e_1 \otimes T_1(\alpha_1)$$

which has rank two, and is linearly independent of any of the other factors appearing in the image, so the rank is at least $2 \cdot \mathrm{rank} T_1(\alpha_0)$. □

Theorem 5.6.3.2. [**BL16**] *For all $q \geq 2$, $\underline{\mathbf{R}}(T_{cw,q}^{\otimes n}) \geq (q+1)^n + 2^n - 1$.*

Proof. Note that $T_{cw,q}^{\otimes n} = T_{cw,q} \otimes T_{cw,q}^{\otimes (n-1)}$. Apply the Lemma iteratively and use Corollary 5.4.1.4. □

Remark 5.6.3.3. As was pointed out implicitly in [**BCS97**] and explicitly in [**BL16**], if the asymptotic border rank (see Definition 3.4.6.1) of $T_{cw,2}$ is the minimal 3, then the exponent of matrix multiplication is 2. The bound in the theorem does not rule this out.

Chapter 6

Valiant's conjecture I: permanent v. determinant and the complexity of polynomials

Recall from the introduction that for a polynomial P, the *determinantal complexity of* P, denoted $\mathrm{dc}(P)$, is the smallest n such that P is an affine linear projection of the determinant, and Valiant's conjecture 1.2.4.2 that $\mathrm{dc}(\mathrm{perm}_m)$ grows faster than any polynomial in m. In this chapter I discuss the conjecture, progress towards it, and its Geometric Complexity Theory variant.

I begin, in §6.1, with a discussion of circuits, context for Valiant's conjecture, definitions of the complexity classes **VP** and **VNP**, and the strengthening of Valiant's conjecture of [**MS01**] that is more natural for algebraic geometry and representation theory. In particular, I explain why it might be considered as an algebraic analog of the famous $\mathbf{P} \neq \mathbf{NP}$ conjecture (although there are other conjectures in the Boolean world that are more closely related to it).

Our study of matrix multiplication indicates a strategy for Valiant's conjecture: look for polynomials on the space of polynomials that vanish on the determinant and not on the permanent. One should look for such polynomials with the aid of geometry and representation theory. Here there is extra geometry available: a polynomial $P \in S^d V$ defines a *hypersurface*

$$Z(P) := \{[\alpha] \in \mathbb{P}V^* \mid P(\alpha) = 0\} \subset \mathbb{P}V^*.$$

Hypersurfaces in projective space have been studied for hundreds of years and much is known about them.

In §6.2 I discuss the simplest polynomials on spaces of polynomials, the *catalecticants* that date back to Sylvester.

One approach to Valiant's conjecture discussed at several points in this chapter is to look for pathologies of the hypersurface $Z(\det_n)$ that persist under degeneration, and that are not shared by $Z(\ell^{n-m}\operatorname{perm}_m)$. The simplest pathology of a hypersurface is its singular set. I discuss the singular loci of the permanent and determinant, as well as a few general remarks on singularities in §6.3.

I then present the classical and recent lower bounds on $\operatorname{dc}(\operatorname{perm}_m)$ of von zur Gathen and Alper-Bogart-Velasco in §6.3.3. These lower bounds on $\operatorname{dc}(\operatorname{perm}_m)$ rely on a key regularity result observed by von zur Gathen. These results cannot extend to the Mulmuley-Sohoni measure $\overline{dc}(\operatorname{perm}_m)$ defined in §6.1.6 because of the regularity result.

The best general lower bound on $\operatorname{dc}(\operatorname{perm}_m)$, namely $\operatorname{dc}(\operatorname{perm}_m) \geq \frac{m^2}{2}$, comes from local differential geometry: the study of Gauss maps. It is presented in §6.4. This bound does extend to $\overline{dc}(\operatorname{perm}_m)$ after some work. The extension is presented in §6.5. To better utilize geometry and representation theory, we will examine the symmetries of the permanent and determinant. Given $P \in S^d V$, let $G_P := \{g \in GL(V) \mid g \cdot P = P\}$ denote the symmetry group of the polynomial P.

Since $\det(AXB) = \det(X)$ if A, B are $n \times n$ matrices with determinant one, and $\det(X^T) = \det(X)$, writing $V = E \otimes F$ with $E, F = \mathbb{C}^n$, we have a map

$$(SL(E) \times SL(F)) \rtimes \mathbb{Z}_2 \to G_{\det_n}$$

where the $\mathbb{Z}_2$ is transpose.

Similarly, letting $T_E \subset SL(E)$ denote the diagonal matrices, we have a map

$$[(T_E \rtimes \mathfrak{S}_n) \times (T_F \rtimes \mathfrak{S}_n)] \rtimes \mathbb{Z}_2 \to G_{\operatorname{perm}_n}.$$

In §6.6, I show that both maps are surjective.

Just as it is interesting and useful to study the difference between rank and border rank, it is worthwhile to study the difference between dc and $\overline{\mathrm{dc}}$, which I discuss in §6.7.

Finally, although it is not strictly related to complexity theory, I cannot resist a brief discussion of determinantal hypersurfaces - those degree n polynomials P with $\operatorname{dc}(P) = n$ in §6.8.

In this chapter I emphasize material that is not widely available to computer scientists, and do not present proofs that already have excellent expositions in the literature such as the completeness of the permanent for **VNP**.

This chapter may be read mostly independently of chapters 2-5.

6.1. Circuits and definitions of VP and VNP

In this section I give definitions of $\mathbf{VP}, \mathbf{VNP}$ via arithmetic circuits and show $(\det_n) \in \mathbf{VP}$. I first discuss why Valiant's conjecture is a cousin of $\mathbf{P} \neq \mathbf{NP}$, namely I show that the permanent can compute the number of perfect matchings of a bipartite graph, something considered difficult, while the determinant can be computed by a polynomial size circuit.

6.1.1. The permanent can do things considered difficult. A standard problem in graph theory, for which the only known algorithms are exponential in the size of the graph, is to count the number of perfect matchings of a bipartite graph, that is, a graph with two sets of vertices and edges only joining vertices from one set to the other.

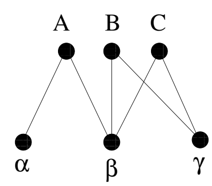


Figure 6.1.1. A bipartite graph, Vertex sets are $\{A, B, C\}$ and $\{\alpha, \beta, \gamma\}$.

A *perfect matching* is a subset of the edges such that each vertex shares an edge from the subset with exactly one other vertex.

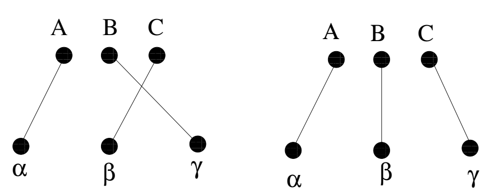


Figure 6.1.2. Two perfect matchings of the graph from Figure 6.1.1.

To a bipartite graph one associates an incidence matrix x^i_j, where $x^i_j = 1$ if an edge joins the vertex i above to the vertex j below and is zero otherwise.

For example the graph above has incidence matrix

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

A perfect matching corresponds to a matrix constructed from the incidence matrix by setting some of the entries to zero so that the resulting matrix has exactly one 1 in each row and column, i.e., is a matrix obtained by applying a permutation to the columns of the identity matrix.

Exercise 6.1.1.1: (1) Show that if x is the incidence matrix of a bipartite graph, then $\mathrm{perm}_n(x)$ indeed equals the number of perfect matchings.

For example, $\mathrm{perm}_3 \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = 2.$

Thus a classical problem: determine the complexity of counting the number of perfect matchings of a bipartite graph (which is complete for the complexity class $\sharp\mathbf{P}$, see [**BCS97**, p. 574]), can be studied via algebra - determine the complexity of evaluating the permanent.

6.1.2. Circuits.

Definition 6.1.2.1. An *arithmetic circuit* $\mathcal{C}$ is a finite, directed, acyclic graph with vertices of in-degree 0 or 2 and exactly one vertex of out-degree 0. The vertices of in-degree 0 are labeled by elements of $\mathbb{C} \cup \{x_1, \dots, x_n\}$, and called *inputs*. Those of in-degree 2 are labeled with $+$ or $*$ and are called *gates*. If the out-degree of v is 0, then v is called an *output gate*. The *size* of $\mathcal{C}$ is the number of edges.

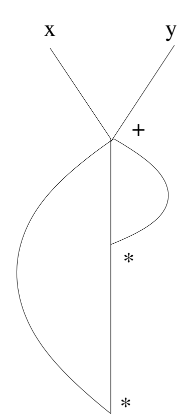


Figure 6.1.3. Circuit for $(x + y)^3$

To each vertex v of a circuit $\mathcal{C}$, associate the polynomial that is computed at v, which will be denoted $\mathcal{C}_v$. In particular the polynomial associated with the output gate is called the polynomial computed by $\mathcal{C}$.

At first glance, circuits do not look geometrical, as they depend on a choice of coordinates. While computer scientists always view polynomials as being given in some coordinate expression, in geometry one is interested in properties of objects that are independent of coordinates. These perspectives are compatible because with circuits one is not concerned with the precise size of a circuit, but its size up to, e.g., a polynomial factor. Reducing the size at worst by a polynomial factor, we can think of the inputs to our circuits as arbitrary affine linear or linear functions on a vector space.

6.1.3. Arithmetic circuits and complexity classes.

Definition 6.1.3.1. Let $d(n), N(n)$ be polynomials and let $f_n \in \mathbb{C}[x_1, \dots, x_{N(n)}]_{\leq d(n)}$ be a sequence of polynomials. We say $(f_n) \in \mathbf{VP}$ if there exists a sequence of circuits $\mathcal{C}_n$ of size polynomial in n computing f_n.

Often the phrase "there exists a sequence of circuits $\mathcal{C}_n$ of size polynomial in n computing f_n" is abbreviated "there exists a polynomial sized circuit computing (f_n)".

The class **VNP**, which consists of sequences of polynomials whose coefficients are "easily" described, has a more complicated definition:

Definition 6.1.3.2. A sequence (f_n) is in **VNP** if there exists a polynomial p and a sequence $(g_n) \in \mathbf{VP}$ such that

$$f_n(x) = \sum_{\epsilon \in \{0,1\}^{p(n)}} g_n(x, \epsilon).$$

One may think of the class **VP** as a bundle over **VNP** where elements of **VP** are thought of as sequences of maps, say $g_n : \mathbb{C}^{N(n)} \to \mathbb{C}$, and elements of **VNP** are projections of these maps by eliminating some of the variables by averaging or "integration over the fiber". In algebraic geometry, it is well known that projections of varieties can be far more complicated than the original varieties. See [**Bas14**] for more on this perspective.

Definition 6.1.3.3. One says that a sequence $(g_m(y_1, \dots, y_{M(m)}))$ can be (polynomially) *reduced* to $(f_n(x_1, \dots, x_{N(n)}))$ if there exists a polynomial $n(m)$ and affine linear functions $X_1(y_1, \dots, y_M), \dots, X_N(y_1, \dots, y_M)$ such that $g_m(y_1, \dots, y_{M(m)}) = f_n(X_1(y), \dots, X_{N(n)}(y))$. A sequence (p_n) is *hard* for a complexity class **C** if (p_n) can be reduced to every $(f_m) \in \mathbf{C}$, and it is *complete* for **C** if furthermore $(p_n) \in \mathbf{C}$.

Exercise 6.1.3.4: (1) Show that every polynomial of degree d can be reduced to x^d.

Theorem 6.1.3.5. *[Valiant]* [**Val79b**] (perm_m) *is complete for* **VNP**.

There are many excellent expositions of the proof, see, e.g. [**BCS97**] or [**Gat87**].

Thus Conjecture 1.2.1.1 is equivalent to:

Conjecture 6.1.3.6. *[Valiant]*[**Val79b**] *There does not exist a polynomial size circuit computing the permanent.*

Now for the determinant:

Proposition 6.1.3.7. $(\det_n) \in$ **VP**.

Remark 6.1.3.8. $\det_n$ would be **VP** complete if $\mathrm{dc}(p_m)$ grew no faster than a polynomial for all sequences $(p_m) \in$ **VP**.

One can compute the determinant quickly via Gaussian elimination: one uses the group to put a matrix in a form where the determinant is almost effortless to compute (the determinant of an upper triangular matrix is just the product of its diagonal entries). However this algorithm as presented is not a circuit (there are divisions and one needs to check if pivots are zero). After a short detour on symmetric polynomials, I prove Proposition 6.1.3.7 in §6.1.5.

6.1.4. Symmetric polynomials. An ubiquitous class of polynomials are the *symmetric polynomials*: let $\mathfrak{S}_N$ act on $\mathbb{C}^N$ by permuting basis elements, which induces an action on the polynomial ring $\mathbb{C}[x_1, \ldots, x_N]$. Let $\mathbb{C}[x_1, \ldots, x_N]^{\mathfrak{S}_N}$ denote the subspace of polynomials invariant under this action. What follows are standard facts and definitions about symmetric functions. For proofs, see, e.g., [**Mac95**, §I.2].

The *elementary symmetric functions* (or elementary symmetric polynomials) are

$$e_n = e_{n,N} = e_n(x_1, \ldots, x_N) := \sum_{J \subset [N] | |J| = n} x_{j_1} \cdots x_{j_n}. \tag{6.1.1}$$

If the number of variables is understood, I write e_n for $e_{n,N}$. They generate the ring of symmetric polynomials. They have the *generating function*

$$E_N(t) := \sum_{k \geq 0} e_k(x_1, \ldots, x_N) t^k = \prod_{i=1}^{N} (1 + x_i t). \tag{6.1.2}$$

Exercise 6.1.4.1: (1) Verify the coefficient of t^n in $E_N(t)$ is $e_{n,N}$.

The *power sum symmetric functions* are

$$p_n = p_{n,N} = p_{n,N}(x_1, \ldots, x_N) = x_1^n + \cdots + x_N^n. \tag{6.1.3}$$

They also generate the ring of symmetric polynomials. They have the generating function

$$P_N(t) = \sum_{k\geq 1} p_k t^{k-1} = \frac{d}{dt}\ln[\prod_{j=1}^{N}(1-x_j t)^{-1}]. \tag{6.1.4}$$

Exercise 6.1.4.2: (2) Verify that the coefficient of t^n in $P_N(t)$ is indeed $p_{n,N}$. ⊚

Exercise 6.1.4.3: (2) Show that

$$P_N(-t) = -\frac{E_N'(t)}{E_N(t)}. \circledcirc \tag{6.1.5}$$

Exercise 6.1.4.3, together with a little more work (see, e.g. [**Mac95**, p. 28]) shows that

$$p_n = \det{}_n \begin{pmatrix} e_1 & 1 & 0 & \cdots & 0 \\ 2e_2 & e_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & & 1 \\ ne_n & e_{n-1} & e_{n-2} & \cdots & e_1 \end{pmatrix}. \tag{6.1.6}$$

Similarly

$$e_n = \frac{1}{n!}\det{}_n \begin{pmatrix} p_1 & 1 & 0 & \cdots & 0 \\ p_2 & p_1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{n-1} & p_{n-2} & & \cdots & n-1 \\ p_n & p_{n-1} & & \cdots & p_1 \end{pmatrix}. \tag{6.1.7}$$

6.1.5. Proof of Proposition 6.1.3.7. Here is a construction of a small circuit for the determinant that appeared in [**Csa76**]:

The determinant of a linear map $f : V \to V$ is the product of its eigenvalues $\lambda_1, \dots, \lambda_{\mathbf{v}}$, i.e., $e_{\mathbf{v}}(\lambda) = \lambda_1 \cdots \lambda_{\mathbf{v}}$.

On the other hand, recall that trace(f) is the sum of the eigenvalues of f, and more generally, letting f^k denote the composition of f with itself k times,

$$\mathrm{trace}(f^k) = p_k(\lambda) = \lambda_1^k + \cdots + \lambda_{\mathbf{v}}^k.$$

The quantities trace(f^k) can be computed with small circuits.

Exercise 6.1.5.1: (2) Write down a circuit for the polynomial $A \mapsto \mathrm{trace}(A^2)$ when A is an $n\times n$ matrix with variable entries.

Thus we can compute $\det_n$ via small circuits and (6.1.7). While (6.1.7) is still a determinant, it is almost lower triangular and its naïve computation, e.g., with Laplace expansion, can be done with an $O(n^3)$-size circuit and the full algorithm for computing $\det_n$ can be executed with an $O(n^4)$ size circuit.

Remark 6.1.5.2. A more restrictive class of circuits are *formulas* which are circuits that are trees. Let $\mathbf{VP}_e$ denote the sequences of polynomials that admit a polynomial size formula. The circuit in the proof above is not a formula because results from computations are used more than once. It is known that the determinant admits a *quasi-polynomial* size formula, that is, a formula of size $n^{O(\log n)}$, and it is complete for the complexity class $\mathbf{VQP} = \mathbf{VP}_s$ consisting of sequences of polynomials admitting a quasi-polynomial size formula see, e.g., [**BCS97**, §21.5] (or equivalently, a polynomial sized "skew" circuit, see ****). It is not known whether or not the determinant is complete for $\mathbf{VP}$.

6.1.6. The Geometric Complexity Theory (GCT) variant of Valiant's conjecture. Recall that when we used polynomials in the study of matrix multiplication, we were actually proving lower bounds on tensor border rank rather than tensor rank. In the case of matrix multiplication, at least as far as the exponent was concerned, this changed nothing. In the case of determinant versus permanent, *it is not known* if using polynomial methods actually leads to a stronger separation of complexity classes. In any case, it will be best to make additional definitions to clarify the two different types of lower bounds.

I also need to address something I swept under the rug earlier: we are looking for polynomials on spaces of polynomials. When the number of variables changes, we can still use the "same" polynomials (by including the smaller space in the larger), but a more serious problem occurs regarding degree, which necessitates the introduction of *padding*. Recall that $\mathrm{End}(V)$ acts on S^nV by $X\cdot(x_{i_1}\cdots x_{i_n}) = (Xx_{i_1})\cdots(Xx_{i_n})$ and extend linearly.

Conjecture 6.1.6.1. *[Rephrasing of Valiant's conjecture] Let ℓ be a linear coordinate on $\mathbb{C}^1$ and consider any linear inclusion $\mathbb{C}^1\oplus\mathbb{C}^{m^2}\to\mathbb{C}^{n^2}$, so in particular $\ell^{n-m}\,\mathrm{perm}_m\in S^n\mathbb{C}^{n^2}$. Let $n(m)$ be a polynomial. Then for all sufficiently large m,*

$$[\ell^{n-m}\,\mathrm{perm}_m]\not\in \mathrm{End}(\mathbb{C}^{n^2})\cdot[\det_{n(m)}].$$

To see the equivalence of the formulations, if $\mathrm{perm}(y^i_j) = \det_n(\Lambda + \sum_{i,j}A_{ij}y_{i,j})$, then $\ell^{n-m}\,\mathrm{perm}_m(y_{i,j}) = \det_n(\ell\Lambda + \sum_{i,j}A_{ij}y_{i,j})$. Such an expression is equivalent to setting each entry of the $n\times n$ matrix to a linear combination of the variables $\ell, y_{i,j}$, which is precisely what the elements of rank m^2+1 in $\mathrm{End}(\mathbb{C}^{n^2})$ can accomplish. Moreover $\ell^{n-m}\,\mathrm{perm}_m = X\cdot\det_{n(m)}$ for some $X\in\mathrm{End}(\mathbb{C}^{n^2})$ implies X has rank m^2+1.

In order to use more tools from algebraic geometry and representation theory to separate complexity classes, the following conjecture appeared in [**MS01**]:

Conjecture 6.1.6.2. [**MS01**] *Let ℓ be a linear coordinate on $\mathbb{C}^1$ and consider any linear inclusion $\mathbb{C}^1 \oplus \mathbb{C}^{m^2} \to \mathbb{C}^{n^2}$, so in particular $\ell^{n-m}\operatorname{perm}_m \in S^n\mathbb{C}^{n^2}$. Let $n(m)$ be a polynomial. Then for all sufficiently large m,*

$$[\ell^{n-m}\operatorname{perm}_m] \notin \overline{GL_{n^2}\cdot[\det_{n(m)}]}.$$

Note that $\overline{GL_{n^2}\cdot[\det_n]} = \overline{\operatorname{End}(\mathbb{C}^{n^2})\cdot[\det_n]}$ so this is a strengthening of Conjecture 6.1.6.1. It will be useful to rephrase the conjecture slightly, to highlight that it is a question about determining whether one orbit closure is contained in another. Let

$$\mathcal{D}et_n := \overline{GL_{n^2}\cdot[\det_n]},$$

and let

$$\mathcal{P}erm_n^m := \overline{GL_{n^2}\cdot[\ell^{n-m}\operatorname{perm}_m]}.$$

Conjecture 6.1.6.3. [**MS01**] *Let $n(m)$ be a polynomial. Then for all sufficiently large m,*

$$\mathcal{P}erm_{n(m)}^m \not\subset \mathcal{D}et_{n(m)}.$$

The equivalence of Conjectures 6.1.6.3 and 6.1.6.2 follows as $\ell^{n-m}\operatorname{perm}_m \notin \mathcal{D}et_n$ implies $GL_{n^2}\cdot\ell^{n-m}\operatorname{perm}_m \not\subset \mathcal{D}et_n$, and since $\mathcal{D}et_n$ is closed and both sides are irreducible, there is no harm in taking closure on the left hand side, as you showed in Exercise 3.3.1.1.

Now the goal is clear: both varieties are *invariant* under GL_{n^2} so their ideals will be GL_{n^2}-modules, as was mentioned in §1.1.13. We look for a GL_{n^2}-module M such that $M \subset I[\mathcal{D}et_n]$ and $M \not\subset I[\mathcal{P}erm_n^m]$.

In §8.8 I explain the original program to solve this conjecture. Although that program cannot work as stated, I believe that the re-focusing of a problem of separating complexity classes to questions in algebraic geometry and representation theory as they proposed, is the most viable path to resolving Valiant's conjecture.

6.2. Flattenings: our first polynomials on the space of polynomials

In this section I discuss the most classical polynomials on the space of polynomials, that were first introduced by Sylveseter in 1852 and called *catalecticants* by him. They are also called *flattenings* and in the computer science literature the polynomials induced by the *method of partial derivatives.*

6.2.1. Three perspectives on $S^d\mathbb{C}^M$. I review our perspectives on $S^d\mathbb{C}^M$. We have seen $S^d\mathbb{C}^M$ is the space of symmetric tensors in $(\mathbb{C}^M)^{\otimes d}$. Given a symmetric tensor $T \in S^d\mathbb{C}^M$, we may form a polynomial P_T on $\mathbb{C}^{M*}$ by, for $v \in \mathbb{C}^{M*}$, $P_T(v) := T(v, \ldots, v)$. I use this identification repeatedly without further mention.

One can also recover T from P_T via *polarization*. Then (up to universal constants) $T(v_{i_1}, \ldots, v_{i_M})$ where $1 \leq i_1 \leq \cdots \leq i_M$ is the coefficient of $t_{i_1} \cdots t_{i_M}$ in $P_T(t_1v_1 + \cdots + t_Mv_M)$. See [**Lan12**, Chap. 2] for details.

As was mentioned in Exercise 2.3.0.4, we may also think of $S^d\mathbb{C}^M$ as the space of homogeneous differential operators of order d on $Sym(\mathbb{C}^{M*}) := \oplus_{j=0}^{\infty} S^j\mathbb{C}^{M*}$.

Thus we may view an element of $S^d\mathbb{C}^M$ as a homogeneous polynomial of degree d on $\mathbb{C}^{M*}$, a symmetric tensor, and as a homogeneous differential operator of order d on the space of polynomials $Sym(\mathbb{C}^{M*})$.

6.2.2. Catalecticants, a.k.a. The method of partial derivatives. Now would be a good time to read §3.1 if you have not already done so. I review a few essential points from it.

The simplest polynomials in $S^n\mathbb{C}^N$ are just the n-th powers of linear forms. Their zero set is a hyperplane (counted with muliplicity n). Let $P \in S^n\mathbb{C}^N$. How can one test if P is an n-th power of a linear form, $P = \ell^n$ for some $\ell \in \mathbb{C}^N$?

Exercise 6.2.2.1: (1!) Show that $P = \ell^n$ for some $\ell \in \mathbb{C}^N$ if and only if $\dim\langle \frac{\partial P}{\partial x^1}, \ldots, \frac{\partial P}{\partial x^N} \rangle = 1$, where $x^1, \ldots, x^N$ are coordinates on $\mathbb{C}^N$.

Note that Exercise 6.2.2.1 is indeed a polynomial test: The dual space $\mathbb{C}^{N*}$ may be considered as the space of first order homogeneous differential operators on $S^n\mathbb{C}^N$, and the test is that the 2×2 minors of the map $P_{1,n-1} : \mathbb{C}^{N*} \to S^{n-1}\mathbb{C}^N$, given by $\frac{\partial}{\partial x^j} \mapsto \frac{\partial P}{\partial x^j}$ are zero.

Exercise 6.2.2.1 may be phrased without reference to coordinates: recall the inclusion $S^nV \subset V \otimes S^{n-1}V = \mathrm{Hom}(V^*, S^{n-1}V)$. For $P \in S^nV$, write $P_{1,n-1} \in \mathrm{Hom}(V^*, S^{n-1}V)$. We may interpret Exercise 6.2.2.1 as saying that P is an n-th power of a linear form if and only if $\mathrm{rank}(P_{1,n-1}) = 1$.

Recall that the *n-th Veronese variety* is

$$v_n(\mathbb{P}V) := \{[P] \in \mathbb{P}S^nV \mid P = \ell^n \text{ for some } \ell \in V\} \subset \mathbb{P}(S^nV).$$

Exercise 6.2.2.1 shows that the Veronese variety is indeed an algebraic variety and by definition, it is invariant under the action of $GL(V)$ on $\mathbb{P}S^nV$. In fact it is homogenous - a single $GL(V)$-orbit.

More generally define the *subspace variety*

$$Sub_k(S^nV) := \mathbb{P}\{P \in S^nV \mid \mathrm{rank}(P_{1,n-1}) \leq k\}.$$

Note that $[P] \in Sub_k(S^nV)$ if and only if there exists a coordinate system where P can be expressed using only k of the $\dim V$ variables. The subspace variety $Sub_k(S^nV) \subset \mathbb{P}S^nV$ has the geometric interpretation as the polynomials whose zero sets in projective space are *cones* with a $\mathbf{v}-k$ dimensional vertex. (In affine space the zero set looks like a cylinder, such as the surface $x^2+y^2=1$ in $\mathbb{R}^3$.) If $[P] \in Sub_k(S^nV)$, then there exist linear coordinates on V such that the expression of P only involves at most k of the coordinates. Consider the hypersurface $X_P \subset \mathbb{P}^{k-1}$ cut out by restricting P to these variables. Then points of $Z(P) \subset \mathbb{P}V^*$ are of the form $[x+y]$ where $x \in \hat{X}_P$ and $y \in \mathbb{P}^{\mathbf{v}-k-1}$ is any point in the complementary space. See §6.4.2 for more details. Equations for $Sub_k(S^nV)$ are the size $k+1$ minors of

$$P_{1,n-1} : V^* \to S^{n-1}V.$$

In fact these generate the ideal, see §8.4.1.

The *symmetric rank* of $P \in S^nV^*$ is $\mathbf{R}_{v_n(\mathbb{P}V)}(P) = \mathbf{R}_S(P)$, the smallest r such that $P = \ell_1^n + \cdots + \ell_r^n$ for $\ell_j \in V$. The *symmetric border rank* of P is $\underline{\mathbf{R}}_{v_n(\mathbb{P}V)}(P) = \underline{\mathbf{R}}_S(P)$, which, in the language of §4.8.1, is the smallest r such that $[P] \in \sigma_r(v_n(\mathbb{P}V))$, the r-th secant variety of the Veronese variety. Symmetric rank will appear naturally in the study of Valiant's conjecture and its variants. In the language of circuits introduced in Chapter 7, it is (essentially) the size of the smallest homogeneous $\Sigma\Lambda\Sigma$-circuit computing P.

How would one test if P is the sum of two n-th powers, $P = \ell_1^n + \ell_2^n$ for some $\ell_1, \ell_2 \in \mathbb{C}^N$?

Exercise 6.2.2.2: (1) Show that $P = \ell_1^n + \ell_2^n$ for some $\ell_j \in \mathbb{C}^N$ implies $\dim \operatorname{span}\{\frac{\partial P}{\partial x^1}, \dots, \frac{\partial P}{\partial x^N} \mid 1 \leq i,j \leq N\} \leq 2$.

Exercise 6.2.2.3: (2) Show that any polynomial vanishing on all polynomials of the form $P = \ell_1^n + \ell_2^n$ for some $\ell_j \in \mathbb{C}^N$ also vanishes on $x^{n-1}y$. ⊚

Exercise 6.2.2.3 reminds us that $\sigma_2(v_n(\mathbb{P}V))$ also includes points on tangent lines.

The condition in Exercise 6.2.2.2 is not sufficient to determine membership in $\sigma_2(v_n(\mathbb{P}V))$, in other words, $\sigma_2(v_n(\mathbb{P}V)) \subsetneq Sub_2(S^nV)$: Consider $P = \ell_1^{n-2}\ell_2^2$. It has $\operatorname{rank}(P_{1,n-1}) = 2$ but $P \notin \sigma_2(v_n(\mathbb{P}V))$ as can be seen by the following exercises:

Exercise 6.2.2.4: (1) Show that $P = \ell_1^n + \ell_2^n$ for some $\ell_j \in \mathbb{C}^N$ implies $\dim \operatorname{span}\{\frac{\partial^2 P}{\partial x^i \partial x^j}\} \leq 2$.

Exercise 6.2.2.5: (1) Show that $P = \ell_1^{n-2}\ell_2^2$ for some distinct $\ell_j \in \mathbb{C}^N$ implies $\dim \operatorname{span}\{\frac{\partial^2 P}{\partial x^i \partial x^j}\} > 2$.

Let $P_{2,n-2} : S^2\mathbb{C}^{N*} \to S^{n-2}\mathbb{C}^N$ denote the map with image $\langle \frac{\partial^2 P}{\partial x^i \partial x^j}\rangle$. Vanishing of the size three minors of $P_{1,n-1}$ and $P_{2,n-2}$ are necessary and sufficient conditions for $P \in \sigma_2(v_n(\mathbb{P}V))$, as was shown by Gundelfinger in 1886 [**Gun**].

More generally, one can consider the polynomials given by the minors of the maps $S^k\mathbb{C}^{N*} \to S^{n-k}\mathbb{C}^N$, given by $D \mapsto D(P)$. Write these maps as $P_{k,n-k} : S^kV^* \to S^{n-k}V$. These equations date back to Sylvester [**Syl52**] and are called the *method of partial derivatives* in the complexity literature, e.g. [**CKW10**]. The ranks of these maps gives a complexity measure on polynomials.

Exercise 6.2.2.6: (1!) What does the method of partial derivatives tell us about the complexity of $x_1 \cdots x_n$, $\det_n$ and perm_n, e.g., taking $k = \lfloor \frac{n}{2} \rfloor$? ⊚

Exercise 6.2.2.6 provides an exponential lower bound for the permanent in the complexity measure of symmetric border rank $\underline{\mathbf{R}}_S$, but we obtain the *same* lower bound for the determinant. Thus this measure will not be useful for separating the permanent from the determinant. It still gives interesting information about other polynomials such as symmetric functions, which we will examine.

The variety of homogeneous polynomials of degree n that are products of linear forms will also play a role in complexity theory. Recall the *Chow variety* of polynomials that decompose into a product of linear forms from §3.1.2:

$$Ch_n(V) := \mathbb{P}\{P \in S^nV \mid P = \ell_1 \cdots \ell_n \text{ for } \ell_j \in V\}.$$

One can define a complexity measure for writing a polynomial as a sum of products of linear forms. The "Zariski closed" version of this condition is membership in $\sigma_r(Ch_n(V))$. In the language of circuits, $\mathbf{R}_{Ch_n(V)}(P)$ is (essentially) the size of the smallest homogeneous $\Sigma\Pi\Sigma$ circuit computing a polynomial P. I discuss this in §7.1.

Exercise 6.2.2.6 gives a necessary test for a polynomial $P \in S^n\mathbb{C}^N$ to be a product of n linear forms, namely $\mathrm{rank}(P_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}) \le \binom{n}{\lfloor \frac{n}{2} \rfloor}$. A question to think about: how would one develop a necessary and sufficient condition to show a polynomial $P \in S^n\mathbb{C}^N$ is a product of n linear forms? See §9.1.6 for an answer.

Unfortunately we have very few techniques for finding good spaces of polynomials on polynomials. One such that generalizes flattenings, called *Young flattenings* is discussed in §8.2.

A natural question is whether or not all flattenings are non-trivial. I address this in §6.2.4 below after defining *conormal spaces*, which will be needed for the proof.

6.2.3. Conormal spaces. Recall the definition of the tangent space to a point on a variety $X \subset \mathbb{P}V$ or $X \subset V$, $\hat{T}_x X \subset V$, from §3.1.3. The *conormal space* $N^*_x X \subset V^*$ is simply defined to be the annhilator of the tangent space: $N^*_x X = (\hat{T}_x X)^\perp$.

Exercise 6.2.3.1: (2!) Show that in $\hat{\sigma}^0_r(Seg(\mathbb{P}^{u-1} \times \mathbb{P}^{v-1}))$, the space of $u \times v$ matrices of rank r,

$$\hat{T}_M \sigma^0_r(Seg(\mathbb{P}^{u-1} \times \mathbb{P}^{v-1})) = \{X \in Mat_{u\times v} \mid X \ker(M) \subset \mathrm{Image}(M)\}.$$

Give a description of $N^*_M \sigma^0_r(Seg(\mathbb{P}^{u-1} \times \mathbb{P}^{v-1}))$. ⊚

6.2.4. All flattenings give non-trivial equations. One idea to show all flattenings give non-trivial equations would be to find some explicit polynomial for which they are all of maximal rank. Unfortunately this is an open question:

Problem 6.2.4.1. Find an explicit sequence of polynomials $P_{d,n} \in S^d\mathbb{C}^n$, such that for all $1 \le k \le \lfloor \frac{d}{2} \rfloor$, $\mathrm{rank}((P_{d,n})_{k,d-k}) = \binom{n+k-1}{k}$.

Exercise 6.2.4.2: (1) Show that if $P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil}$ is of maximal rank, then all $P_{k,d-k}$ are of maximal rank.

Theorem 6.2.4.3. [**Gre78, IE78**] *For a general polynomial $P \in S^dV$, all the maps $P_{k,d-k} : S^kV^* \to S^{d-k}V$ are of maximal rank.*

Proof. (This proof is adapted from [**IK99**].) By Exercise 6.2.4.2 it is sufficient to consider the case $k = \lfloor \frac{d}{2} \rfloor$. For each $0 \le t \le \binom{\mathbf{v}+\lfloor \frac{d}{2} \rfloor -1}{\lfloor \frac{d}{2} \rfloor}$, let

$$Gor(t) := \{P \in S^dV \mid \mathrm{rank} P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil} = t\}.$$

("Gor" is after Gorenstein, see [**IK99**].) Note that $S^dV = \sqcup_t Gor(t)$. Since this is a finite union there must be at least one (and exactly one by semi-continuity) t_0 such that $\overline{Gor(t_0)} = S^dV$. We want to show that $t_0 = \binom{\mathbf{v}+\lfloor \frac{d}{2} \rfloor -1}{\lfloor \frac{d}{2} \rfloor}$. We will do this by computing conormal spaces as we must have $N^*_P Gor(t_0) = 0$ for $P \in Gor(t_0)$. Now, for any t, the subspace $N^*_P Gor(t) \subset S^dV$ satisfies

$$N^*_P Gor(t) \subset N^*_{P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil}} \sigma_t = N^*_{P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil}} \sigma_t(Seg(\mathbb{P}S^{\lfloor \frac{d}{2} \rfloor}V \times S^{\lceil \frac{d}{2} \rceil}V)) \subset S^{\lfloor \frac{d}{2} \rfloor}V \otimes S^{\lceil \frac{d}{2} \rceil}V,$$

and $N^*_P Gor(t)$ is simply the image of $N^*_{P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil}} \sigma_t$ under the symmetrization map to S^dV^*. On the other hand, by Exercise 6.2.3.1, $N^*_{P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil}} \sigma_t = \ker P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil} \otimes \ker P_{\lceil \frac{d}{2} \rceil, \lfloor \frac{d}{2} \rfloor}$. In order for $N^*_P Gor(t)$ to be zero, we need $N^*_{P_{\lfloor \frac{d}{2} \rfloor, \lceil \frac{d}{2} \rceil}} \sigma_t$ to be zero (otherwise there will be something nonzero in the image of the symmetrization map: if d is odd, the two degrees are different and this is clear. If d is even, the conormal space is the tensor product of a vector space with itself), which implies $\ker P_{\lceil \frac{d}{2}\rceil,\lfloor \frac{d}{2}\rfloor} = 0$ □

Note that the maximum symmetric border rank (in all but a few known exceptions) is $\frac{1}{\mathbf{v}}\binom{\mathbf{v}+d-1}{d}$, whereas flattenings only give equations up to symmetric border rank $\binom{\mathbf{v}+\lfloor \frac{d}{2}\rfloor-1}{\lfloor \frac{d}{2}\rfloor}$.

6.2.5. Jacobian varieties. While the ranks of symmetric flattenings are the same for the permanent and determinant, by looking more closely at the maps, we can extract geometric information that distinguishes them.

First, for $P \in S^nV$, consider the images $P_{k,n-k}(S^kV^*) \subset S^{n-k}V$. This is a space of polynomials and we can consider the common zero set of these polynomials, the *k-th Jacobian variety of P*:

$$Z(P)_{Jac,k} := \{[\alpha] \in \mathbb{P}V^* \mid q(\alpha) = 0 \ \forall q \in P_{k,n-k}(S^kV^*)\}.$$

It is easy to see that $Z(\det_n)_{Jac,k}$ is simply $\sigma_{n-k-1}(Seg(\mathbb{P}^{n-1}\times\mathbb{P}^{n-1}))$, the matrices of rank at most $n-k-1$. It is *not known* what the varieties $Z(\mathrm{perm}_n)_{Jac,k}$ are in general. I explicitly determine $Z(\mathrm{perm}_n)_{Jac,n-2}$ in §6.3.2 below as it is used to prove the symmetries of the permanent are what we expect them to be.

In §**??** I discuss further information that one can extract from the image of $P_{k,n-k}$.

6.3. Singular loci

As mentioned above, the geometry of the hypersurfaces $Z(\det_n)$ and $Z(\mathrm{perm}_m)$ will aid us in comparing the complexity of the determinant and permanent. A simple invariant that will be useful is the dimension of the singular set of a hypersurface. We will need a more subtle definition than that presented in §3.1.3. This new dimension is **upper?** semi-continuous under degenerations of polynomials.

6.3.1. Definition of the (scheme theoretic) singular locus.

Definition 6.3.1.1. Say a variety $X = \{P_1 = 0, \dots, P_s = 0\} \subset \mathbb{P}V$ has codimension c, using the definition of codimension in §dimsubsect. Then $x \in X$ is a *singular point* if $dP_{1,x}, \dots, dP_{s,x}$ fail to span a space of dimension c. Let $X_{sing} \subset X$ denote the singular points of X. In particular, if $X = Z(P)$ is a hypersuface and $x \in X$, then $x \in X_{sing}$ if and only if $dP_x = 0$. Note that X_{sing} is also the zero set of a collection of polynomials.

Warning: This definition is a property of the ideal generated by the polynomials $P_1, \dots, P_s$, not of X as a set. For example every point of $(x_1^2 +$

$\cdots + x_n^2)^2 = 0$ is a singular point. In the language of algebraic geometry, one refers to the singular point of the *scheme* associated to $\{P_1 = 0, \dots, P_s = 0\}$.

"Most" hypersurfaces $X \subset \mathbb{P}V$ are smooth, in the sense that $\{P \in \mathbb{P}S^dV \mid Z(P)_{sing} \neq \emptyset\} \subset \mathbb{P}S^dV$ is a hypersurface, as proven below in §**??**. The size of the singular locus of $Z(P)$ is a measure of the pathology of P.

Singular loci will also be used in the determination of symmetry groups.

6.3.2. Singularities of $Z(\mathrm{perm}_m)$. In contrast to the determinant, the singular set of the permanent is not understood, even its codimension is not known! The problem is more difficult because unlike in the determinant case, we do not have normal forms for points on $Z(\mathrm{perm}_m)$.

Exercise 6.3.2.1: (1!) Show that the permanent admits a "Laplace type" expansion similar to that of the determinant.

Exercise 6.3.2.1 implies:

Proposition 6.3.2.2. *$Z(\mathrm{perm}_m)_{sing}$ consists of the $m \times m$ matrices with the property that all size $m-1$ sub-matrices of it have permanent zero.*

Exercise 6.3.2.3: (1) Show that $Z(\mathrm{perm}_m)_{sing}$ has codimension at most $2m$ in $\mathbb{C}^{m^2}$. ⊚

Since $Z(\mathrm{perm}_2)_{sing} = \emptyset$, let's start with perm_3. Since we will need it later, I prove a more general result:

Lemma 6.3.2.4. *The variety $Z(\mathrm{perm}_m)_{Jac,m-2}$ is the union of the following varieties:*

(1) *Matrices A with all entries zero except those in a single size 2 submatrix, and that submatrix has zero permanent.*

(2) *Matrices A with all entries zero except those in the j-th row for some j.*

(3) *Matrices A with all entries zero except those in the j-th column for some j.*

In other words, let $X \subset Mat_m(\mathbb{C})$ denote the subvariety of matrices that are zero except in the upper 2×2 corner and that 2×2 submatrix has zero permanent, and let Y denote the variety of matrices that are zero except in the first row, then

$$Z(\mathrm{perm}_m)_{Jac,m-2} = \bigcup_{\sigma \in (\mathfrak{S}_m \times \mathfrak{S}_m) \rtimes \mathbb{Z}_2} \sigma \cdot X \cup \sigma \cdot Y. \tag{6.3.1}$$

The proof is straight-forward. Here is the main idea: Take a matrix with entries that don't fit that pattern, e.g., one that begins

$$\begin{matrix} a & b & e \\ * & d & * \end{matrix}$$

and note that it is not possible to fill in the two unknown entries and have all size two sub-permanents, even in this corner, zero. There are just a few such cases since we are free to act by $(\mathfrak{S}_m \times \mathfrak{S}_m) \rtimes \mathbb{Z}_2 \subset G_{\mathrm{perm}_m}$.

Corollary 6.3.2.5.

$$\{\mathrm{perm}_3 = 0\}_{sing} = \bigcup_{\sigma \in (\mathfrak{S}_3 \times \mathfrak{S}_3) \ltimes \mathbb{Z}_2} \sigma \cdot X \cup \sigma \cdot Y.$$

In particular, all the irreducible components of $\{\mathrm{perm}_3 = 0\}_{sing}$ *have the same dimension and and* $codim(\{\mathrm{perm}_3 = 0\}_{sing}, \mathbb{C}^9) = 6$.

This equidimensionality property already fails for perm_4: consider

$$\left\{ \begin{pmatrix} x^1_1 & x^1_2 & 0 & 0 \\ x^2_1 & x^2_2 & 0 & 0 \\ 0 & 0 & x^3_3 & x^3_4 \\ 0 & 0 & x^4_3 & x^4_4 \end{pmatrix} \mid x^1_1 x^2_2 + x^2_1 x^1_2 = 0,\ x^3_3 x^4_4 + x^4_3 x^3_4 = 0 \right\}$$

This defines a six dimensional irreducible component of $\{\mathrm{perm}_4 = 0\}_{sing}$ which is not contained in either a space of matrices with just two nonzero rows (or columns) or the set of matrices that are zero except for in some 3×3 submatrix which has zero permanent. In [**vzG87**] von zur Gathen states that all components of $\{\mathrm{perm}_4 = 0\}_{sing}$ are either of dimension six or eight.

Although we do not know the codimension of $Z(\mathrm{perm}_m)_{sing}$, the following estimate will suffice for the application of von zur Gathen's regularity theorem 6.3.3.1 below.

Proposition 6.3.2.6 (von zur Gathen [**vzG87**])**.**

$$codim(Z(\mathrm{perm}_m)_{sing}, \mathbb{C}^{m^2}) \geq 5.$$

Proof. I work by induction on m, the case $m = 2$ is ok as $Z(\mathrm{perm}_2)_{sing} = \emptyset$. Let I, J be multi-indices of the same size and let $sp(I|J)$ denote the subpermanent of the $(m - |I|, m - |I|)$ submatrix omitting the index sets (I, J). Let $C \subset Z(\mathrm{perm}_m)_{sing}$ be an irreducible component of the singular set. If $sp(i_1, i_2 | j_1, j_2)|_C = 0$ for all $(i_1, i_2 | j_1, j_2)$, we are done by induction as then $C \subset \cup Z(\mathrm{perm}_{m-1})_{sing}$ where the union is over all size $m-1$ submatrices. So assume there is at least one size $m - 2$ subpermanent that is not identically zero on C, without loss of generality assume it is $sp(m-1, m|m-1, m)$. We have, via permanental Laplace expansions,

$$\begin{aligned}0 &= sp(m,m)|_C\\ &= \sum_{j=1}^{m-2} x^j_{m-1} sp(i,m|m-1,m) + x^{m-1}_{m-1} sp(m-1,m|m-1,m)\end{aligned}$$

so on a Zariski open subset of C, x^{m-1}_{m-1} is a function of the m^2-4 variables x^s_t, $(s,t)\notin\{(m-1,m-1),(m-1,m),(m,m-1),(m,m)\}$, Similar expansions give us x^{m-1}_m, x^m_{m-1}, and x^m_m as functions of the other variables, so we conclude $\dim C\leq m^2-4$. We need to find one more nonzero polynomial that vanishes identically on C that does not involve the variables $x^{m-1}_{m-1}, x^m_{m-1}, x^{m-1}_m, x^m_m$ to obtain another relation and to conclude $\dim C\leq m^2-5$. Consider

$$\begin{aligned}&sp(m-1,m|m-1,m)sp(m-2,m) - sp(m-2,m|m-1,m)sp(m-1,m)\\ &\quad - sp(m-2,m-1|m-1,m)sp(m,m)\\ &= -2x^{m-2}_{m-1} sp(m-2,m-1|m-1,m)sp(m-2,m|m-1,m)\\ &\quad + \text{terms not involving } x^{m-2}_{m-1},\end{aligned}$$

where we obtained the second line by permanental Laplace expansions in the size $m-1$ subpermanents in the expression, and arranged things such that all terms with $x^{m-1}_{m-1}, x^m_{m-1}, x^{m-1}_m, x^m_m$ appearing cancel. Since this expression is a sum of terms divisible by size $m-1$ subpermanents, it vanishes identically on C. But $2x^{m-2}_{m-1}sp(m-2,m-1|m-1,m)sp(m-2,m|m-1,m)$ is not the zero polynomial, so the whole expression is not the zero polynomial. Thus we obtain another nonzero polynomial that vanishes identically on C and is independent of the previous four as it does not involve any of $x^{m-1}_{m-1}, x^m_{m-1}, x^{m-1}_m, x^m_m$. $\square$

Although one expects that in general $\operatorname{codim}(Z(\operatorname{perm}_m)_{sing})$ to be greater than 5, Proposition 6.3.2.6 is sufficient for the hypothesis of Proposition 6.3.3.1 below.

6.3.3. von zur Gathen's regularity theorem and its consequences for lower bounds.

Proposition 6.3.3.1 (von zur Gathen [**vzG87**], also see [**ABV15**]). *Let $M>4$, and let $P\in S^m\mathbb{C}^M$ satisfy $\operatorname{codim}(\{P=0\}_{sing},\mathbb{C}^M)\geq 5$. If $P=\det_n\circ\tilde{A}$, where $\tilde{A}=\Lambda+A:\mathbb{C}^M\to\mathbb{C}^{n^2}$ is an affine linear map with Λ constant and A linear, then* $\operatorname{rank}\Lambda=n-1$.

Proof. I first claim that if $\tilde{A}(y)\in Z(\det_n)_{sing}$ then $y\in Z(P)_{sing}$. To see this, note that for any $y\in\mathbb{C}^M$, the differential of P at y satisfies (by the chain rule)

$$dP|_y = d(\det_n \circ \tilde{A})|_y = A^T(d(\det_n)|_{\tilde{A}(y)}),$$

where I have used that $d(\det_n)|_{\tilde{A}(y)} \in T^*_{\tilde{A}(y)}\mathbb{C}^{n^2} \simeq \mathbb{C}^{n^2*}$ and $A^T : \mathbb{C}^{n^2*} \to \mathbb{C}^{M^*}$ is the transpose of the differential of $\tilde{A}$. In particular, if $d(\det_n)|_{\tilde{A}(y)} = 0$ then $dP_y = 0$, which is what we needed to show.

Now by Theorem 3.1.5.1, the set

$$\tilde{A}(\mathbb{C}^M) \cap Z(\det_n)_{sing} \subset \mathbb{C}^{n^2}$$

is either empty or of dimension at least $\dim(\tilde{A}(\mathbb{C}^M)) + \dim(Z(\det_n)_{sing}) - n^2 = M + (n^2-4) - n^2 = M-4$, so the same is true for $\tilde{A}^{-1}(\tilde{A}(\mathbb{C}^M) \cap Z(\det_n)_{sing})$. But this latter set is contained in $Z(P)_{sing}$, which is of dimension at most $M-5$, so we conclude it is empty.

Thus for all $y \in \mathbb{C}^M$, $\operatorname{rank}\tilde{A}(y) \geq n-1$. In particular $\operatorname{rank}\tilde{A}(0) \geq n-1$, but $\tilde{A}(0) = \Lambda$. Finally equality holds because if Λ had rank n, then $\det(\tilde{A}(\mathbb{C}^M))$ would have a constant term. □

Exercise 6.3.3.2: (1) Prove that any polynomial $p \in S^d\mathbb{C}^M$ with singular locus of codimension greater than four must have $\mathrm{dc}(p) > d$.

Proposition 6.3.3.3. [**Cai90**] *Let $F \subset Mat_n(\mathbb{C})$ be an affine linear subspace such that for all $X \in F$, $\operatorname{rank}(F) \geq n-1$. Then $\dim F \leq \binom{n+1}{2} + 1$.*

For the proof, see [**Cai90**]. Note that Proposition 6.3.3.3 is near optimal as consider F the identity matrix plus free variables in the strictly upper-triangular slots, which has dimension $\binom{n}{2}$.

Exercise 6.3.3.4: (2) Use Proposition 6.3.3.3 to show $\mathrm{dc}(\mathrm{perm}_m) \geq \sqrt{2}m$.

Exercise 6.3.3.5: (2) Let $Q \subset \mathbb{P}^{n+1}$ be a smooth quadric hypersurface of dimension n. Show that the maximum dimension of a linear projective space contained in Q is $\lfloor \frac{n}{2} \rfloor$. ⊚

Theorem 6.3.3.6 (Alper-Bogart-Velasco [**ABV15**])**.** *Let $P \in S^d\mathbb{C}^M$ with $d \geq 3$ and such that $codim(Z(P)_{sing}, \mathbb{C}^M) \geq 5$. Then $\mathrm{dc}(P) \geq codim(Z(P)_{sing}, \mathbb{C}^M) + 1$.*

Proof. Let $n = \mathrm{dc}(P)$. Say $P = \det_n \circ \tilde{A}$, with $\tilde{A} = \Lambda + A$. By Proposition 6.3.3.1, $\operatorname{rank}(\Lambda) = n-1$, and using $G_{\det_n}$, we may assume Λ is normalized to the matrix that is zero everywhere but the diagonal, where it has one's except in the $(1,1)$-slot where it is also zero. Expand $\det(\tilde{A}(y)) = p_0 + p_1 + \cdots + p_n$ as a sum of homogeneous polynomials. Since the right hand side equals P, we must have $p_j = 0$ for $j < d$. Then $p_0 = \det(\Lambda) = 0$ and $p_1 = A^1_1$. Now $p_2 = \sum_{i=2}^n A^1_i A^i_1 = 0$ and more generally, each p_j is a sum of monomials, each of which contains an element in the first column and an element in the first row of A. Each A^i_j is a linear form on $\mathbb{C}^M$ and as such, we can consider the intersection of their kernels. Write $\Gamma = \cap_{i=1}^{n-1}(\ker A^i_1) \cap (\ker A^1_i)$. Then $\Gamma \subset Z(P)_{sing}$. Consider the A^1_i, A^j_1 as coordinates on $\mathbb{C}^{2(n-1)}$, p_2 defines a smooth quadric hypersurface in $\mathbb{P}^{2(n-1)-1}$. By Exercise 6.3.3.5, the maximum dimension of a linear space on such a quadric is $n-1$, so the rank of the linear map $\mathbb{C}^M \to \mathbb{C}^{2(n-1)}$ given by $y \mapsto (A^1_i(y), A^j_1(y))$ is at most $n-1$. But Γ is the kernel of this map. We have

$$n - 1 \geq \mathrm{codim}\Gamma \geq \mathrm{codim}(Z(P)_{sing}, \mathbb{C}^M)$$

and recalling that $n = \mathrm{dc}(P)$ we conclude. □

Exercise 6.3.3.7: (2) Prove that $\mathrm{codim}(\mathrm{perm}_m) = 2m$ when $m = 3, 4$.

Corollary 6.3.3.8. [**ABV15**] $\mathrm{dc}(\mathrm{perm}_3) = 7$ *and* $\mathrm{dc}(\mathrm{perm}_4) \geq 9$.

The upper bound for $\mathrm{dc}(\mathrm{perm}_3)$ is from (1.2.3).

Remark 6.3.3.9. Even if one could prove $\mathrm{codim}(\mathrm{perm}_m) = 2m$ for all m, the above theorem would only give a linear bound on $\mathrm{dc}(\mathrm{perm}_m)$. This bound would be obtained from taking one derivative. In the next section, I show that taking two derivatives, one can get a quadratic bound. Unfortunately, taking three derivatives does not appear to improve the situation further.

6.4. Geometry and the state of the art regarding dc(perm_m)

In mathematics, one often makes transforms to reorganize information, such as the Fourier transform. There are geometric transforms to "reorganize" the information in an algebraic variety. Taking the Gauss image (dual variety) of a hypersurface is one such, as I now describe.

6.4.1. Gauss maps. A classical construction for the geometry of surfaces in 3-space, is the *Gauss map* that maps a point of the surface to its unit normal vector on the unit sphere as in Figure 3.

This Gauss image can be defined for a surface in $\mathbb{P}^3$ without the use of a distance function if one instead takes the union of all *conormal lines* (see §6.2.3) in $\mathbb{P}^{3*}$. Let $S^\vee \subset \mathbb{P}^{3*}$ denote this Gauss image. One loses qualitiative information in this setting, however one still has the information of the *dimension* of $S^\vee$.

This dimension will drop if through all points of the surface there is a curve along which the tangent plane is constant. For example, if M is a cylinder, i.e., the union of lines in three space perpendicular to a plane curve, the Gauss image is a curve:

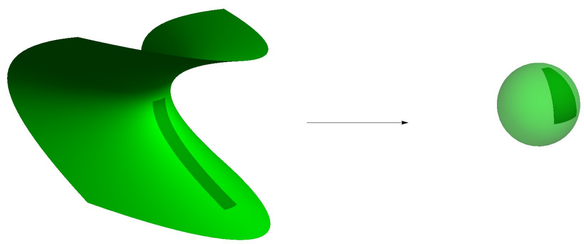


Figure 6.4.1. The shaded area of the surface maps to the shaded area of the sphere.

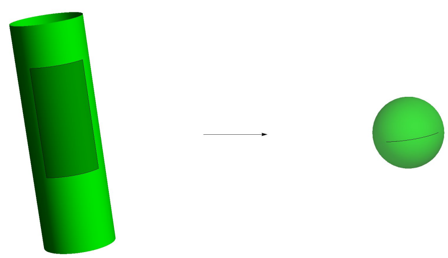
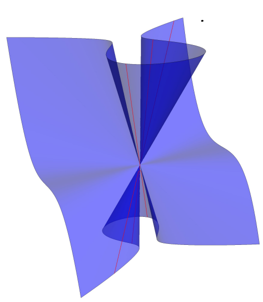


Figure 6.4.2. Lines on the cylinder are collapsed to a point.

The extreme case is when the surface is a plane, then its Gauss image is just a point.

6.4.2. What do surfaces with degenerate Gauss maps "look like"? Here is a generalization of the cylinder above: Consider a curve $C \subset \mathbb{P}^3$, and a point $p \in \mathbb{P}^3$. Define *the cone over C with vertex p*,

$$J(C, p) := \{[x] \in \mathbb{P}^3 \mid x = y + \overline{p} \text{ for some } y \in C,\ \overline{p} \in \hat{p}\}.$$



Exercise 6.4.2.1: (1) Show that if $p \neq y$, $\hat{T}_x J(C,p) = \mathrm{span}\{\hat{T}_y C, \hat{p}\}$.

Thus the tangent space to the cone is constant along the rulings, and the surface only has a curves worth of tangent (hyper)-planes, so its dual variety is degenerate.

Exercise 6.4.2.2: (2) More generally, let $X \subset \mathbb{P}V$ be an irreducible variety and let $L \subset \mathbb{P}V$ be a linear space. Define $J(X,L)$, the *cone* over X with vertex L analogously. Show that given $x \in X_{smooth}$, with $x \notin L$, the tangent space to $J(X,L)^\vee$ at $\overline{x} + \overline{\ell}$ is constant for all $\ell \in L$.

Here is another type of surface with a degenerate Gauss map: Consider again a curve $C \subset \mathbb{P}^3$, and this time let $\tau(C) \subset \mathbb{P}^3$ denote the Zariski closure of the union of all points on $\mathbb{P}\hat{T}_x C$ as x ranges over the smooth points of C. The variety $\tau(C)$ is called the *tangential variety* to the curve C.

Exercise 6.4.2.3: (2) Show that if $y_1, y_2 \in \tau(C)$ are both on a tangent line to $x \in C$, then $\hat{T}_{y_1}\tau(C) = \hat{T}_{y_2}\tau(C)$, and thus $\tau(C)^\vee$ is degenerate. ⊚

In 1910 C. Segre proved that the above two examples are the only surfaces with degenerate dual varieties:

Theorem 6.4.2.4. [**Seg10**, p. 105] *Let $S^2 \subset \mathbb{P}^3$ be a surface with degenerate Gauss image . Then S is one of the following:*

(1) *A linearly embedded $\mathbb{P}^2$,*

(2) *A cone over a curve C,*

(3) *A tangential variety to a curve C.*

(1) is a special case of both (2) and (3) and is the only intersection of the two.

The proof is differential-geometric, see [**?**, §3.4] **check sect**.

6.4.3. Dual varieties. If $X \subset \mathbb{P}V$ is an irreducible hypersurface, the Zariski closure of its Gauss image will be a projective subvariety of $\mathbb{P}V^*$. Gauss images of hypersurfaces are special cases of *dual varieties*. For an irreducible variety $X \subset \mathbb{P}V$, define $X^\vee \subset \mathbb{P}V^*$, the *dual variety of* X, by

$$X^\vee := \overline{\{H \in \mathbb{P}V^* \mid \exists x \in X_{smooth},\ \hat{T}_x X \subseteq \hat{H}^\perp\}}$$

$$\overline{\{H \in \mathbb{P}V^* \mid \exists x \in X_{smooth},\ H \in \mathbb{P}N_x^* X\}}$$

Here H refers both to a point in $\mathbb{P}V^*$ and the hyperplane $\mathbb{P}(\hat{H}^\perp) \subset \mathbb{P}V$.

That the dual variety is indeed a variety may be seen by considering the following *incidence correspondence*:

$$\mathcal{I} := \overline{\{(x, H) \in X_{smooth} \times \mathbb{P}V^* \mid \mathbb{P}\hat{T}_x X \subseteq H\}} \subset \mathbb{P}V \times \mathbb{P}V^*$$

and note that its image under the two projections are respectively X and $X^\vee$. When X is smooth, $\mathcal{I} = \mathbb{P}N^*X$, the projectivized conormal bundle. Both projections are surjective regular maps, so in by Theorem 3.1.4.1, $X^\vee$ is an irreducible variety.

Exercise 6.4.3.1: (2) Show

$$\mathcal{I} = \overline{\{(x, H) \in \mathbb{P}V \times (X^\vee)_{smooth} \mid \mathbb{P}\hat{T}_H X^\vee \subseteq x\}} \subset \mathbb{P}V \times \mathbb{P}V^*$$

and thus $(X^\vee)^\vee = X$. (This is called the *reflexivity theorem* and dates back to C. Segre.)

For our purposes, the most important property of dual varieties is that for a smooth hypersurface other than a hyperplane, its dual variety is also a hypersurface. This will be a consequence of the B. Segre dimension formula 6.4.5.1 below. If the dual of $X \subset \mathbb{P}V$ is not a hypersurface, one says that $X^\vee$ is *degenerate*. It is a classical problem to study the varieties with degenerate dual varieties.

Exercise 6.4.2.2 shows that higher dimensional cones have degenerate dual varieties. Griffiths and Harris [**GH79**] vaguely conjectured a higher dimensional generalization of C. Segre's theorem, namely that a variety with a degenerate dual is "built out of " cones and tangent developables. For example, $Z(\det_n)$ may be thought of as the union of tangent lines to tangent lines to ... to the Segre variety $Seg(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1})$, and we will see that it indeed has a degenerate dual variety.

Segre's theorem indicates that if we take the Zariski closure in $\mathbb{P}S^dV^*$ of the set of irreducible hypersurfaces of degree d with degenerate dual varieties, we will obtain a reducible variety. This will complicate the use of dual varieties for Valiant's conjecture.

For more on dual varieties see [**Lan12**, §8.2].

6.4.4. $Z(\det_n)_{sing}$**.** As far as singularities are concerned, the determinant is quite pathological: Thanks to $G_{\det_n}$, the determination of $Z(\det_n)_{sing}$ is easy to describe. Any point of $Z(\det_n)$ is in the $G_{\det_n}$-orbit of some

$$p_r := \begin{pmatrix} \mathrm{Id}_r & 0 \\ 0 & 0 \end{pmatrix} \tag{6.4.1}$$

where $1 \leq r \leq n-1$ and the blocking is $(r, n-r) \times (r, n-r)$. The nature of the singularity of $x \in Z(\det_n)$ is the same as that of the corresponding p_r.

Recall that $\sigma_r = \sigma_r(Seg(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1})) \subset \mathbb{P}(\mathbb{C}^n \otimes \mathbb{C}^n)$ is the set of matrices (up to scale) of rank at most r.

The smooth points of $Z(\det_n) = \sigma_{n-1}$ are those in the $G_{\det_n}$-orbit of p_{n-1}, as shown by the following exercises:

Exercise 6.4.4.1: (1) Show that $d(\det_n)_{p_{n-1}} = dx^n_n$.

Exercise 6.4.4.2: (1) Show that $Z(\det_n)_{sing} = \sigma_{n-2}$.

Exercise 6.4.4.3: (1) Show that $\sigma_r = Z(\det_n)_{Jac,n-r}$.

Exercise 6.2.3.1 implies $\dim \sigma_r(Seg(\mathbb{P}^{u-1} \times \mathbb{P}^{v-1})) = r(u+v-r) - 1$.

6.4.5. What does this have to do with complexity theory? Having a degenerate dual variety is a pathology, and our dimension calculation below will show that if $Q \in S^m\mathbb{C}^M$ is an irreducible polynomial such that Q is an affine linear degeneration of an irreducible polynomial P, then $\dim(Z(Q)^\vee) \leq \dim(Z(P))^\vee$.

To determine the dual variety of $Z(\det_n) \subset \mathbb{P}(E \otimes F)$, recall that any smooth point of $Z(\det_n)$ is $G_{\det_n}$-equivalent to

$$p_{n-1} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & 0 \end{pmatrix} \in Z(\det_n).$$

and that

$$N^*_{p_{n-1}} Z(\det_n) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * \end{pmatrix}.$$

Since any smooth point of $Z(\det_n)$ can be moved to p_{n-1} by a change of basis, we conclude that the tangent hyperplanes to $Z(\det_n)$ are parametrized by the rank one matrices $Seg(\mathbb{P}E^* \otimes \mathbb{P}F^*)$, the space of which has dimension $2n-1$ (or $2n-2$ in projective space), because they are obtained by multiplying a column vector by a row vector.

Proposition 6.4.5.1 (B. Segre, see e.g., [**GKZ94**]). *Let $P \in S^dV^*$ be irreducible and let $[x] \in Z(P)$ be a general point. Then*

$$\dim Z(P)^\vee = \mathrm{rank}(P_{d-2,2}(x^{d-2})) - 2.$$

Here $(P_{d-2,2}(x^{d-2})) \in S^2V^*$, and we are computing the rank of this symmetric matrix. In coordinates, $P_{d-2,2}$ may be written as a symmetric matrix whose entries are polynomials of degree $d-2$ in the coordinates of x, and is called the *Hesssian*.

Proof. Let $x \in \hat{Z}(P) \subset V$ be a smooth point, so $P(x) = \overline{P}(x, \dots, x) = 0$ and $dP_x = \overline{P}(x, \dots, x, \cdot) \neq 0$ and take $h = dP_x \in V^*$, so $[h] \in Z(P)^\vee$. Now consider a curve $h_t \subset \hat{Z}(P)^\vee$ with $h_0 = h$. There must be a corresponding (possibly stationary) curve $x_t \in \hat{Z}(P)$ such that $h_t = \overline{P}(x_t, \dots, x_t, \cdot)$ and thus $h_0' = (d-1)\overline{P}(x^{d-2}, x_0', \cdot)$. Thus the dimension of $\hat{T}_h Z(P)^\vee$ is the rank of $P_{d-2,2}(x^{d-2})$ minus one (we subtract one because we are only allowed to feed in vectors x_0' that are tangent to $Z(P)$). Now just recall that $\dim X = \dim \hat{T}_x X - 1$. $\square$

Exercise 6.4.5.2: (1) Show that if $Q \in S^m\mathbb{C}^M$ and there exists $\tilde{A} : \mathbb{C}^M \to \mathbb{C}^N$ such that $Q(y) = P(\tilde{A}(y))$ for all $y \in \mathbb{C}^{M*}$, then $\mathrm{rank} Q_{m-2,2}(y) \leq \mathrm{rank} P_{m-2,m}(\tilde{A}(y))$.

Exercise 6.4.5.3: (1) Show that every $P \in Sub_k(S^dV)$ has $\dim Z(P)^\vee \leq k-2$.

Exercise 6.4.5.4: (2) Show that $\sigma_3(Ch_n(\mathbb{C}^{n^2})) \not\subset \mathcal{D}et_n$.

Exercise 6.4.5.5: (2) Show that $\sigma_{2n+1}(v_n(\mathbb{P}^{n^2-1})) \not\subset \mathcal{D}et_n$.

Exercise 6.4.5.6: (2) Show that $\{x_1 \cdots x_n + y_1 \cdots y_n = 0\} \subset \mathbb{P}^{2n-1}$ is self dual, in the sense that it is isomorphic to its own dual variety.

To show a hypersurface has a nondegenerate dual variety, it suffices to find a point where the Hessian of its defining equation has maximal rank.

6.4.6. Permanent case. Consider the point

$$y_0 = \begin{pmatrix} 1-m & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ & \vdots & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Exercise 6.4.6.1: (1!) Show $\mathrm{perm}(y_0) = 0$. ⊚

Now compute $(\mathrm{perm}_m)_{m-2,m}(y_0)$: First note that

$$\frac{\partial}{\partial y^i_j}\frac{\partial}{\partial y^k_l}\,\mathrm{perm}_m(y) = \begin{cases} 0 & \text{if } i=k \text{ or } j=l \\ \mathrm{perm}_{m-2}(y^{\hat{i}\hat{k}}_{\hat{j}\hat{l}}) & \text{otherwise} \end{cases}$$

where $y^{\hat{i}\hat{k}}_{\hat{j}\hat{l}}$ is the $(m-2)\times(m-2)$ size matrix obtained by removing rows i,k and columns j,l.

Exercise 6.4.6.2: (2) Show that if we order indices $y^1_1,\dots,y^m_1,y^1_2,\dots,y^m_2,\dots,y^m_m$, then the Hessian matrix of the permanent at y_0 takes the form

$$\begin{pmatrix} 0 & Q & Q & \cdots & Q \\ Q & 0 & R & \cdots & R \\ Q & R & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & R \\ Q & R & \cdots & R & 0 \end{pmatrix} \tag{6.4.2}$$

where

$$Q=(m-2)\begin{pmatrix} 0 & 1 & \cdots 1 & \\ 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix},\ R=\begin{pmatrix} 0 & m-2 & m-2 & \cdots & m-2 \\ m-2 & 0 & -2 & \cdots & -2 \\ m-2 & -2 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -2 \\ m-2 & -2 & \cdots & -2 & 0 \end{pmatrix}.$$

Lemma 6.4.6.3. *Let Q,R be invertible $m\times m$ matrices and let M be an $m^2\times m^2$ matrix of the form* (6.4.2). *Then M is invertible.*

Proof. Without loss of generality, we may assume $Q=\mathrm{Id}_m$ (multiply the left and right by the block diagonal matrix whose block diagonals are $Q^{-1},\mathrm{Id}_m,\dots,\mathrm{Id}_m$). Let $v=(v_1,\dots,v_m)^T$, where $v_j\in\mathbb{C}^m$, be a vector in the kernel. Then we have the equations

$$\begin{aligned} &v_2+\cdots+v_m=0,\\ &v_1+Rv_3+\cdots+Rv_m=0,\\ &\vdots\\ &v_1+Rv_2+\cdots+Rv_{m-1}=0. \end{aligned}$$

i.e.,

$$\begin{aligned} &v_2+\cdots+v_m=0,\\ &v_1-Rv_2=0,\\ &\vdots\\ &v_1-Rv_m=0. \end{aligned}$$

Multiply the first line by R to conclude $(m-1)v_1 = 0$ and hence $v_1 = 0$, and the remaining equations imply the other $v_j = 0$. $\square$

Thus the permanent hypersurface $Z(\mathrm{perm}_m) \subset \mathbb{P}^{m^2-1}$. has a non-degenerate Gauss map. When one includes $\mathbb{C}^{m^2} \subset \mathbb{C}^{n^2}$, so the equation $Z(\mathrm{perm}_m)$ becomes an equation in a space of n^2 variables that only uses m^2 of the variables, one gets a cone with vertex $\mathbb{P}^{n^2-m^2-1}$ corresponding to the unused variables, in particular, the Gauss image will have dimension m^2-2.

If one makes an affine linear substitution $X = X(Y)$, by the chain rule, the Gauss map of $\{\det(X(Y)) = 0\}$ will be at least as degenerate as the Gauss map of $\{\det(X) = 0\}$. Using this, one obtains:

Theorem 6.4.6.4 (Mignon-Ressayre [**MR04**]). *If $n(m) < \frac{m^2}{2}$, then there do not exist affine linear functions $x^i_j(y^s_t)$, $1 \le i, j \le n$, $1 \le s, t \le m$ such that*
$\mathrm{perm}_m(Y) = \det_n(X(Y))$. *I.e.,* $\mathrm{dc}(\mathrm{perm}_m) \ge \frac{m^2}{2}$.

Remark 6.4.6.5. We saw a linear lower bound by taking one derivative and a quadratic lower bound by taking two. Unfortunately it does not appear to be possible to improve the Mignon-Ressayre bound by taking three derivatives.

6.5. Extension of the Mignon-Ressayre result to $\overline{\mathrm{dc}}$

To extend the Mignon-Ressayre theorem to $\overline{\mathrm{dc}}$ we will need to find polynomials on $\mathbb{P}S^nV$ that vanish on the hypersurfaces with degenerate dual varieties. This was a classically studied question whose answer was known only in a very few number of small cases. In this section I present an answer to the classical question and its application to Conjecture 1.2.5.2.

6.5.1. First steps towards equations. Let $P \in S^dV^*$ be irreducible. Segre's formula 6.4.5.1 may be restated as: $\dim Z(P)^\vee \le k$ if and only if, for all $w \in V$,

$$P(w) = 0 \quad \Rrightarrow \quad \det{}_{k+3}(P_{d-2,2}(w^{d-2})|_F) = 0 \ \forall F \in G(k+3, V). \tag{6.5.1}$$

Equivalently, for any $F \in G(k+3, V)$, the polynomial P must divide $\det_{k+3}(P_{d-2,2}|_F) \in S^{(k+3)(d-2)}V^*$, where $\det_{k+3}$ is evaluated on the S^2V^* factor in $S^2V^* \otimes S^{d-2}V^*$.

Thus to find polynomials on S^dV^* characterizing hypersurfaces with degenerate duals, we need polynomials that detect if a polynomial P divides a polynomial Q. Now $P \in S^dV^*$ divides $Q \in S^eV^*$ if and only if $Q \in P \cdot S^{e-d}V^*$, i.e.

$$x^{I_1}P \wedge \cdots \wedge x^{I_D}P \wedge Q = 0$$

where x^{I_j}, is a basis of $S^{e-d}V$ (and $D = \binom{\mathbf{v}+e-d-1}{e-d}$). Let $\mathcal{D}_{k,d,N} \subset \mathbb{P}S^d\mathbb{C}^N$ denote the zero set of these equations when $Q = \det_{k+3}(P_{d-2,2}|_F)$ as F ranges over $G(k+3, V)$.

Define $Dual_{k,d,N} \subset \mathbb{P}(S^dV^*)$ as the Zariski closure of the set of irreducible hypersurfaces of degree d in $\mathbb{P}V \simeq \mathbb{P}^{N-1}$, whose dual variety has dimension at most k. Our discussion above implies $Dual_{k,d,N} \subseteq \mathcal{D}_{k,d,N}$.

Note that $[\det_n] \in Dual_{2n-2,n,n^2} \subseteq \mathcal{D}_{2n-2,n,n^2}$.

6.5.2. The lower bound on $\overline{dc}(\mathrm{perm}_m)$. The calculation of §6.4.6 shows that $\mathrm{perm}_{m-2,2}(y_0^{m-2})$ is of maximal rank. Here we don't have perm_m, but rather $\ell^{n-m}\,\mathrm{perm}_m$.

Proposition 6.5.2.1. *Let $U = \mathbb{C}^M$, let $R \in S^mU^*$ be irreducible, let ℓ be a coordinate on $\mathbb{C}$ be nonzero, let $U^* \oplus \mathbb{C} \subset \mathbb{C}^{N*}$ be a linear inclusion.*

If $[R] \in \mathcal{D}_{\kappa,m,M}$ and $[R] \notin \mathcal{D}_{\kappa-1,m,M}$, then $[\ell^{d-m}R] \in \mathcal{D}_{\kappa,d,N}$ and $[\ell^{d-m}R] \notin \mathcal{D}_{\kappa-1,d,N}$.

Proof. Let $u_1, \ldots, u_M, v, w_{M+2}, \ldots, w_N$ be a basis of $\mathbb{C}^N$ adapted to the inclusions $\mathbb{C}^M \subset \mathbb{C}^{M+1} \subset \mathbb{C}^N$, so $(U^*)^\perp = \langle w_{M+2}, \ldots, w_N\rangle$ and $(L^*)^\perp = \langle u_1, \ldots, u_M, w_{M+2}, \ldots, w_N\rangle$. Let $c = (d-m)(d-m-1)$. In these coordinates, the matrix of $(\ell^{d-m}R)_{d-2,2}$ in $(M, 1, N-M-1) \times (M, 1, N-M-1)$-block form:

$$(\ell^{d-m}R)_{d-2,2} = \begin{pmatrix} \ell^{d-m}R_{m-2,2} & \ell^{d-m-1}R_{m-1,1} & 0 \\ \ell^{d-m-1}R_{m-1,1} & c\ell^{d-m-2}R & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

First note that $\det_{M+1}((\ell^{d-m}R)_{d-2,2}|_F)$ for any $F \in G(M+1, \mathbb{C}^N)$ is either zero or a multiple of $\ell^{d-m}R$. If $\dim Z(R)^\vee = M-2$ (the expected dimension), then for a general $F \in G(M+1, \mathbb{C}^N)$, $\det_M((\ell^{d-m}R)_{d-2,2}|_F)$ will not be a multiple of $(\ell^{d-m}R)_{d-2,2}$, and more generally if $\dim Z(R)^\vee = \kappa$, then for a general $F \in G(\kappa+2, \mathbb{C}^N)$, $\det_{\kappa+2}((\ell^{d-m}R)_{d-2,2}|_F)$ will not be a multiple of $\ell^{d-m}R$ but for any $F \in G(\kappa+3, \mathbb{C}^N)$, $\det_{\kappa+3}((\ell^{d-m}R)_{d-2,2}|_F)$ will be a multiple of $\ell^{d-m}R$. This shows $[R] \notin \mathcal{D}_{\kappa-1,m,M}$, implies $[\ell^{d-m}R] \notin \mathcal{D}_{\kappa-1,d,N}$.

Exercise 6.5.2.2: (1) Show that $[R] \in \mathcal{D}_{\kappa,m,M}$, implies $[\ell^{d-m}R] \in \mathcal{D}_{\kappa,d,N}$. ⊚

□

Proposition 6.5.2.1 implies:

Theorem 6.5.2.3. [**LMR13**] *$\mathcal{P}erm_n^m \not\subset \mathcal{D}_{2n-2,n,n^2}$ when $m < \frac{n^2}{2}$. In particular, $\overline{dc}(\mathrm{perm}_m) \geq \frac{m^2}{2}$.*

On the other hand, by Exercise 6.4.5.3 cones have degenerate duals, so $\ell^{n-m}\operatorname{perm}_m \in \mathcal{D}_{2n-2,n,n^2}$ whenever $m \geq \frac{n^2}{2}$.

The next step from this perspective would be:

Problem 6.5.2.4. Find equations that distinguish cones (e.g. $Z(\ell^{n-m}\operatorname{perm}_m) \subset \mathbb{P}^{n^2-1}$) from tangent developables (e.g., $Z(\det_n) \subset \mathbb{P}^{n^2-1}$). More precisely, find equations that are zero on tangent developables but nonzero on cones.

6.5.3. A better module of equations. The equations above are of enormous degree. I now derive equations of much lower degree. Since $P \in S^d\mathbb{C}^N$ divides $Q \in S^e\mathbb{C}^N$ if and only if for each $L \in G(2,\mathbb{C}^N)$, $P|_L$ divides $Q|_L$, it will be sufficient to solve this problem for polynomials on $\mathbb{C}^2$. This will have the advantage of producing polynomials of much lower degree.

Let $V = \mathbb{C}^2$, let $d \leq e$, let $P \in S^dV$ and $Q \in S^eV$. If P divides Q then $S^{e-d}V \cdot P$ will contain Q. That is, the vectors $x^{e-d}P, x^{e-d-1}yP, \dots, y^{e-d}P, Q$ in S^eV will fail to be linearly independent, i.e.,

$$x^{e-d}P \wedge x^{e-d-1}yP \wedge \cdots \wedge y^{e-d}P \wedge Q = 0.$$

Since $\dim S^eV = e+1$, these potentially give a $\binom{e+1}{e-d+2}$-dimensional vector space of equations, of degree $e-d+1$ in the coefficients of P and linear in the coefficients of Q.

By taking our polynomials to be $P = P|_L$ and $Q = \det_{k+3}(P_{n-2,2}|_F)|_L$ for $F \in G(k+3,V)$ and $L \in G(2,F)$ (or, for those familiar with flag varieties, better to say $(L,F) \in Flag_{2,k+3}(V)$) we now have equations parametrized by the pairs (L,F). Note that $\deg(Q) = e = (k+3)(d-2)$.

Remark 6.5.3.1. More generally, with $V = \mathbb{C}^2$, given $P \in S^dV$, $Q \in S^eV$, one can ask if P,Q have at least r roots in common (counting multiplicity). Then P,Q having r points in common says the spaces $S^{e-r}V \cdot P$ and $S^{d-r}V \cdot Q$ intersect. That is,

$$x^{e-r}P \wedge x^{e-r-1}yP \wedge \cdots \wedge y^{e-r}P \wedge x^{d-r}Q \wedge x^{d-r-1}yQ \wedge \cdots \wedge y^{d-r}Q = 0.$$

In the case $r = 1$, we get a single polynomial, called the *resultant*, which is of central importance. In particular, the proof of Noether normalization from §3.1.4, that the projection of a projective variety $X \subset \mathbb{P}W$ from a point $y \in \mathbb{P}W$ with $y \notin X$, to $\mathbb{P}(W/\hat{y})$ is still a projective variety relies on the resultant to produce equations for the projection.

6.6. Symmetries of the determinant and permanent

The permanent and determinant both have the property that they are *characterized* by their symmetry groups in the sense described in ***. I expect these symmetry groups to play a central role in the study of Valiant's conjecture in future work. For example, the only known exponential separation of the permanent from the determinant in any restricted model (as defined in Chapter 7), is the model of *equivariant determinantal complexity*, which is defined in terms of symmetry groups.

6.6.1. Symmetries of the determinant.

Theorem 6.6.1.1 (Frobenius [**Fro97**]). *Write $\rho : GL_{n^2} \to GL(S^n\mathbb{C}^{n^2})$ for the induced action. Let $\phi \in GL_{n^2}$ be such that $\rho(\phi)(\det_n) = \det_n$. Then, identifying $\mathbb{C}^{n^2}$ with the space of $n \times n$ matrices,*

$$\phi(z) = \begin{cases} gzh, & \text{or} \\ gz^T h \end{cases}$$

for some $g, h \in GL_n$, with $\det_n(g)\det_n(h) = 1$. Here z^T denotes the transpose of z.

I will present the proof from [**Die49**].

Write $\mathbb{C}^{n^2} = E \otimes F = \mathrm{Hom}(E^*, F)$ with $E, F = \mathbb{C}^n$. Let $\mathbb{Z}_n$ denote the cyclic group on n elements and consider the inclusion $\mathbb{Z}_n \times \mathbb{Z}_n \subset GL(E) \times GL(F)$ given by the n-th roots of unity times the identity matrix. Let μ_n denote the kernel of the product map $(\mathbb{Z}_n)^{\times 2} \to \mathbb{Z}_n$.

Corollary 6.6.1.2. $G_{\det_n} = (SL(E) \times SL(F))/\mu_n \rtimes \mathbb{Z}_2$

To prove the Corollary, just note that the $\mathbb{C}^*$ corresponding to $\det(g)$ above and μ_n are the kernel of the map $\mathbb{C}^* \times SL(E) \times SL(F) \to GL(E \otimes F)$.

Exercise 6.6.1.3: (2) Prove the $n = 2$ case of Corollary 6.6.1.2. ⊚

Lemma 6.6.1.4. *Let $U \subset E \otimes F$ be a linear subspace such that $U \subset Z(\det_n)$. Then $\dim U \leq n^2 - n$. The subvariety of the Grassmannian $G(n^2 - n, E \otimes F)$ consisting of maximal linear spaces on $Z(\det_n)$ has two irreducible components, call them Σ_α and Σ_β, where*

$$\Sigma_\alpha = \{X \in G(n^2 - n, E \otimes F) \mid \ker(X) = \hat{L} \text{ for some } L \in \mathbb{P}E^*\}, \textit{and} \tag{6.6.1}$$

$$\Sigma_\beta = \{X \in G(n^2 - n, E \otimes F) \mid \mathrm{Image}(X) = \hat{H} \text{ for some } H \in \mathbb{P}F^*\}. \tag{6.6.2}$$

Here for $f \in X$, $f : E^* \to F$ is considered as a linear map, $\ker(X)$ means the intersections of the kernels of all $f \in X$ and $\mathrm{Image}(X)$ is the span of all the images.

Moreover, for any two distinct $X_j \in \Sigma_\alpha$, $j = 1, 2$, and $Y_j \in \Sigma_\beta$ we have

$$\dim(X_1 \cap X_2) = \dim(Y_1 \cap Y_2) = n^2 - 2n, \textit{and} \tag{6.6.3}$$

$$\dim(X_i \cap Y_j) = n^2 - 2n + 1. \tag{6.6.4}$$

Exercise 6.6.1.5: (2) Prove Lemma 6.6.1.4.

One can say more: each element of Σ_α corresponds to a left ideal and each element of Σ_β corresponds to a right ideal in the space of $n \times n$ matrices.

Proof of theorem 6.6.1.1. Let $\Sigma = \Sigma_\alpha \cup \Sigma_\beta$. Then the automorphism of $G(n^2 - n, E \otimes F)$ induced by ϕ must preserve Σ. By the conditions (6.6.3),(6.6.4) of Lemma 6.6.1.4, in order to preserve dimensions of intersections, either every $U \in \Sigma_\alpha$ must map to a point of Σ_α, in which case every $V \in \Sigma_\beta$ must map to a point of Σ_β, or, every $U \in \Sigma_\alpha$ must map to a point of Σ_β, and every $V \in \Sigma_\beta$ must map to a point of Σ_α. If we are in the second case, replace ϕ by $\phi \circ T$, where $T(z) = z^T$, so we may now assume ϕ preserves both Σ_α and Σ_β.

Now $\Sigma_\alpha \simeq \mathbb{P}E^*$, so ϕ induces an algebraic map $\phi_E : \mathbb{P}E^* \to \mathbb{P}E^*$.

Exercise 6.6.1.6: (2) Show that if $L_1, L_2, L_3 \in \mathbb{P}E$ lie on a $\mathbb{P}^1$, then $\dim(U_{L_1} \cap U_{L_2} \cap U_{L_3}) = n^2 - 2n$.

In order for ϕ to preserve $\dim(U_{L_1} \cap U_{L_2} \cap U_{L_3})$, the images of the L_j under ϕ_E must also lie on a $\mathbb{P}^1$, and thus ϕ_E must take lines to lines (and similarly hyperplanes to hyperplanes). But then, (see, e.g., [**Har95**, §18, p. 229]) $\phi_E \in PGL(E)$, and similarly, $\phi_F \in PGL(F)$, where $\phi_F : \mathbb{P}F^* \to \mathbb{P}F^*$ is the corresponding map. Here $PGL(E)$ denotes $GL(E)/\mathbb{C}^*$, the image of $GL(E)$ in its action on projective space. Write $\hat{\phi}_E \in GL(E)$ for any choice of lift and similarly for F.

Consider the map $\tilde{\phi} \in GL(E \otimes F)$ given by $\tilde{\phi}(X) = \hat{\phi}_E{}^{-1}\phi(X)\hat{\phi}_F{}^{-1}$. The map $\tilde{\phi}$ sends each $U \in \Sigma_\alpha$ to itself as well as each $V \in \Sigma_\beta$, in particular it does the same for all intersections. Hence it preserves $Seg(\mathbb{P}E \times \mathbb{P}F) \subset \mathbb{P}(E \otimes F)$ point-wise, so it is up to scale the identity map because $E \otimes F$ is spanned by points of $\hat{S}eg(\mathbb{P}E \times \mathbb{P}F)$. $\square$

6.6.2. Symmetries of the permanent. Write $\mathbb{C}^{n^2} = E \otimes F$. Let $\Gamma_n^E := T_E^{SL} \rtimes \mathfrak{S}_n$, and similarly for F. Then it is easy to see $(\Gamma_n^E \times \Gamma_n^F) \rtimes \mathbb{Z}_2 \to G_{\mathrm{perm}_n}$, where the nontrivial element of $\mathbb{Z}_2$ acts by sending a matrix to its transpose. We would like to show this map is surjective and determine its kernel. However, it is not when $n = 2$.

Exercise 6.6.2.1: What is G_{perm_2}? ⊚

Theorem 6.6.2.2. [**MM62**] *For $n \geq 3$, $G_{\mathrm{perm}_n} = (\Gamma_n^E \times \Gamma_n^F)/\mu_n \rtimes \mathbb{Z}_2$.*

Proof. I follow [**Ye11**]. Recall the description of $Z(\mathrm{perm}_n)_{Jac,n-2}$ from Lemma 6.3.2.4. Any linear transformation preserving the permanent must send a component of $Z(\mathrm{perm}_n)_{Jac,n-2}$ of type (1) to another of type (1). It must send a component C^j either to some C^k or some C_i. But if $i \neq j$, $C^j \cap C^i = 0$ and for all i, j, $\dim(C^i \cap C_j) = 1$. Since intersections must be mapped to intersections, either all components C^i are sent to components C_k or all are permuted among themselves. By composing with an element of $\mathbb{Z}_2$, we may assume all the C^i's are sent to C^i's and the C_j's are sent to

C_j's. Similarly, by composing with an element of $\mathfrak{S}_n \times \mathfrak{S}_n$ we may assume each C_i and C^j is sent to itself. But then their intersections are sent to themselves. So we have, for all i, j,

$$(x^i_j) \mapsto (\lambda^i_j x^i_j) \tag{6.6.5}$$

for some λ^i_j and there is no summation in the expression. Consider the image of a size 2 submatrix, e.g.,

$$\begin{matrix} x^1_1 & x^1_2 \\ x^2_1 & x^2_2 \end{matrix} \mapsto \begin{matrix} \lambda^1_1 x^1_1 & \lambda^1_2 x^1_2 \\ \lambda^2_1 x^2_1 & \lambda^2_2 x^2_2 \end{matrix}. \tag{6.6.6}$$

In order that the map (6.6.5) be in G_{perm_n}, when $(x^i_j) \in Z(\mathrm{perm}_n)_{Jac,n-2}$, the permanent of the matrix on the right hand side of (6.6.6) must be zero. Using that $x^1_1 x^2_2 + x^1_2 x^2_1 = 0$, the permanent of the right hand side of (6.6.6) is $\lambda^1_1\lambda^2_2 x^1_1 x^2_2 + \lambda^2_1\lambda^1_2 x^1_2 x^2_1 = x^1_1 x^2_2(\lambda^1_1\lambda^2_2 - \lambda^2_1\lambda^1_2)$ which implies $\lambda^1_1\lambda^2_2 - \lambda^1_2\lambda^2_1 = 0$, thus all the 2×2 minors of the matrix (λ^i_j) are zero, so it has rank one and is the product of a column vector and a row vector, but then it is an element of $T_E \times T_F$. $\square$

6.6.3. Do optimal determinantal expressions see symmetry? Recall from Chapter 4 that the symmetries of the matrix multiplication tensor appear in the optimal and conjecturally optimal rank expressions for it. Will the same be true for determinantal expressions of polynomials, in particular of the permanent?

The best known determinantal expression of perm_m is of size $2^m - 1$ and is due to Grenet [**Gre11**]. (Previously Valiant [**Val79a**] had shown there was an expression of size 4^m.) We saw (Corollary 6.3.3.8) that when $m = 3$ this is the best expression. This motivated N. Ressayre and myself to try to understand Grenet's expression. We observed the following *equivariance* property:

Recall $\Gamma^E_m \subset T^{SL}_E \rtimes \mathfrak{S}_m \subset G_{\mathrm{perm}_m}$ from §6.6.2.

Proposition 6.6.3.1. [**LR15**] *Grenet's expressions $\tilde{A}_{Grenet} : Mat_m(\mathbb{C}) \to Mat_n(\mathbb{C})$ such that $\mathrm{perm}_m(Y) = \det_n(\tilde{A}_{Grenet}(Y))$ are Γ^E_m-equivariant. Namely, given $g \in \Gamma^E_m$, there exist $n \times n$ matrices B, C such that $\tilde{A}_{Grenet,m}(g \cdot Y) = B\tilde{A}_{Grenet,m}(Y)C$. , i.e, there exists an injective group homomorphism $\psi : \Gamma^E_m \to G_{\det_n}$ such that $\tilde{A}_{Grenet,m}(Y) = \psi(g)(\tilde{A}_{Grenet,m}(gY))$.*

For example, let

$$g(t) = \begin{pmatrix} t_1 & & \\ & t_2 & \\ & & t_3 \end{pmatrix}.$$

Then $A_{Grenet,3}(g(t)Y) = B(t)A_{Grenet,3}(Y)C(t)$, where

$$B(t) = \begin{pmatrix} t_3 & & & & & & \\ & t_1t_3 & & & & & \\ & & t_1t_3 & & & & \\ & & & t_1t_3 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{pmatrix} \quad \text{and} \quad C(t) = B(t)^{-1}.$$

Exercise 6.6.3.2: (2) Determine $B(g)$ and $C(g)$ when $g \in \Gamma_3^E$ is the permutation $(1,2)$.

In fact, via this equivariance, one can give an invariant description of Grenet's expressions:

Let let $k \in [m]$. The space S^kE is an irreducible $GL(E)$-module but it is is not irreducible as a Γ_m^E-module. For example, let $e_1, \dots, e_m$ be a basis of E, and let $(S^kE)_{reg}$ denote the span of $\prod_{i\in I} e_i$, for $I \subset [m]$ of cardinality k (the space spanned by the square-free monomials, also known as the space of *regular* weights): $(S^kE)_{reg}$ is an irreducible Γ_m^E-submodule of S^kE. Moreover, there exists a unique Γ_m^E-equivariant projection π_k from S^kE to $(S^kE)_{reg}$.

For $v \in E$, define $s_k(v) : (S^kE)_{reg} \to (S^{k+1}E)_{reg}$ to be multiplication by v followed by π_{k+1}. Alternatively, $(S^{k+1}E)_{reg}$ is a Γ_m^E-submodule of $E \otimes (S^kE)_{reg}$, and $s_k : E \to (S^kE)^*_{reg} \otimes (S^{k+1}E)_{reg}$ is the unique Γ_m^E-equivariant inclusion. Let $\mathrm{Id}_W : W \to W$ denote the identity map on the vector space W. Fix a basis $f_1, \dots, f_m$ of F^*.

Proposition 6.6.3.3. [**LR15**] *The following is Grenet's determinantal representation of* perm_m. *Let* $\mathbb{C}^n = \bigoplus_{k=0}^{m-1} (S^kE)_{reg}$, *so* $n = 2^m - 1$, *and identify* $S^0E \simeq (S^mE)_{reg}$ *(both are trivial* Γ_m^E*-modules). Set*

$$\Lambda_0 = \sum_{k=1}^{m-1} \mathrm{Id}_{(S^kE)_{reg}}$$

and define

$$\tilde{A} = \Lambda_0 + \sum_{k=0}^{m-1} s_k \otimes f_{k+1}. \tag{6.6.7}$$

Then $(-1)^{m+1}\,\mathrm{perm}_m = \det_n \circ \tilde{A}$. *To obtain the permanent exactly, replace* $\mathrm{Id}_{(S^1E)_{reg}}$ *by* $(-1)^{m+1}\,\mathrm{Id}_{(S^1E)_{reg}}$ *in the formula for* Λ_0.

In bases respecting the block decomposition induced from the direct sum, the linear part, other than the last term which lies in the upper right block, *lies just below the diagonal blocks, and all blocks other than the upper right block and the diagonal and sub-diagonal blocks, are zero.*

Moreover the map $\tilde{A}$ *is* Γ_m^E*-equivariant.*

I prove Proposition 6.6.3.3 in §8.12.4.

6.7. dc v. $\overline{\mathrm{dc}}$

Is conjecture 6.1.6.2 really stronger than Valiant's conjecture 6.1.6.1? That is, do there exist sequences (P_m) of polynomials with $\overline{\mathrm{dc}}(P_m)$ bounded by a polynomial in m but $\mathrm{dc}(P_m)$ growing super-polynomially?

K. Mulmuley [**Mul**] conjectures that this is indeed the case, and the existence of such sequences "explains" why Valiant's conjecture is so difficult.

Before addressing this conjecture, one should at least find a sequence P_m with $\mathrm{dc}(P_m) > \overline{\mathrm{dc}}(P_m)$. At least such a sequence is known as I now describe.

Warning: this section is in very rough form

begin with general discussion of finding components on the boundary*

6.7.1. On the boundary of the orbit of the determinant. **get rid of module structure, postphone

Recall that the transposition $\tau \in G_{\det_n}$ allows us to write $\mathbb{C}^{n^2} = E \otimes E = S^2E \oplus \Lambda^2 E$, where the decomposition is into the ± 1 eigenspaces for τ. For $M \in E \otimes E$, write $M = M_S + M_\Lambda$ reflecting this decomposition.

Define a polynomial $P_\Lambda \in S^n(\mathbb{C}^{n^2})^*$ by

$$P_\Lambda(M) = \overline{\det}_n(M_\Lambda, \dots, M_\Lambda, M_S).$$

Let $Pf_i(M_\Lambda)$ denote the Pfaffian of the skew-symmetric matrix, obtained from M_Λ by suppressing its i-th row and column. Write $M_S = (s_{ij})$.

Exercise 6.7.1.1: Show that

$$P_\Lambda(M) = \sum_{i,j} s_{ij} Pf_i(M_\Lambda) Pf_j(M_\Lambda).$$

In particular, $P_\Lambda = 0$ if n is even but is not identically zero when n is odd.

Proposition 6.7.1.2. [**LMR13**] $P_{\Lambda,n} \in \mathcal{D}et_n$. *Moreover,* $\overline{GL(W) \cdot P_\Lambda}$ *is an irreducible codimension one component of the boundary of* $\mathcal{D}et_n$*, not contained in* $\mathrm{End}(W) \cdot [\det_n]$*. In particular* $\overline{dc}(P_{\Lambda,m}) = m < dc(P_{\Lambda,m})$.

The proof of Proposition 6.7.1.2 is given in §8.5.1

The hypersurface defined by P_Λ has interesting properties.

Proposition 6.7.1.3. [**LMR13**]

$$Z(P_\Lambda)^\vee = \overline{\mathbb{P}\{v^2 \oplus v \wedge w \in S^2\mathbb{C}^n \oplus \Lambda^2\mathbb{C}^n,\ v, w \in \mathbb{C}^n\}} \subset \mathbb{P}^{n^2-1}.$$

Proof. Note that

$$P_\Lambda(M) = \lim_{t\to 0} \frac{1}{t} \det(M_\Lambda + tM_S).$$

$\square$

As expected, $Z(P_\Lambda)^\vee$ resembles $Seg(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1})$.

Remark 6.7.1.4. For those familiar with the notation, $Z(P_\Lambda)$ can be defined as the image of the projective bundle $\pi : \mathbb{P}(E) \to \mathbb{P}^{n-1}$, where $E = \mathcal{O}(-1) \oplus Q$ is the sum of the tautological and quotient bundles on $\mathbb{P}^{n-1}$, by a sub-linear system of $\mathcal{O}_E(1) \otimes \pi^*\mathcal{O}(1)$. This sub-linear system contracts the divisor $\mathbb{P}(Q) \subset \mathbb{P}(E)$ to the Grassmannian $G(2, n) \subset \mathbb{P}\Lambda^2\mathbb{C}^n$.

Remark 6.7.1.5. A second way to realize the polynomial $P = ***$ from Example **??** is via P_Λ: take

$$M_\Lambda = \begin{pmatrix} 0 & x_3 & x_2 \\ -x_3 & 0 & x_1 \\ -x_2 & -x_1 & 0 \end{pmatrix}, \quad M_S = \begin{pmatrix} x_1 & 0 & 0 \\ 0 & x_4 & 0 \\ 0 & 0 & x_2 \end{pmatrix}.$$

6.7.2. Mulmuley's conjectures on the wildness of the boundary.

Give conj, discuss why other models may be too weak, what is known about gap between rank and border rank, include ABV example here with $\overline{\mathrm{dc}} > \mathrm{dc}$.

Theorem 6.7.2.1 (Alper-Bogart-Velasco [**ABV15**])**.** $\mathrm{dc}(x_1^3 + x_2^2 x_3 + x_2 x_4^2) \geq 6$.

Remark 6.7.2.2. Note that in contrast $\hat{\mathcal{D}}et_3 \supset \mathbb{P}S^3\mathbb{C}^4$ **put in proof??***. In particular a smooth cubic in four variables has determinantal complexity three. Since $x_1^3 + x_2^2 x_3 + x_2 x_4^2$ can degenerate to a polynomial with determinantal complexity three (in fact it is the unique cubic in four variables with determinantal complexity greater than three), we see the failure of semi-continuity ... say more....

6.7.3. Hüttenhain's $\det_3$ theorem. **to be written**

6.8. Determinantal hypersurfaces

Classically, there was interest in determining which smooth hypersurfaces of degree d were expressible as a $d \times d$ determinant. The result in the first nontrivial case shows how daunting GCT might be.

Theorem 6.8.0.1 (Letao Zhang and Zhiyuan Li)**.** *The variety* $\mathbb{P}\{P \in S^4\mathbb{C}^4 \mid [P] \in \mathcal{D}et_4\} \subset \mathbb{P}S^4\mathbb{C}^4$ *is a hypersurface of degree* $640, 224$.

The rest of this subsection uses more advanced language from algebraic geometry and can be safely skipped.

The following "folklore" theorem was made explicit in [**Bea00**, Cor. 1.12]:

Theorem 6.8.0.2. *Let* $U = \mathbb{C}^{n+1}$*, let* $P \in S^dU$*, and let* $Z = Z(P) \subset \mathbb{CP}^n$ *be the corresponding hypersurface of degree* d*. Assume* Z *is smooth and choose any inclusion* $U \subset \mathbb{C}^{d^2}$.

If $P \in \mathrm{End}(\mathbb{C}^{d^2}) \cdot [\det_d]$*, we may form a map between vector bundles* $M : \mathcal{O}_{\mathbb{P}^n}(-1)^d \to \mathcal{O}_{\mathbb{P}^n}^d$ *whose cokernel is a line bundle* $L \to Z$ *with the properties:*

i) $H^i(Z, L(j)) = 0$ *for* $1 \leq i \leq n-2$ *and all* $j \in \mathbb{Z}$

ii) $H^0(X, L(-1)) = H^{n-1}(X, L(j)) = 0$

Conversely, if there exists $L \to Z$ *satisfying properties i) and ii), then* Z *is determinantal via a map* M *as above whose cokernel is* L.

If we are concerned with the hypersurface being in $\mathcal{D}et_n$, the first case where this is not automatic is for quartic surfaces, where it is a codimension one condition:

Proposition 6.8.0.3. [**Bea00**, Cor. 6.6] *A smooth quartic surface is determinantal if and only if it contains a nonhyperelliptic curve of genus* 3 *embedded in* $\mathbb{P}^3$ *by a linear system of degree* 6.

Proof of 6.8.0.1. From Proposition 6.8.0.3, the hypersurface is the locus of quartic surfaces containing a (Brill-Noether general) genus 3 curve C of degree six. This translates into the existence of a lattice polarization

$$\begin{array}{ccc} & h & C \\ h & 4 & 6 \\ C & 6 & 4 \end{array}$$

of discriminant $-(4^2 - 6^2) = 20$. By the Torelli theorems, the $K3$ surfaces with such a lattice polarization have codimension one in the moduli space of quartic $K3$ surfaces.

Let $D_{3,6}$ denote the locus of quartic surfaces containing a genus 3 curve C of degree six in $\mathbb{P}^{34} = \mathbb{P}(S^4\mathbb{C}^4)$. It corresponds to the Noether-Lefschetz divisor NL_{20} in the moduli space of the degree four $K3$ surfaces. Here NL_d denotes the Noether-Lefschetz divisor, parameterizing the degree 4 $K3$ surfaces whose Picard lattice has a rank 2 sub-lattice containing h with discriminant $-d$. (h is the polarization of the degree four $K3$ surface, $h^2 = 4$.)

The Noether-Lefschetz number n_{20}, which is defined by the intersection number of NL_{20} and a line in the moduli space of degree four $K3$ surfaces, equals the degree of $D_{3,6}$ in $\mathbb{P}^{34} = \mathbb{P}(S^4\mathbb{C}^4)$.

The key fact is that n_d can be computed via the modularity of the generating series for any integer d. More precisely, the generating series $F(q) := \sum_d n_d q^{d/8}$ is a modular form of level 8, and can be expressed by a polynomial of $A(q) = \sum_n q^{n^2/8}$ and $B(q) = \sum_n (-1)^n q^{n^2/8}$.

The explicit expression of $F(q)$ is in [**MP**, Thm 2]. As an application, the Noether-Lefschetz number n_{20} is the coefficient of the term $q^{20/8} = q^{5/2}$, which is $640,224$. $\square$

Chapter 7

Valiant's conjecture II: Restricted models and other approaches

This chapter continues the discussion of Valiant's conjecture and its variants. So far we have approached Valiant's conjecture and its variants by trying to improve benchmarks such as proving lower bounds for $\mathrm{dc}(\mathrm{perm}_m)$. Another approach to these conjectures to to prove them under supplementary hypotheses, which are called *restricted models* in the computer science literature. In the case of the restricted models of shallow circuits introduced in §7.1, there is a path to proving the full conjecture by proving lower complexity bounds that are stronger than super-polynomial, as explained in §7.1. I begin the section with a detour for readers not familiar with big numbers as different levels of super-polynomial growth need to be compared both for statements and proofs, and a discussion of the geometry of one of the simplest class of shallow circuits, the $\Sigma\Lambda\Sigma$-circuits whose complexity essentially measures symmetric tensor rank. In §7.2 I explain the geometry associated to the depth 3,4, and 5 circuits that arise in [**GKKS13a**], as interesting lower bounds have been proven in those models. I return to them in §7.5, proving the lower bounds of [**GKKS13a**] for the permanent and determinant in those models and analyze the method of proof, *shifted partial derivatives* is in detail. There are several complexity measures that are equivalent to determinantal complexity, such as *algebraic branching programs* and *iterated matrix multiplication complexity* These are discussed in §7.3. Several additional restricted models are presented in §7.4: Shpilka's [**Shp02**] *depth-2* *symmetric arithmetic circuits*, Aravind and Joegelkar's *rank k determinantal expressions* of [**AJ15**], a beautiful result of Glynn [**Gly13**] on a certain class of expressions for the permanent, Nisan's non-commutative circuits [**Nis91**], and the equivariant determinantal complexity of [**LR15**].

As pointed out by Shpilka and Yehudayoff in [**SY09**], restricted circuits of polynomial size only compute polynomials with "simple" structure. Thus to understand them one needs to determine the precise meaning of "simple" for a give restricted class, and then find an "explicit" polynomial without such structure. One could rephrase this geometrically as restricted circuits of a fixed size define an algebraic variety in $S^n\mathbb{C}^N$ that is the closure of the set of polynomials computable with a restricted circuit of that size. The goal becomes to find an equation of that variety and an explicit polynomial not satisfying that equation.

7.1. Shallow Circuits

The *depth* of a circuit $\mathcal{C}$ is the length of (i.e., the number of edges in) the longest path in $\mathcal{C}$ from an input to its output. If a circuit has small depth, it is called a *shallow circuit*, and the polynomial it computes can be computed quickly in parallel. When one studies circuits of bounded depth, one must allow gates to have an arbitrary number of edges coming in to them ("unbounded fanin"). For such circuits, multiplication by constants is considered "free ".

There are *depth reduction theorems* described in §7.1.3 that enable one substitute the problem of e.g., showing that there does not exist a small circuit computing the permanent to the problem of showing that there does not exists a "slightly less small" shallow circuit computing the permanent. These classes of shallow circuits have algebraic varieties associated to them: the *depth three* or $\Sigma\Pi\Sigma$ circuits, which consist of depth three formulas where the first layer of gates consist of additions, the second of multiplications, and the last gate is an addition gate, the $\Sigma\Lambda\Sigma\Lambda\Sigma$ circuits, which are depth five circuits where the first layer of gates are additions, the second layer consists of "powering gates", where a powering gate takes f to f^δ for some natural number δ, the third layer addition gates, the fourth layer again powering gates, and the fifth layer is an addition gate, and the the depth four $\Sigma\Pi\Sigma\Pi$ circuits which are similarly defined. I describe the associated varieties to these classes of circuits in §7.2.1, §7.2.2, and §**??**. A $\Sigma\Lambda^\alpha\Sigma\Lambda^\beta\Sigma$ means the powers are respectively β and α, and other superscripts are to be similarly interpreted.

One can restrict one's class of circuits further by requiring that they are *homogeneous* in the sense that each gate computes a homogeneous polynomial. It turns out that for $\Sigma\Lambda\Sigma\Lambda\Sigma$ circuits, this is not restrictive for the questions of interest, but for $\Sigma\Pi\Sigma$ circuits, there is a tremendous loss of computing power described in §**??**. As described in §**??**, this loss of computing power can be overcome by something we are already familiar with: computing padded polynomials.

7.1.1. Detour for those not familiar with big numbers. When dealing with shallow circuits, we will have to distinguish between different rates of super-polynomial growth, both in statements and proofs of theorems. This detour is for those readers not used to comparing large numbers.

$$n! \gtrapprox \sqrt{2\pi n}(\frac{n}{e})^n \tag{7.1.1}$$

$$\ln(n!) = n\ln(n) - O(\ln(n)) \tag{7.1.2}$$

$$\binom{2n}{n} \gtrapprox \frac{4^n}{\sqrt{\pi n}} \tag{7.1.3}$$

$$\ln\binom{\alpha n}{\beta n} = \alpha H_e(\frac{\beta}{\alpha})n - O(\ln n) \tag{7.1.4}$$

$$\binom{\alpha n}{\beta n} \sim [\frac{\alpha^\alpha}{\beta^\beta(\alpha-\beta)^{\alpha-\beta}}]^n \tag{7.1.5}$$

where $H_e(x) := -x\ln x - (1-x)\ln(1-x)$ is the *Shannon entropy*. All these identities follow from (7.1.1), which follows from *Stirling's formula*, which gives an approximation for the Gamma function, e.g., for $x > 0$,

$$\Gamma(x) = \sqrt{2\pi}x^{x-\frac{1}{2}}e^{-x}e^{\frac{\theta(x)}{12x}}$$

where $0 < \theta(x) < 1$. Stirling's formula may be proved via complex analysis (estimating a contour integral), see, e.g. [**Ahl78**, §5.2.5].

Exercise 7.1.1.1: (1) Show $a^{\log(b)} = b^{\log(a)}$.

Exercise 7.1.1.2: (1) Consider the following sequences of n:

$$\log_2(n),\ n,\ 100n,\ n^2,\ n^3,\ n^{\log_2(n)},\ 2^{[\log_2(n)]^2},\ n^{\sqrt{\log_2(n)}},\ 2^n,\ \binom{2n}{n},\ n!,\ n^n.$$

In each case, determine for which n, the sequence surpasses the number of atoms in the known universe. (It is estimated that there are between 10^{78} and 10^{82} atoms in the known universe.)

Exercise 7.1.1.3: (1) Compare the sizes of $s^{\sqrt{d}}$ and $2^{\sqrt{d\log ds}}$.

Exercise 7.1.1.4: (1) Compare the sizes of $\binom{n^2+\frac{n}{2}-1}{\frac{n}{2}}$ and $\binom{n}{\frac{n}{2}}^2$.

7.1.2. $\sigma_r(v_d(\mathbb{P}V)$ and $\Sigma\Lambda\Sigma$ circuits. Recall the definition of $\mathbf{R}_S$ from §6.2.2. One of the simplest class of shallow circuits are the $\Sigma\Lambda\Sigma$ circuits mentioned in §6.2, where a polynomial $P \in S^nV$ admits a size $O(r)$ $\Sigma\Lambda\Sigma$

circuit, i.e., $P = \ell_1^n + \cdots + \ell_{O(r)}^n$ for some $\ell_j \in V$ by definition means $[P] \in \sigma^0_{O(r)}(v_n(\mathbb{P}V))$, where the superscript denotes the Zariski open subset of $\sigma_{O(r)}(v_n(\mathbb{P}V))$ consisting of points on an honest secant $\mathbb{P}^{r-1}$.

In this subsection I describe upper and lower bounds for $\mathbf{R}_S$ and $\underline{\mathbf{R}}_S$ for several basic polynomials. First for a monomial, there is Fischer's formula [**Fis94**]:

$$x_1 \cdots x_n = \frac{1}{2^{n-1} n!} \sum_{\epsilon \in \{-1,1\}^{n-1}} (x_1 + \epsilon_1 x_2 + \cdots + \epsilon_{n-1} x_n)^n \epsilon_1 \cdots \epsilon_{n-1} \tag{7.1.6}$$

Remark 7.1.2.1. In §10.6, I show that (7.1.6) is optimal, i.e., that $R_S(x_1 \cdots x_n) = 2^{n-1}$, which was first shown in [**RS11**].

Exercise 7.1.2.2: (1) Verify (7.1.6).

Note that $x_1 \cdots x_n = e_{n,n}$. Here is a generalization of Fischer's formula for odd degree due to H. Lee [**Lee16**] (which is also optimal, see §**??**). First, when when $n = 2k+1$ is odd, rewrite Fischer's formula as:

$$x_1 x_2 \cdots x_n = \frac{1}{2^{n-1} n!} \sum_{I \subset [n], |I| \le k} (-1)^{|I|} (\delta(I,1)x_1 + \delta(I,2)x_2 + \cdots + \delta(I,n)x_n)^n. \tag{7.1.7}$$

For an integer set I and an integer i, define

$$\delta(I,i) = \begin{cases} -1 & i \in I \\ 1 & i \notin I \end{cases}.$$

Theorem 7.1.2.3. [**Lee16**] *Let $d = 2k+1$ and let $N \ge d$. Then*

$$e_{d,N} = \frac{1}{2^{d-1} d!} \sum_{I \subset [N], |I| \le k} (-1)^{|I|} \binom{N-k-|I|-1}{k-|I|} (\delta(I,1)x_1 + \delta(I,2)x_2 + \cdots + \delta(I,N)x_N)^d.$$

In particular, for d odd, $\mathbf{R}_S(e_{d,N}) \le \sum_{i=0}^{\lfloor \frac{d}{2} \rfloor} \binom{N}{i}$.

Proof. We work by downwards induction, the case $d = N$ is Fischer's formula. Let $d < N$ and let $F_{d,N}$ denote the right hand side of the expression.

Observe that $F_{d,d} = e_{d,d}$ and $F_{d,N-1} = F_{d,N}(x_1, \ldots, x_{N-1}, 0)$ up to a constant. In particular $F_{d,d} = F_{d,N}(x_1, \ldots, x_d, 0, \ldots, 0)$. The analogous statement holds setting any subset of the variables to zero. This implies that $F_{d,N}$ is an expression that has all the square-free monomials in $e_{d,N}$ appearing in it. Moreover, there are no other monomials appearing in $F_{d,N}$ as otherwise there would be a monomial involving fewer than d variables that would appear in some specialization to some $e_{d,d}$. Checking the constant is correct, we conclude. $\square$

Using the flattening (see §6.2), $(\det_n)_{\lceil\frac{n}{2}\rceil,\lfloor\frac{n}{2}\rfloor} : S^{\lceil\frac{n}{2}\rceil}W \to S^{\lfloor\frac{n}{2}\rfloor}W$ and writing $W = E\otimes F = \mathbb{C}^n\otimes\mathbb{C}^n$, the image of $(\det_n)_{\lceil\frac{n}{2}\rceil,\lfloor\frac{n}{2}\rfloor}$ is $\Lambda^{\lfloor\frac{n}{2}\rfloor}E\otimes\Lambda^{\lfloor\frac{n}{2}\rfloor}F$, the minors of size $\lfloor\frac{n}{2}\rfloor$. For the permanent one similarly gets sub-permanents. Thus,

$$\underline{\mathbf{R}}_S(\det_n) \geq \binom{n}{\lfloor\frac{n}{2}\rfloor}^2, \quad \underline{\mathbf{R}}_S(\mathrm{perm}_n) \geq \binom{n}{\lfloor\frac{n}{2}\rfloor}^2. \tag{7.1.8}$$

Exercise 7.1.2.4: (1) Find a lower bound for $\underline{\mathbf{R}}_S(x_1\cdots x_n)$.

The following proposition bounds $\Sigma\Lambda\Sigma$ complexity by $\overline{\mathrm{dc}}$:

Proposition 7.1.2.5. *$\ell^{n-m}\sigma_r(v_m(\mathbb{P}V)) \subset \mathcal{D}et_n$ when $n > rm$.*

Exercise 7.1.2.6: (2) Prove Proposition 7.1.2.5.⊚

7.1.3. Depth reduction theorems. A major result in the study of shallow circuits was [**VSBR83**], where it was shown that if a polynomial of degree d can be computed by a circuit of size s, then it can be computed by a circuit of depth $O(\log d \log s)$ and size polynomial in s.

Here are the relevant results relevant for our discussion. They combine results of [**Bre74, GKKS13b, ?, Koi, AV08**]:

Theorem 7.1.3.1. *Let $N = N(n)$ be a polynomial and let $P_n \in S^n\mathbb{C}^N$ be a sequence of polynomials that can be computed by a circuit of size $s = s(n)$.*

Then:

1. *P is computable by a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit of size $2^{O(\sqrt{n\log(ns)\log(N)})}$.*
2. *P is computable by a $\Sigma\Pi\Sigma$ circuit of size roughly $s^{\sqrt{n}}$, more precisely of size $2^{O(\sqrt{n\log(N)\log(ns)})}$.*
3. *P is computable, by a homogeneous $\Sigma\Lambda\Sigma\Lambda\Sigma$ circuit of size roughly $s^{\sqrt{n}}$, more precisely of size $2^{O(\sqrt{n\log(ns)\log(N)})}$, and both powering gates of size of roughly $\sqrt{n}$.*

Here are ideas towards the proof: In [**GKKS13b**] they prove upper bounds for the size of an inhomogeneous depth three circuit computing a polynomial, in terms of the size of an arbitrary circuit computing the polynomial. They first apply the work of [**Koi, AV08**], which allows one to reduce an arbitrary circuit of size s computing a polynomial of degree d in n variables to a formula of size $2^{O(\log s \log d)}$ and depth d.

The next step is via the iterated matrix multiplication polynomial. In §7.3 we will see that formula size is at least as large as iterated matrix multiplication complexity. Say we can compute $f \in S^m\mathbb{C}^M$ via m matrix multiplications of $n\times n$ matrices with linear entries. Group the entries into groups of $\lceil\frac{m}{a}\rceil$ for some a. To simplify the discussion, assume $\frac{m}{a}$ is an integer, otherwise adjust accordingly. Write

$$X_1 \cdots X_m = (X_1 \cdots X_{\frac{m}{a}})(X_{\frac{m}{a}+1} \cdots X_{2\frac{m}{a}}) \cdots (X_{m-\frac{m}{a}+1} \cdots X_m).$$

Each term in parenthesis can be computed (brutally) via a $\Sigma\Pi^{\frac{m}{a}}$-circuit of size $n^{\frac{m}{a}}$. After getting the resulting matrices, we can compute the rest via a $\Sigma\Pi^a$ circuit of size n^a. This reduces one to a depth four circuit of size $s' = 2^{O(\sqrt{d \log d \log s \log n})}$ Then we can get a depth five powering circuit using (7.1.6).

The new circuit has size $O(s')$ and is of the form $\Sigma\Lambda\Sigma\Lambda\Sigma$. Finally, they use (6.1.6) to convert the power sums to elementary symmetric functions which keeps the size at $O(s')$ and drops the depth to three.

7.2. Geometry and shallow circuits

There is a simple geometric reformulation of $\Sigma\Lambda\Sigma\Lambda\Sigma$ circuits given in §7.2.1. There is a natural geometric reformulation of homogeneous depth three circuits described in §7.2.2, namely via the variety $\sigma_r(Ch_d(\mathbb{P}V))$. Unfortunately, homogeneous depth three circuits are next to useless, as is explained in §7.2.4. To make use of the variety $\sigma_r(Ch_d(\mathbb{P}V))$, despite it being only useful for homogeneous depth three circuits while Theorem 7.1.3.1 requires arbitrary depth three circuits, one works with padded polynomials, as I explain in §7.2.5.

7.2.1. Geometric reformulation of homogeneous $\Sigma\Lambda\Sigma\Lambda\Sigma$ circuits. Recall that computer scientists always work in bases and the inputs to the circuits are constants and variables. For homogeneous circuits, the inputs are simply the variables. The first layer of such a circuit is just to obtain arbitrary linear forms from these variables, so it plays no role in the geometry. The second layer sends a linear form ℓ to ℓ^δ, i.e., we are forming points of $v_\delta(\mathbb{P}V)$. The next layer consists of addition gates, which means we obtain sums of d-th powers, i.e., points of $\sigma_r(v_\delta(\mathbb{P}V))$. Then at the next layer, we take Veronese re-embeddings of these secant varieties to obtain points of $v_{\delta'}(\sigma_r(v_\delta(\mathbb{P}V)))$, and in the final addition gate we obtain a point of $\sigma_{r'}(v_{\delta'}(\sigma_r(v_\delta(\mathbb{P}V))))$. Thus we may rephrase Theorem 7.1.3.1(2) as:

Proposition 7.2.1.1. [**Lan14a**] *Let $d = n^{O(1)}$ and let $P \in S^d\mathbb{C}^n$ be a polynomial sequence that can be computed by a circuit of size s. Then $[P] \in \sigma_{r_1}(v_{\frac{d}{\delta}}(\sigma_{r_2}(v_\delta(\mathbb{P}^{n-1}))))$ with roughly $\delta \sim \sqrt{d}$ and $r_1 r_2 \sim s^{\sqrt{d}}$, more precisely $r_1 r_2 \delta = 2^{O(\sqrt{d \log(ds) \log(n)})}$.*

Corollary 7.2.1.2. [**GKKS13b**] *If for all but finitely many m, $\delta \simeq \sqrt{m}$, and all r_1, r_2 such that $r_1 r_2 = 2^{\sqrt{m}\log(m)\omega(1)}$, one has $[\mathrm{perm}_m] \notin \sigma_{r_1}(v_{m/\delta}(\sigma_{r_2}(v_\delta(\mathbb{P}^{m^2-1}))))$,* *then there is no circuit of polynomial size computing the permanent, i.e.,* **VP** $\neq$ **VNP**.

Problem 7.2.1.3. Find equations for $\sigma_{r_1}(v_\delta(\sigma_{r_2}(v_\delta(\mathbb{P}^{m^2-1}))))$.

7.2.2. Multiplicative joins and depth four circuits. Following [**Lan10**], for varieties $X \subset \mathbb{P}S^aW$ and $Y \subset \mathbb{P}S^bW$, define the *multiplicative join* of X and Y, $MJ(X,Y) := \{[xy] \mid [x] \in X,\ [y] \in Y\} \subset \mathbb{P}S^{a+b}W$, and define $MJ(X_1, \dots, X_k)$ similarly. Let $\mu_k(X) = MJ(X_1, \dots, X_k)$ when all the $X_j = X$, which is a multiplicative analog of the secant variety. Note that $\mu_k(\mathbb{P}W) = Ch_k(W)$. The varieties associated to the polynomials computable by bounded depth formulas are of the form $\sigma_{r_k}(\mu_{d_{k-1}}(\sigma_{r_{k-2}}(\cdots \mu_{d_1}(\mathbb{P}W)\cdots)))$, and $\mu_{d_{k+1}}(\sigma_{r_k}(\mu_{d_{k-1}}(\sigma_{r_{k-2}}(\cdots \mu_{d_1}(\mathbb{P}W)\cdots))))$. In particular, a $\Sigma^r\Pi^\alpha\Sigma^s\Pi^\beta$ circuit computes (general) points of $\sigma_r(\mu_\alpha(\sigma_s(\mu_\beta(\mathbb{P}W))))$.

7.2.3. Secant varieties and homogeneous depth three circuits. The relation between secant varieties of Chow varieties and depth three circuits is as follows:

Proposition 7.2.3.1. *A polynomial $P \in S^nW$ in $\sigma_r^0(Ch_n(W))$ is computable by a homogeneous depth three circuit of size $r + nr(1+\mathbf{w})$. If $P \notin \sigma_r^0(Ch_n(W))$, then P cannot be computed by a homogeneous depth three circuit of size $n(r+1)+(r+1)$.*

Proof. In the first case, $P = \sum_{j=1}^r (x_{1j}\cdots x_{nj})$ for some $x_{sj} \in W$. Expressed in terms of a fixed basis of W, each x_{sj} is a linear combination of at worst $\mathbf{w}$ basis vectors, thus to create each one requires at worst $nr\mathbf{w}$ additions. Then to multiply them in groups of n is nr multiplications, and finally to add these together is r further additions. In the second case, at best P is in $\sigma_{r+1}^0(Ch_n(W))$, in which case, even if each of the x_{sj}'s is a basis vector (so no initial additions are needed), we still must perform $n(r+1)$ multiplications and $r+1$ additions. $\square$

I first explain why the computer science literature generally allows inhomogeneous depth three circuits, and then why one does not need to do so.

7.2.4. Why homogeneous depth three circuits do not appear useful at first glance. Recalling (see §7.1.1) that $\binom{2m}{m} \sim \frac{4^m}{\sqrt{\pi m}}$, by (7.1.8) we have $[\det_n], [\mathrm{perm}_n] \notin \sigma_{O(\frac{4^n}{n})} v_n(\mathbb{P}W)$. On the other hand, (7.1.6), implies

$$\sigma_r(Ch_n(W)) \subset \sigma_{r2^n}(v_n(\mathbb{P}W)).$$

We conclude, for any constant C and n sufficiently large, that

$$\det_n \notin \sigma_{C\frac{2^n}{n}}(Ch_n(W)),$$

and similarly for the permanent. By Proposition 7.2.3.1, we conclude:

Proposition 7.2.4.1. [**NW97**] *The polynomial sequences* $\det_n$ *and* perm_n *do not admit homogeneous depth three circuits of size* 2^n.

Thus homogeneous depth three circuits at first sight do not seem that powerful because a polynomial sized homogeneous depth 3 circuit cannot compute the determinant.

To make matters worse, consider the polynomial corresponding to iterated matrix multiplication of three by three matrices $IMM_k^3 \in S^k(\mathbb{C}^{9k})$. It is complete for $\mathbf{VP}_e$ of Remark 6.1.5.2, and also has an exponential lower bound for its Chow border rank:

Exercise 7.2.4.2: Use flattenings to show $\underline{\mathbf{R}}_S(IMM_k^3) \geq (const.)3^k$, and conclude $IMM_k^3 \notin \sigma_{poly(k)}(Ch_k(W))$.

By Exercise 7.2.4.2, sequences of polynomials admitting polynomial size formulas do not in general have polynomial size depth three circuits.

7.2.5. Homogeneous depth three circuits for padded polynomials. If one works with padded polynomials instead of polynomials (as we did with $\mathcal{D}et_n$), the power of homogeneous depth three circuits increases dramatically. (As mentioned above, in [**GKKS13b**] and elsewhere they consider inhomogeneous polynomials and circuits instead of padding.) The following geometric version of a result of Ben-Or (presented below as a Corollary) was suggested by K. Efremenko:

Proposition 7.2.5.1. *Let* $\mathbb{C}^{m+1}$ *have coordinates* $\ell, x_1, \dots, x_m$ *and let* $e_m^k = e_m^k(x_1, \dots, x_m)$. *For all* $k \leq m$, $\ell^{m-k}e_m^k \in \sigma_m^0(Ch_m(\mathbb{C}^{m+1}))$.

Proof. Fix an integer $u \in \mathbb{Z}$ and define

$$\begin{aligned} g_u(x,\ell) &= (u\ell)^m E_m(\frac{1}{u\ell}) \\ &= \prod_{i=1}^m (x_i + u\ell) \\ &= \sum_k u^{m-k} e_m^k(x) \ell^{m-k}. \end{aligned}$$

Note $g_u(x,\ell) \in Ch_m(\mathbb{C}^{m+1})$. Letting $u = 1, \dots, m$, we may use the inverse of the Vandermonde matrix to write each $\ell^{m-k}e_m^k$ as a sum of m points in $Ch_m(\mathbb{C}^{m+1})$ because

$$\begin{pmatrix} 1^0 & 1^1 & \cdots & 1^m \\ 2^0 & 2^1 & \cdots & 2^m \\ & \vdots & & \\ m^0 & m^1 & \cdots & m^m \end{pmatrix} \begin{pmatrix} \ell^{m-1}e_m^1 \\ \ell^{m-2}e_m^2 \\ \vdots \\ \ell^0 e_m^m \end{pmatrix} = \begin{pmatrix} g_1(x,\ell) \\ g_2(x,\ell) \\ \vdots \\ g_m(x,\ell) \end{pmatrix}.$$

$\square$

Corollary 7.2.5.2 (Ben-Or). *$\ell^{m-k}e^k_m$ can be computed by a homogeneous depth three circuit of size $3m^2 + m$.*

Proof. As remarked above, for any point of $\sigma_r Ch_n(\mathbb{C}^{m+1})$ one gets a circuit of size at most $r + nr + rn(m+1)$, but here at the first level all the addition gates have fanin two (i.e., there are two inputs to each addition gate) instead of the possible $m + 1$. $\square$

Remark 7.2.5.3. The best lower bound for computing the e^k_n via a $\Sigma\Pi\Sigma$ circuit is $\Omega(n^2)$ [**SW01**], so Corollary 7.2.5.2 is very close to (and may well be) sharp.

Proposition 7.2.5.4. *Say $P \in S^m\mathbb{C}^M$ is computable by a depth three circuit of size s. Then $\ell^{n-m}P$ is computable by a homogeneous depth three circuit of size $O(s^2)$.*

Proof. Start with the inhomogeneous circuit computing P. At the first level, add a homogenizing variable ℓ, so that the affine linear outputs become linear in our original variables plus ℓ, the product gates will each produce a homogeneous polynomial. While the different product gates may produce polynomials of different degrees, if we were trying to produce a homogeneous polynomial, when we add them up what remains must be a sum of homogeneous polynomials, such that when we set $\ell = 1$, we obtain the desired homogeneous polynomial. Say the largest power of ℓ appearing in this sum is q_L. Note that $q_L < s$. For each other term there is some other power of ℓ appearing, say q_i for the i-th term. Then to the original circuit, add $q_L - q_i$ inputs to the i-th product gate, where each input is ℓ. This will not change the size of the circuit by more than $q_L r < s^2$. Our new homogeneous depth three circuit will output $\ell^{q_L}P$. $\square$

In geometric language:

Proposition 7.2.5.5. [**Lan14a**] *Let $d = N^{O(1)}$ and let $P \in S^d\mathbb{C}^N$ be a polynomial that can be computed by a circuit of size s.*

Then $[\ell^{n-d}P] \in \sigma_r(Ch_n(\mathbb{C}^{N+1}))$ with roughly $rn \sim s^{\sqrt{d}}$, more precisely, $rn = 2^{O(\sqrt{d\log(N)\log(ds)})}$.

Corollary 7.2.5.6. [**GKKS13b**] *$[\ell^{n-m}\det_m] \in \sigma_r(Ch_n(\mathbb{C}^{m^2+1}))$ where $rn = 2^{O(\sqrt{m}\log m)}$.*

Proof. The determinant admits a circuit of size m^4, so it admits a $\Sigma\Pi\Sigma$ circuit of size

$$2^{O(\sqrt{m\log(m)\log(m*m^4)})} = 2^{O(\sqrt{m}\log m)},$$

so its padded version lies in $\sigma_r(Ch_n(\mathbb{C}^{m^2+1}))$ where $rn = 2^{O(\sqrt{m}\log m)}$. $\square$

Corollary 7.2.5.7. [**GKKS13b**] *If for all but finitely many m and all r,n with $rn = 2^{\sqrt{m}\log(m)\omega(1)}$, one has $[\ell^{n-m}\operatorname{perm}_m] \notin \sigma_r(Ch_n(\mathbb{C}^{m^2+1}))$, then there is no circuit of polynomial size computing the permanent, i.e.,* $\mathbf{VP} \neq \mathbf{VNP}$.

Proof. One just needs to observe that the number of edges in the first layer (which are invisible from the geometric perspective) is dominated by the number of edges in the other layers. □

Remark 7.2.5.8. The expected dimension of $\sigma_r(Ch_m(W))$ is $rm\mathbf{w}+r-1$. If we take n and work instead with padded polynomials $\ell^{n-m}P$, the expected dimension of $\sigma_r(Ch_n(W))$ is $rn\mathbf{w}+r-1$. In contrast, the expected dimension of $\sigma_r(v_{d-a}(\sigma_\rho(v_a(\mathbb{P}W))))$ does not change when one increases the degree, which gives some insight as to why padding is so useful for homogeneous depth three circuits but not for $\Sigma\Lambda\Sigma\Lambda\Sigma$ circuits.

7.3. Algebraic branching programs and determinants

7.3.1. Algebraic branching programs and iterated matrix multiplication.

Definition 7.3.1.1 (Nisan [**Nis91**]). An *Algebraic Branching Program* (ABP) over $\mathbb{C}$ is a directed acyclic graph Γ with a single source s and exactly one sink t. Each edge e is labeled with an affine linear function ℓ_e in the variables $\{y^i | 1 \leq i \leq M\}$. Every directed path $p = e_1e_2\cdots e_k$ represents the product $\Gamma_p := \prod_{j=1}^k \ell_{e_j}$. For each vertex v the polynomial Γ_v is defined as $\sum_{p\in\mathcal{P}_{s,v}} \Gamma_p$ where $\mathcal{P}_{s,v}$ is the set of paths from s to v. We say that Γ_v *is computed by* Γ *at* v. We also say that Γ_t *is computed by* Γ or that Γ_t *is the output of* Γ.

The *size* of Γ is the number of vertices. Let $\operatorname{abpc}(P)$ denote the smallest size of an algebraic branching program that computes P.

An ABP is *layered* if we can assign a layer $i \in \mathbb{N}$ to each vertex such that for all i, all edges from layer i go to layer $i+1$. Let $\operatorname{labpc}(P)$ denote the the smallest size of a layered algebraic branching program that computes P. Of course $\operatorname{labpc}(P) \geq \operatorname{abpc}(P)$.

An ABP is *homogeneous* if the polynomials computed at each vertex are all homogeneous.

A homogeneous ABP Γ is *degree layered* if Γ is layered and the layer of a vertex v coincides with the degree of v. For a homogeneous P let $\operatorname{dlabpc}(P)$ denote the the smallest size of a degree layered algebraic branching program that computes P. Of course $\operatorname{dlabpc}(P) \geq \operatorname{labpc}(P)$.

Definition 7.3.1.2. The *iterated matrix multiplication complexity* of a polynomial $P(y)$ in M variables, $\operatorname{immc}(P)$ is the smallest n such that there exists affine linear maps $B_j : \mathbb{C}^M \to \mathrm{Mat}_n(\mathbb{C})$, $j = 1, \ldots, n$, such that $P(y) = \mathrm{trace}(B_n(y) \cdots B_1(y))$. The *homogeneous iterated matrix multiplication complexity* of a degree m homogeneous polynomial $P \in S^m\mathbb{C}^M$, $\mathrm{himmc}(P)$, is the smallest n such that there exist natural numbers $n_1, \ldots, n_m$ with $1 = n_1$, and $n = n_1 + \cdots + n_m$, and linear maps $A_s : \mathbb{C}^M \to \mathrm{Mat}_{n_s \times n_{s+1}}$, $1 \leq s \leq m$, with the convention $n_{m+1} = n_1$, such that $P(y) = A_m(y) \cdots A_1(y)$.

In this section we describe how to obtain a size $O(m^3)$ regular determinantal expression for $\det_m$. We use standard techniques about algebraic branching programs and an algorithm described by Mahajan and Vinay [**MV97**].

Proposition 7.3.1.3. *Let P be a polynomial. Then* $\mathrm{dc}(P) \leq \mathrm{labpc}(P) - 1$. *Moreover, if the constant term of P is zero, then we also have* $\mathrm{rdc}(P) \leq \mathrm{labpc}(P) - 1$.

Proof. From a layered algebraic branching program Γ^{algbp} we create a directed graph Γ^{root} by identifying the source and the sink vertex and by calling the resulting vertex the root vertex. From Γ^{root} we create a directed graph Γ^{loops} by adding at each non-root vertex a loop that is labeled with the constant 1. Let A denote the adjacency matrix of Γ^{loops}. Since Γ^{algbp} is layered, each path from the source to the sink in Γ^{algbp} has the same length. If that length is even, then $\det(A)$ equals the output of Γ^{algbp}, otherwise $-\det(A)$ equals the output of Γ^{algbp}. This proves the first part.

Now assume P has no constant term. Let Λ denote the constant part of A, so Λ is a complex square matrix. Since Γ^{algbp} is layered we ignore all edges coming out of the sink vertex of Γ^{algbp} and order all vertices of Γ^{algbp} topologically, i.e., if there is an edge from vertex u to vertex v, then u precedes v in the order. We use this order to specify the order in which we write down Λ. Since the order is topological, Λ is lower triangular with one exception: The first row can have additional nonzero entries. By construction of the loops in Γ^{loops} the main diagonal of Λ is filled with 1s everywhere but at the top left where Λ has a 0. Thus $\mathrm{corank}(\Lambda) = 1$ or $\mathrm{corank}(\Lambda) = 0$. But if $\mathrm{corank}(\Lambda) = 0$, then the constant term of P is $\det(\Lambda) \neq 0$, which is a contradiction to the assumption. $\square$

Proposition 7.3.1.4. $\mathrm{labpc}(\det_m) \leq \frac{m^3}{3} - \frac{m}{3} + 2$.

Proof. This is an analysis of the algorithm in [**MV97**] with all improvements that are described in the article. We construct an explicit layered ABP Γ. Each vertex of Γ is a triple of three nonnegative integers (h, u, i), where i indicates its layer. The following triples appear as vertices in Γ.

- The source $(1, 1, 0)$.

- For all $1 \le i < m$:
  - The vertex $(i+1, i+1, i)$.
  - For each $2 \le u \le m$ and each $1 \le h \le \min(i,u)$ the vertex (h,u,i).
- The sink $(1,1,m)$.

Lemma 7.3.1.5. *The number of vertices in Γ is $\frac{m^3}{3} - \frac{m}{3} + 2$. There is only the source vertex in layer 0 and only the sink vertex in layer m. The number of vertices in layer $i \in \{1, \dots, m-1\}$ is $i(i+1)/2 + i(m-1)$.*

Proof. By the above construction, the number of vertices in Γ equals

$$2 + \sum_{i=1}^{m-1} \Big(1 + \sum_{u=2}^{m} \min(i,u)\Big) = 1 + m + \sum_{i=1}^{m-1} \sum_{u=2}^{m} \min(i,u).$$

We see that $\sum_{i=1}^{m-1} \sum_{u=2}^{m} \min(i,u) = (m-2)(m-1)/2 + \sum_{i=1}^{m-1} \sum_{u=1}^{m-1} \min(i,u)$. It is easy to see that $\sum_{i=1}^{m-1} \sum_{u=1}^{m-1} \min(i,u)$ yields the square pyramidal numbers (OEIS[1] A000330): $m(m-1)(m-\frac{1}{2})/3$. Therefore

$$1+m+\sum_{i=1}^{m-1} \sum_{u=2}^{m} \min(i,u) = 1+m+m(m-1)(m-\tfrac{1}{2})/3+(m-2)(m-1)/2 = \tfrac{m^3}{3}-\tfrac{m}{3}+2.$$

To analyze a single layer $1 \le i \le m-1$ we observe

$$1 + \sum_{u=2}^{m} \min(i,u) = \sum_{u=1}^{m} \min(i,u) = i(i+1)/2 + i(m-i).$$

□

We now describe the edges in Γ. The vertex (h,u,i) is positioned in the ith layer with only edges to the layer $i+1$, with the exception that layer $m-1$ has edges only to the sink. From (h,u,i) we have the following outgoing edges.

- If $i+1 < m$:
  - for all $h+1 \le v \le m$ an edge to $(h,v,i+1)$ labeled with x_v^u.
  - for all $h+1 \le h' \le m$ an edge to $(h',h',i+1)$ labeled with $-x_h^u$.
- If $i+1 = m$: An edge to the sink labeled with αx_h^u, where $\alpha = 1$ if m is odd and $\alpha = -1$ otherwise.

The fact that Γ actually computes $\det_m$ follows from [**MV97**]. □

As an illustration for $m = 3,4,5$ we include the adjacency matrices of the Γ^{loops} that come out of the combination of the constructions in Proposition 7.3.1.4 and Proposition 7.3.1.3.

[1] http://oeis.org/

0	0	0	0	x21	x31	x22	x32	x33
x12	1	0	0	0	0	0	0	0
x13	0	1	0	0	0	0	0	0
-x11	0	0	1	0	0	0	0	0
0	x22	x32	0	1	0	0	0	0
0	x23	x33	0	0	1	0	0	0
0	-x21	-x31	0	0	0	1	0	0
0	0	0	x23	0	0	0	1	0
0	-x21	-x31	-x22	0	0	0	0	1

0	0	0	0	0	0	0	0	0	0	0	0	-x21	-x31	-x41	-x22	-x32	-x42	-x33	-x43	-x44
x12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x13	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
x14	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-x11	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	x22	x32	x42	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	x23	x33	x43	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	x24	x34	x44	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	-x21	-x31	-x41	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	x23	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	x24	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	-x21	-x31	-x41	-x22	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	x22	x32	x42	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	x23	x33	x43	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	x24	x34	x44	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	-x21	-x31	-x41	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	x23	x33	x43	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	x24	x34	x44	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	-x21	-x31	-x41	-x22	-x32	-x42	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	x34	0	0	0	0	0	0	0	1	0
0	0	0	0	0	-x21	-x31	-x41	-x22	-x32	-x42	-x33	0	0	0	0	0	0	0	0	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	x21	x31	x41	x51	x22	x32	x42	x5
x12	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
x13	0	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
x14	0	0	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
x15	0	0	0	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
-x11	0	0	0	0	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	x22	x32	x42	x52	0	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	x23	x33	x43	x53	0	0	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	x24	x34	x44	x54	0	0	0	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	x25	x35	x45	x55	0	0	0	0	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	-x21	-x31	-x41	-x51	0	0	0	0	0	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	x23	0	0	0	0	0	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	x24	0	0	0	0	0	0	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	x25	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	-x21	-x31	-x41	-x51	-x22	0	0	0	0	0	0	0	0	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	x22	x32	x42	x52	0	0	0	0	0	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	x23	x33	x43	x53	0	0	0	0	0	0	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	x24	x34	x44	x54	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	x25	x35	x45	x55	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	-x21	-x31	-x41	-x51	0	0	0	0	0	0	0	0	0	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	x23	x33	x43	x53	0	0	0	0	0	0	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	x24	x34	x44	x54	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	x25	x35	x45	x55	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	-x21	-x31	-x41	-x51	-x22	-x32	-x42	-x52	0	0	0	0	0	0	0	0	0	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	x34	0	0	0	0	0	0	0	0	0	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	x35	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	-x21	-x31	-x41	-x51	-x22	-x32	-x42	-x52	-x33	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x22	x32	x42	x52	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x23	x33	x43	x53	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x24	x34	x44	x54	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x25	x35	x45	x55	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-x21	-x31	-x41	-x51	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x23	x33	x43	x53	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x24	x34	x44	x54	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	x25	x35	x45	x55	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-x21	-x31	-x41	-x51	-x22	-x32	-x42	-x52	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	x34	x44	x54	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	x35	x45	x55	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-x21	-x31	-x41	-x51	-x22	-x32	-x42	-x52	-x33	-x43	-x53	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	x45	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-x21	-x31	-x41	-x51	-x22	-x32	-x42	-x52	-x33	-x43	-x53	-x44	0	0	0	0	0	0	0	0

Add discussion on appearances, motivating read k and rank k

7.3.2. Determinantal complexity and ABP's. The following result, while "known to the experts", is not easily accessible in the literature.

Moreover, we give a precise formulation to facilitate measuring benchmark progress in different models.

In the following theorem note that himmc and dlabpc are only defined for homogeneous polynomials.

Theorem 7.3.2.1. *The complexity measures* rdc, dc, labpc, immc, abpc, himmc, *and* dlabpc *are all polynomially related. More precisely, let P be any polynomial. Let $\varphi(m) := \frac{m^3}{3} - \frac{m}{3} + 2$ denote the layered ABP size of the Mahajan-Vinay construction for* $\det_m$. *Then*

(1) $\mathrm{dc}(P) \leq \mathrm{labpc}(P) - 1$. *If P has no constant part, then* $\mathrm{rdc}(P) \leq \mathrm{labpc}(P) - 1$.

(2) $\mathrm{labpc}(P) \leq \varphi(\mathrm{dc}(P))$.

(3) *By definition* $\mathrm{dc}(P) \leq \mathrm{rdc}(P)$. *If P has no constant part, then* $\mathrm{rdc}(P) \leq \varphi(\mathrm{dc}(P)) - 1$. *If codim$(P_{sing}) \geq 5$, then* $\mathrm{rdc}(P) = \mathrm{dc}(P)$.

(4) $\mathrm{labpc}(P) = \mathrm{immc}(P) + 1$. *If P is homogeneous, then* $\mathrm{dlabpc}(P) = \mathrm{himmc}(P) + 1$.

(5) *definition* $\mathrm{abpc}(P) \leq \mathrm{labpc}(P) \leq \mathrm{dlabpc}(P)$, *where* $\mathrm{dlabpc}(P)$ *is defined only if P is homogeneous. If P is homogeneous of degree d then* $\mathrm{dlabpc}(P) \leq (d+1)\,\mathrm{abpc}(P)$.

Remark 7.3.2.2. It is an important and perhaps tractable open problem to prove an $\omega(m^2)$ lower bound for $\mathrm{dc}(\mathrm{perm}_m)$. By Theorem 7.3.2.1, it would suffice to prove an $\omega(m^6)$ lower bound for $\mathrm{himmc}(\mathrm{perm}_m)$.

Remark 7.3.2.3. The computation model of homogeneous iterated matrix multiplication has the advantage that one is comparing the homogeneous iterated matrix multiplication polynomial himm directly with the permanent, whereas with the determinant $\det_n$, one must compare with the *padded permanent* $\ell^{n-m}\,\mathrm{perm}_m$. The padding causes insurmountable problems if one wants to find *occurrence obstructions* in the sense of [**MS01, MS08**]. The problem was first observed in [**KL14**] and then proved insurmountable in [**IP15**] and [**BIP16**]. Thus *a priori* it might be possible to prove Valiant's conjecture via occurrence obstructions in the himmc model. However, with the determinant already one needed to understand difficult properties about three factor Kronecker coefficients, and for the himmc model, one would need to prove results about m-factor Kronecker coefficients, which are not at all understood.

Regarding the geometric search for separating equations, the advantage one gains by removing the padding is offset by the disadvantage of dealing with the himmc polynomial that for all known equations such as Young flattenings (which includes the method of shifted partial derivatives as a special case) and equations for degenerate dual varieties, behaves far more generically than the determinant.

Remark 7.3.2.4. One can also show that if P is any polynomial of degree d, then $\mathrm{labpc}(P) \leq d(\mathrm{abpc}(P)^2)$.

Proof. (1) is Proposition 7.3.1.3.

Proof of (2): We first write the determinant polynomial $\det_{\mathrm{dc}(P)}$ as a size $\varphi(\mathrm{dc}(P))$ layered ABP Γ using 7.3.1.4. The projection that maps $\det_{\mathrm{dc}(P)}$ to P can now be applied to Γ to yield a size $\varphi(\mathrm{dc}(P))$ layered ABP of P.

Proof of (3): To see the second inequality we combine (1) and (2). The last assertion is von zur Gathen's result [**vzG87**].

Proof of (4): We prove $\mathrm{labpc}(P) \leq \mathrm{immc}(P)+1$. Given $n_1, \dots, n_m$ with $n_1 = 1$ and $n_1 + \cdots + n_m = \mathrm{immc}(P)$ and linear maps B_j, $1 \leq j \leq m$, we construct the ABP Γ that has a single vertex at level $m+1$, n_j vertices at level j, $1 \leq j \leq m$, and is the complete bipartite graph between levels. The labels of Γ are given by the B_j. We now prove $\mathrm{immc}(P) \leq \mathrm{labpc}(P) - 1$. Given a layered ABP Γ with $m+1$ layers, recall that by definition Γ has only 1 vertex in the top layer and only one vertex in the bottom layer. Let n_j denote the number of vertices in layer j, $1 \leq j \leq m$. Define the linear maps B_j by reading off the labels between layer j and layer $j+1$. The proof of the second claim is analogous.

Proof of (5): We first homogenize and then adjust the ABP. Replace each vertex v other than s by $d+1$ vertices $v^1, v^2, \dots, v^{d+1}$ corresponding to the homogeneous parts of Γ_v. Replace each edge e going from a vertex v to a vertex w by $(2d+1)$ edges, where we split the linear and constant parts: If e is labeled by $\ell + \delta$, where ℓ is linear and $\delta \in \mathbb{C}$, the edge from v^i to w^i, $1 \leq i \leq d$, is labeled with δ and the edge from v^i to w^{i+1}, $1 \leq i \leq d-1$, is labeled with ℓ. We now have a homogeneous ABP. Our task is to make it degree layered. As a first approach we assign each degree i vertex to be in layer i, but there may be edges labeled with constants between vertices in the same layer. The edges between vertices of different layers are linear forms. Call the vertices in layer i that have edges incoming from layer $i-1$, *layer i entry vertices.* Remove the non-entry vertices. From entry vertex of layer i to entry vertex of layer $i+1$, use the *linear form* computed by the sub-ABP between them. In other words, for every pair (v, w) of layer i entry vertex v and layer $i+1$ entry vertex w, put an edge from v to w with weight

$$\sum_p \Pi_e \mathrm{weight}(e)$$

where the sum is over paths p from v to w and the product is over edges in the path p. The resulting ABP is degree homogeneous and computes P. $\square$

7.4. Additional restricted models

7.4.1. Elementary symmetric polynomial complexity. Let $P \in S^m V$ and define the *elementary symmetric complexity* of P, $\mathrm{esc}(P)$, to be the smallest N such that $P \in \mathrm{End}(\mathbb{C}^N) \cdot e_{m,N} =: \hat{\mathcal{E}}^0_{m,N}$, and $\overline{\mathrm{esc}}(P)$ to be the smallest N such that $P \in \overline{\mathrm{End}(\mathbb{C}^N) \cdot e_{m,N}} = \overline{GL_N \cdots e_{m,N}} =: \hat{\mathcal{E}}_{m,N}$. Shpilka [**Shp02**] refers to $\mathrm{esc}(P)$ as the "size of the smallest depth two circuit with a symmetric gate at the top and plus gates at the bottom". (His circuits have the output gate at the top.)

First this is a legitimate complexity model: for any polynomial P, $\mathrm{esc}(P)$ is finite. In fact, we have the more precise:

Proposition 7.4.1.1. [**Shp02**] *$\sigma^0_r(v_m(\mathbb{P}V)) \subset \mathcal{E}^0_{m,rm}$ and $\sigma_r(v_m(\mathbb{P}V)) \subset \mathcal{E}_{m,rm}$.*

Proof. Without loss of generality, assume $\mathbf{v} = r$ and let $y_1, \ldots, y_r$ be a basis of V. It will be sufficient to show $\sum y_j^m \in \mathcal{E}^0_{m,mr}$. Let ω be a primitive m-th root of unity. Then I claim

$$\sum y_j^m = -e_{m,rm}(y_1, -\omega y_1, -\omega^2 y_1, \ldots, \omega^{m-1} y_1, -y_2, -\omega y_2, \ldots, -\omega^{m-1} y_r).$$

To see this, evaluate the generating function:

$$\begin{aligned}&E_{rm}(t)(y_1, -\omega y_1, -\omega^2 y_1, \ldots, \omega^{m-1} y_1, -y_2, -\omega y_2, \ldots, -\omega^{m-1} y_r)\\&= \Pi_{i\in[r]}\Pi_{s\in[m]}(1-\omega^s y_i)\\&= \Pi_{i\in[r]}(1-y_i^m t^m)\end{aligned}$$

but the coefficient of t^m on the last line is $-\sum_i y_i^m$. □

Corollary 7.2.5.2 implies that $\mathrm{esc}(P)$ is at least the square root of the size of the smallest depth three circuit computing P.

Shpilka proves lower bounds for esc in the same way the first lower bounds for dc were found: by considering linear spaces on $Z(e_{m,N})$.

Theorem 7.4.1.2. [**Shp02**] *Let $L \subset Z(e_{m,N}) \subset \mathbb{P}V^*$ be a linear space. Then $\dim L \leq \min(\max(N-m, m-1), \frac{m+N}{2}) - 1$.*

Thus if $Z(P)$ has large linear spaces on it we obtain lower bounds for $\mathrm{esc}(P)$.

Proof. The key to the proof is the algebraic independence of the $e_{k,N}$. Note that if we have two sets of variables $(x, y) = (x_1, \ldots, x_k, y_1, \ldots, y_{N-k})$, then

$e_{m,N}(x,y) = \sum_{j=0}^{m} e_{m-j,k}(x)e_{j,N-k}(y)$. We are assuming $e_{m,N}|_{\hat{L}} = 0$, so

$$0 = e_{m,N}(x,\ell)$$
$$= e_{m,k}(x) + \sum_{j=1}^{m} e_{m-j,k}(x)e_j(\ell(x)). \tag{7.4.1}$$

First assume $k-1 = \dim L \geq \max(N-m, m-1)$. Since $e_{k,u} = 0$ if $k > u$, if $N-k < m$ the sum in (7.4.1) is from 1 to $N-k$. Our linear space will have an isomorphic projection onto some coordinate k-plane, without loss of generality, assume it is the first, so that L has equations $x_s = \ell_s(x_1,\dots,x_k)$ for $k+1 \leq s \leq N$.

Now let $\Psi : \mathbb{C}[x_1,\dots,x_k] \to \mathbb{C}[x_1,\dots,x_k]^{\mathfrak{S}_k}$ denote the symmetrization operator and recall that it is a ring homomorphism, so in particular $\Psi(fg) = \Psi(f)\Psi(g)$ and $\Psi(f+g) = \Psi(f) + \Psi(g)$. Apply Ψ to (7.4.1) to obtain

$$0 = e_{m,k}(x) + \sum_{j=1}^{N-k} e_{m-j,k}(x)\Psi(e_j(\ell(x)))$$

but this expresses $e_{m,k}$ as a polynomial in symmetric functions of degree less than k, a contradiction.

Now assume $\dim \hat{L} \geq \frac{m+N}{2}$, so we have

$$0 = e_{m,k}(x) + e_{m,N-k}(\ell(x)) + \sum_{j=1}^{m} e_{m-j,k}(x)e_j(\ell(x)).$$

The idea is again the same, but we must somehow reduce to a smaller space. If we take $D \in \{\ell_1,\dots,\ell_N - k\}^{\perp} \subset V^*$ and apply it, we can eliminate the $e_{m,N-k}(\ell(x))$ term. But if we take a random such D, we will no longer have symmetric functions. However, one can find a D such that, if we restrict to span of the the first $m-1$ coordinate vectors, call this space $V_{m-1} \subset \mathbb{C}^k \subset \mathbb{C}^N$, then $De_{r,k}|_{V_{m-1}} = e_{r-1,m-1}$. Just take D of the form $D = \sum_{j=1}^{m-1} \frac{d}{dx_j} + D'$ (Such a D always exists by counting dimensions.) Unfortunately this is still not good enough, as letting $x' = (x_1,\dots,x_{m-1})$ we now have

$$0 = e_{m-1,m-1}(x') \sum_{j=1}^{m} e_{m-j,k}(x')e_j(\ell(x')).$$

We could argue as before if we could eliminate the $j=1$ term. But we can! since $k \geq \frac{m+N}{2}$, one can also assume $D(e_{1,k}(x)) = 0$. □

Exercise 7.4.1.3: (1) Show $\overline{\text{esc}}(\det_m) \geq 2m^2 - 3m$.

Exercise 7.4.1.4: (1) Show that if $m \geq \frac{N+1}{2}$, there exists a linear space of dimension $d-1$ on $Z(e_{m,N})$. ⊚

Proposition 7.4.1.5. [**Shp02**] *(attributed to Saks) There exists a* $\mathbb{P}^{\lfloor \frac{N}{q} \rfloor - 1} \subset Z(e_{m,N})$*, where* q *is the smallest integer such that* q *does not divide* m*.*

Proof. Let ω be a primitive q-th root of unity. Let $e_1, \ldots, e_N$ denote the standard basis of $\mathbb{C}^N$. Consider

$$\hat{L} := \operatorname{span}\{e_{1+jq} + \omega e_{2+jq} + \omega^2 e_{3+jq} + \cdots + \omega^{q-1} e_{q+jq} \mid j = 0, \ldots, \lfloor \frac{N}{q} \rfloor - q\}$$

Note that all the power sum polynomials $p_{r,N}$, $1 \leq r \leq m$ vanish on $\hat{L}$, so L is contained in the hypersurface defined by any symmetric function of degree m. $\square$

By Exercise 7.4.1.4 and Proposition 7.4.1.5, we see Theorem 7.4.1.2 is close to being sharp.

The following conjecture appeared in [**Shp02**] (phrased differently):

Conjecture 7.4.1.6. [**Shp02**] *There exists a polynomial* $r(m)$ *such that* $\sigma^0_{r(m)} Ch_m(\mathbb{C}^{mr(m)}) \not\subset \mathcal{E}^0_{m,2^m}$*. In fact one might even be able to take* $r(m) \equiv 2$*.*

The second assertion is astonishing, as when $r = 1$ the two sets coincide, and when $r = 2$ the left hand side has dimension about $4m$ and the right hand side has dimension about 4^m.

Exercise 7.4.1.7: (2) Show that $\sigma_2(Ch_m(\mathbb{C}^{2m})) \not\subset \mathcal{E}_{m,\frac{3}{2}m-3}$.

Question 7.4.1.8. [**Shp02**] What is the maximal dimension of a linear subspace $L \subset \mathbb{P}V^*$ such that $L \subset Z(e_{m,\mathbf{v}})$?

Remark 7.4.1.9. ***remove or move** Strassen [**Str75**] proved a lower bound of $\Omega(n \log n)$ for the size of *any* arithmetic circuit computing all the e^j_n simultaneously.

7.4.2. Non-commutative circuits.

7.4.3. A classical exponential lower bound for the permanent (and determinant).

Here the restriction is that one is not allowed to exploit the commutivity of multiplication. Let $\mathbb{C}\{y_1, \ldots, y_N\}$ denote the ring of polynomials in the non-commuting variables $y_1, \ldots, y_N$. Choose an expression for a polynomial P and consider it as in this larger ring. The definition of circuits is the same here, just that we cannot assume $ab = ba$ for expressions a and b.

Define the *non-commutative algebraic branching program complexity* of a polynomial P, $\mathrm{NCabpc}(P)$ to be the size of the smallest non-commutative ABP that commutes P.

Theorem 7.4.3.1. [**Nis91**] $\mathrm{NCabpc}(\det_n) = \mathrm{NCabpc}(\mathrm{perm}_n) = 2^n - 1$.

Proof. Insert proof □

7.4.4. Column-multilinear HIMM. Theorems 7.4.3.1 and 7.4.7.1 are related.

insert here after write-up with Christian*

7.4.5. Glynn's Theorem on expressions for the permanent. Recall, for $P \in S^m\mathbb{C}^M$, $\mathbf{R}_{Ch_m(\mathbb{C}^M)}(P)$ is the smallest r such that $P(y_1, \dots, y_M) = \sum_{s=1}^r \Pi_{u=1}^m (\sum_{a=1}^M \lambda_{sua} y_a)$ for some constants λ_{sua}. This corresponds to the smallest homogeneous $\Sigma^r\Pi\Sigma^M$ circuit that computes P. If P is multilinear, so $M = mw$ and we may write $y_a = (y_{i\alpha})$ where $1 \leq i \leq m$, $1 \leq \alpha \leq w$, and $P = \sum C_\alpha y_{1\alpha} \cdots y_{m\alpha}$ we could restrict to *multi-linear* $\Sigma\Pi\Sigma$ circuits, those of the form $\sum_{s=1}^r \Pi_{i=1}^m (\sum_{\alpha=1}^w \lambda_{su} y_{i,\alpha})$. Write $\mathbf{R}^{ML}_{Ch_m(\mathbb{C}^M)}(P)$ for the smallest multilinear $\Sigma^r\Pi\Sigma^w$ circuit for such a P. We can consider multilinear $\Sigma\Pi\Sigma$-circuit complexity as a restricted model. In this context, we have the following theorem of Glynn:

Theorem 7.4.5.1. **[Gly13]** $\mathbf{R}^{ML}_{Ch_m(\mathbb{C}^M)}(\mathrm{perm}_m) = \mathbf{R}_S(x_1 \cdots x_m) = 2^{m-1}$.

More precisely, constants $\lambda_{s,j}$, $1 \leq s \leq r$, $1 \leq j \leq m$ satisfy:

$$x_1 \cdots x_m = \sum_{s=1}^{r} (\sum_{j=1}^{m} \lambda_{s,j} x_j)^m \tag{7.4.2}$$

if and only if

$$\mathrm{perm}_m(y_{ij}) = m! \sum_{s=1}^{r} \Pi_{i=1}^m (\sum_{j=1}^{m} \lambda_{s,j} y_{ij}). \tag{7.4.3}$$

Proof. Given a Waring decomposition (7.4.2) of $x_1 \cdots x_m$, set $x_j = \sum_k y_{jk} z_k$. The coefficient of $z_1 \cdots z_m$ in the resulting expression on the left hand side is the permanent and the coefficient of $z_1 \cdots z_m$ on the right hand side is the right hand side of (7.4.3).

To see the other direction, given an expression (7.4.3), we specialize to various matrices to show identities among the $\lambda_{s,j}$ that will imply all coefficients but the desired one on the right hand side are zero.

The coefficient of $x_1^{b_1} \cdots x_m^{b_m}$, where $b_1 + \cdots + b_m = m$ in (7.4.2) is $\binom{m}{b_1, \dots, b_m} \sum_s \lambda_{s,1}^{b_1} \cdots \lambda_{s,m}^{b_m}$.

Let y be a matrix where there are b_j 1's in column j and zero elsewhere. Then unless each $b_j = 1$, $\mathrm{perm}(y) = 0$. But (7.4.3) says that $0 = \mathrm{perm}(y)$ is a nonzero constant times $\sum_s \lambda_{s,1}^{b_1} \cdots \lambda_{s,m}^{b_m}$. Thus all these terms are zero and the only potential nonzero coefficient in the right hand side of (7.4.2) is the coefficient of $x_1 \cdots x_m$. This coefficient is $m! = \binom{m}{1, \dots, 1}$ times $\lambda_{s,1} \cdots \lambda_{s,m}$. Plugging in $y = \mathrm{Id}$ shows $1 = m! \lambda_{s,1} \cdots \lambda_{s,m}$. □

7.4.6. Rank k determinantal expressions. ***add here ***

7.4.7. Equivariant determinantal complexity. Motivated by the symmetry of Grenet's expressions for the permanent discussed in §6.6.3, N. Ressayre and I asked, what happens if one *imposes* the Γ_m^E-equivariance? We found:

Theorem 7.4.7.1. [**LR15**] *Among Γ_m^E-equivariant determinantal expressions for* perm_m*, Grenet's size $2^m - 1$ expressions are optimal and unique up to trivialities.*

The Γ_m^E-equivariance is peculiar as it only makes sense for the permanent. To fix this, we defined a complexity measure that could be applied to all polynomials:

Definition 7.4.7.2. Let $\tilde{A} : V \longrightarrow \mathrm{Mat}_n(\mathbb{C})$ be a determinantal representation of $P \in S^m V^*$. Define

$$G_A = \{g \in G_{\det_n} \mid g \cdot \Lambda = \Lambda \text{ and } g \cdot A(V) = A(V)\},$$

the *symmetry group of the determinantal representation* $\tilde{A}$ of P.

The group G_A comes with a representation $\rho_A : G_A \longrightarrow GL(A(V))$ obtained by restricting the action to $A(V)$. We assume that P cannot be expressed using $\dim(V)-1$ variables, i.e., that $P \notin S^m V'$ for any hyperplane $V' \subset V^*$. Then $A : V \longrightarrow A(V)$ is bijective. Let $A^{-1} : A(V) \longrightarrow V$ denote its inverse. Set

$$\begin{aligned} \bar{\rho}_A : G_A &\longrightarrow GL(V) \\ g &\longmapsto A \circ \rho_A(g) \circ A^{-1}. \end{aligned} \tag{7.4.4}$$

Definition 7.4.7.3. We say $\tilde{A}$ is an *equivariant representation* of P if $\overline{\rho}_A(G_A) = G_P$.

If G is a subgroup of G_P, we say that $\tilde{A}$ is G*-equivariant* if G is contained in the image of $\bar{\rho}_A$.

Definition 7.4.7.4. For $P \in S^m V^*$, define the *equivariant determinantal complexity* of P, denoted $\mathrm{edc}(P)$, to be the smallest n such that there is an equivariant determinantal representation of P.

Note that if P is a generic polynomial, $\mathrm{edc}(P) = \mathrm{dc}(P)$ because it will have a trivial symmetry group. One also has $\mathrm{edc}(\det_m) = \mathrm{dc}(\det_m)$ because taking $\tilde{A} = \mathrm{Id}$ is equivariant.

Theorem 7.4.7.5. [**LR15**] *There exists a G_{perm_m}-equivariant determinantal expression for* perm_m *of size* $\binom{2m}{m} - 1 \sim 4^m$.

Theorem 7.4.7.6. [**LR15**] *Among G_{perm_m}-equivariant determinatal expressions for* perm_m*, the size* $\binom{2m}{m} - 1$ *expressions are optimal and unique up to trivialities.*

In particular, Valiant's conjecture holds in the restricted model of equivariant expressions.

Proofs are given in §8.12.4.

7.5. Permanent (and determinant) v. Shallow circuits

In this section I describe work of Gupta, Kamath, Kayal, and Saptharishi [**GKKS13a**] that generated considerable excitement, winning the best paper award at the 2013 Conference on Computational Complexity (CCC) because it came tantalizingly close to proving Valiant's conjecture by showing that the permanent does not admit a size $2^{o(\sqrt{n})}$ depth four circuit with bottom fanin bounded by $\sqrt{n}$. Compare this with Theorem 7.1.3.1 that implies it would be sufficient to show that perm_m is not computable by a homogeneous $\Sigma\Pi\Sigma\Pi$ circuit of size $2^{\Omega(\sqrt{d\log(ds)\log(n)})}$. **check fanin issue***

The caveat is that in the same paper, they proved the same lower bound for the determinant. On the other hand, a key estimate they use (7.11.1) is close to being sharp for the determinant but conjecturally far from being sharp for the permanent.

Their method of proof is via a classical subject in algebraic geometry: the study of *Hilbert functions*, and opens the way for using techniques from commutative algebra (study of *syzygies*) in algebraic complexity theory. In §**??** I show that the shifted partial derivative technique alone is not enough for proving **VP** $\neq$ **VNP**, but I also discuss potential extensions of it that could produce stronger results.

7.5.1. Lower complexity bounds for perm_m (and $\det_n$) for depth four circuits.

Theorem 7.5.1.1. [**GKKS13a**] *Any* $\Sigma\Pi^{O(\sqrt{m})}\Sigma\Pi^{O(\sqrt{m})}$ *circuit that computes* perm_m *or* $\det_m$ *must have top fanin at least* $2^{\Omega(\sqrt{m})}$.

In other words $[\mathrm{perm}_m] \notin \sigma_s(MJ^q(\sigma_t(MJ^{m-q}(\mathbb{P}^{m^2-1}))))$, for $s = 2^{o(\sqrt{m})}$ and $q = O(\sqrt{m})$. In fact they show $[\mathrm{perm}_m] \notin \sigma_s(MJ^q(\mathbb{P}S^{m-q}\mathbb{C}^{m^2}))$.

add depth three corollary, say how solved open problem CKW10.

A basic measure of the singularity of a point z of a hypersurface $Z = \mathrm{Zeros}(P) \subset \mathbb{P}V$ is its *multiplicity*. Choose affine linear coordinates in a standard affine open subset (isomorphic to $\mathbb{C}^{\mathbf{v}}$) so that $z = (0, \ldots, 0)$. Write out the Taylor expansion of (the de-homogenized) P in the coordinates centered at z. That $z \in Z$ says the 0-th order term of the series is zero. If z is a singular point, the first order term will vanish. The multiplicity of Z at z is the lowest degree non-vanishing term in the Taylor series. (The *tangent cone* to a point on a hypersurface is the zero set of the lowest degree homogeneous term in the Taylor series (see, e.g. [**Mum95**, §5.1]).)

Recall the Jacobian varieties from §6.2.5. The dimension of $Z_{Jac,k}$ is a measure of the nature of the singularities of Z.

If $P = Q_1 \cdots Q_p$ is the product of p polynomials, and $k \leq p$, then $Z_{Jac,k}$ will be of codimension at most k because it contains $\mathrm{Zeros}(Q_{i_1}) \cap \cdots \cap \mathrm{Zeros}(Q_{i_k})$ for all $(i_1, \ldots, i_k) \subset [p]$.

Now the GKKS model is not polynomials of this form, but sums of such. With the sum of M such, we can arrive at a smooth hypersurface. So the goal is to find a pathology of $Q_1 \cdots Q_p$ that persists even when taking sums. (The goal is to find something that persists even when taking a sum of $2^{\sqrt{m}}$ such!)

Recall the method of partial derivatives/flattenings which has the desired persistence property. In this situation, the dimension of the space of partial derivatives (rank of the flattenings) is not small enough to prove the desired lower bounds. However, the image of the flattening map will be of a very pathological nature, in that all the polynomials in the image are in an ideal generated by a small number of lower degree polynomials. To see this any first derivative is in the span of $S^{q-1}V \cdot (\sum_j Q_1 \cdots \hat{Q}_j \cdots Q_p)$, where the hat denotes omission. The space of k-th derivatives is in the span of $S^{q-k}V \cdot (\sum_{|J|=k} Q_1 \cdots \hat{Q}_{j_1} \cdots \hat{Q}_{j_k} \cdots Q_p)$. In particular, it has dimension at most

$$\binom{p}{k} \dim S^{q-k}V = \binom{p}{k}\binom{\mathbf{v}+q-k-1}{q-k}. \tag{7.5.1}$$

More important than its dimension, is its structure: the ideal it generates, in a given degree D "looks like" the polynomials of degee $D-k$ times a small fixed space of dimension $\binom{p}{k}$.

This is behaviour similar to the ideals that grow the slowest, the *lex-segement ideals* (see, e.g., [**Gre98**, §3]). These are the ideals, say generated by K elements, where the generators are the first K monomials in lexographic order. For $1 \leq K \leq M$, the generators are $x_1^d, x_1^{d-1}x_2, \ldots, x_1^{d-1}x_K$. For $M+1 \leq K \leq 2M$, the generators are $x_1^{d-1}x_j, x_1^{d-2}x_2x_s$, $1 \leq j \leq M$, $2 \leq s \leq K-M$, etc... Among ideals with a fixed number of generators in a fixed degree, these ideals grow the slowest.

Theorem 7.5.1.2 (Macaulay-Gotzmann, see, e.g., [**Got78, Gre98**])**.** *Say* $\mathcal{I} \subset Sym(V)$ *is generated in degree at most* κ, *and* $\dim \mathbb{C}[X]_\kappa = \dim S^\kappa V/I_\kappa =$

Q. Write

$$Q = \binom{a_\kappa}{\kappa} + \binom{a_{\kappa-1}}{\kappa-1} + \cdots + \binom{a_\delta}{\delta} \tag{7.5.2}$$

with $a_\kappa > a_{\kappa-1} > \cdots > a_\delta$ (such an expression exists and is unique), then

$$\dim \mathbb{C}[X]_{\kappa+\tau} \leq \binom{a_\kappa+\tau}{\kappa+\tau} + \binom{a_{\kappa-1}+\tau}{\kappa+\tau-1} + \cdots + \binom{a_\delta+\tau}{\delta+\tau} \tag{7.5.3}$$

Equality is achieved for all τ if equality holds for $\tau = 1$. Equality holds for lex-segment ideals.

In contrast, the fastest possible growth of an ideal generated in degree d by $K < N$ generators is like that of a *complete intersection* *** define: In degree D they are have dimension

$$K\binom{N+(D-d)-1}{D-d} - \binom{K}{2}\binom{N+(D-2d)-1}{D-2d} + \binom{K}{3}\binom{N+(D-3d)-1}{D-3d} + \cdots$$

Fröberg [**Frö85**] conjectures such ideals exist even when $K > N$ and Iarrobino [**Iar97**] conjectures further that the ideal generated by $\ell_1^d, \ldots, \ell_K^d$ has this growth (this is known for $K \leq M$).

The study of the growth of ideals is a classical subject in algebraic geometry. The function $\mathrm{Hilb}_t(\mathcal{I}) := \dim \mathcal{I}_t$ is called the *Hilbert function* of the ideal $\mathcal{I} \subset Sym(V)$.

This suggests comparing the Hilbert functions of the ideal generated by a polynomial computable by a "small" depth four circuit, i.e, of the form $\sum_{j=1}^s Q_{1j} \cdots Q_{pj}$ and the ideal generated by the partial derivatives of the permanent, which are just the sub-permanents. Little is known about the latter, even the dimension of their zero set is not known in general. Nevertheless, we just need a lower bound on its growth, which we can obtain by degenerating it to an ideal that we can estimate.

First we get an upper bound on the growth of the Jacobian variety of $Q_1 \cdots Q_m$: By the discussion above, it has dimension at most

$$\binom{p}{k} \dim S^{q-k+\ell} V = \binom{p}{k}\binom{\mathbf{v}+\ell+q-k-1}{q-k}.$$

To get the lower bound on the growth of the ideal generated by sub-permanents we use a crude estimate: given a polynomial f given in coordinates, its *leading monomial* in some order (say lexographic), is the monomial in its expression that is highest in the order. So if an ideal is generated by $f_1, \ldots, f_q$ in degree d, then in degree $d + \ell$, it is of dimension at most the number of monomials in degree $d + \ell$ that contain a leading monomial from one of the f_j.

If we order the variables in $\mathbb{C}^{m^2}$ by $y_1^1 > y_2^1 > \cdots > y_m^1 > y_1^2 > \cdots > y_m^m$, then the leading monomial of any sub-permanent is the product of the elements on the principal diagonal. Even working with this the estimate is difficult, so in [**GKKS13a**] they restrict further to only look at leading monomials among the variables on the diagonal and super diagonal: $\{y_1^1, \ldots, y_m^m, y_2^1, y_3^2, \ldots, y_m^{m-1}\}$. Among these, they compute that the number of leading monomials of degree δ is $\binom{2m-\delta}{\delta}$. In our case, $\delta = m-k$ and $D = \ell + m - k$. Let $I_d^{\mathrm{perm}_m,k} \subset S^d\mathbb{C}^{m^2}$ denote the degree d component of the ideal generated by the order k partial derivatives of the permanent. We have

$$\dim I_{n-k+\ell}^{\mathrm{perm}_m,k} \geq \binom{m+k}{2k}\binom{m^2+\ell-2k}{\ell}. \tag{7.5.4}$$

Putting the estimates together, if we want to realize the permanent by a depth four circuit as above, for some s, we need

$$s \geq \frac{\binom{m+k}{2k}\binom{m^2+\ell-2k}{\ell}}{\binom{c\sqrt{m}+k}{k}\binom{m^2+\ell+(\sqrt{m}-1)k}{m^2}} \tag{7.5.5}$$

They obtain their result by setting $\ell = m^{\frac{5}{2}}$ and $k = \epsilon m^{\frac{1}{2}}$ where ϵ is a constant defined below. To see this, one calculates (using the estimates of §7.1.1):

$$\ln \frac{\binom{m^2+m^{\frac{5}{2}}-2\epsilon\sqrt{m}}{m^{\frac{5}{2}}}}{\binom{m^2+m^{\frac{5}{2}}+(\sqrt{m}-1)\epsilon\sqrt{m}}{m^2}} = -2\epsilon\sqrt{m}\ln\sqrt{m} - \epsilon\sqrt{m} \pm O(1)$$

$$\ln \frac{\binom{m^2+\epsilon\sqrt{m}}{2\epsilon\sqrt{m}}}{\binom{(c+\epsilon)\sqrt{m}}{\epsilon\sqrt{m}}} = \sqrt{m}[2\epsilon\ln\frac{\sqrt{m}}{2\epsilon} + 2\epsilon$$
$$+ (c+\epsilon)[\frac{\epsilon}{c+\epsilon}\ln(\frac{\epsilon}{c+\epsilon}) + (1-\frac{\epsilon}{c+\epsilon})\ln(1-\frac{\epsilon}{c+\epsilon})] + O(\ln m)$$

Putting these together, we get

$$\ln(s) \geq \epsilon\sqrt{m}\ln\frac{1}{4\epsilon(c+\epsilon)} \pm O(1)$$

so if we choose ϵ such that $\frac{1}{4\epsilon(c+\epsilon)} = e$, we get $\ln(s) \geq \Omega(\sqrt{m})$.

add precise numerical result and compare with Guan

7.6. Shifted partial derivatives cannot separate permanent from determinant

7.6.1. Statement of the result. We prove the method of shifted partial derivatives cannot give better than a quadratic separation of the permanent from the determinant:

Theorem 7.6.1.1. [**ELSW16**] *There exists a constant M such that for all $m > M$ and every $n > 2m^2 + 2m$, any τ and any $k < n$,*

$$\operatorname{rank}((\ell^{n-m} \operatorname{perm}_m)_{(k,n-k)[\tau]}) < \operatorname{rank}((\det_n)_{(k,n-k)[\tau]}).$$

7.6.2. Overview of the proof. The proof of Theorem 7.6.1.1 splits into four cases:

- (C1) Case $k \geq n - \frac{n}{m+1}$,
- (C2) Case $2m \leq k \leq n - 2m$,
- (C3) Case $k < 2m$ and $\tau > \frac{3}{2}n^2 m$,
- (C4) Case $k < 2m$ and $\tau < \frac{n^3}{6m}$.

Note that C1,C2 overlap when $n > 2m^2 + 2m$ and C3,C4 overlap when $n > \frac{m^2}{4}$, so it suffices to take $n > 2m^2 + 2m$.

In the first case, the proof has nothing to do with the padded permanent or its derivatives: it is valid for any polynomial in $m^2 + 1$ variables. Cases C2,C3 only use that we have a padded polynomial. In the case C4, the only property of the permanent that is used is an estimate on the size of the space of its partial derivatives. Case C1 is proved by showing that in this range the partials of the determinant can be degenerated into the space of *all* polynomials of degree $n - k$ in $m^2 + 1$ variables. Cases C2,C3 use that when $k < n - m$, the Jacobian ideal of *any* padded polynomial $\ell^{n-m} P \in S^n W$ is contained in the ideal generated in degree $n - m - k$ by ℓ^{n-m-k}, which has slowest possible growth by Macaulay's theorem as explained below. Case C2 compares that ideal with the Jacobian ideal of the determinant; it is smaller in degree $n - k$ and therefore smaller in all higher degrees by Macaulay's theorem. Case C3 compares that ideal with an ideal with just two generators in degree $n - k$. Case C4 uses a lower bound for the determinant used in [**GKKS13a**] and compares it with a very crude upper bound for the dimension of the space of shifted partial derivatives for the permanent.

7.7. Macaulay's Theorem

We only use Corollary 7.7.0.4 from this section in the proof of Theorem 7.6.1.1.

Theorem 7.7.0.1 (Macaulay, see, e.g., [**Gre98**]). *Let $\mathcal{I} \subset Sym(\mathbb{C}^N)$ be a homogeneous ideal, and let d be a natural number. Write*

$$\dim S^d\mathbb{C}^N/\mathcal{I}_d = \binom{a_d}{d} + \binom{a_{d-1}}{d-1} + \cdots + \binom{a_\delta}{\delta} \tag{7.7.1}$$

with $a_d > a_{d-1} > \cdots > a_\delta$ (such an expression exists and is unique). Then

$$\dim \mathcal{I}_{d+\tau} \geq \binom{N+d+\tau-1}{d+\tau} - \left[\binom{a_d+\tau}{d+\tau} + \binom{a_{d-1}+\tau}{d+\tau-1} + \cdots + \binom{a_\delta+\tau}{\delta+\tau}\right]. \tag{7.7.2}$$

Remark 7.7.0.2. Gotzman [**Got78**] showed that if $\mathcal{I}$ is generated in degree at most d, then equality is achieved for all τ in (7.7.2) if equality holds for $\tau = 1$. Ideals satisfying this minimal growth exist, for example, *lex-segment ideals* satisfy this property, see [**Gre98**].

Remark 7.7.0.3. Usually Macaulay's theorem is stated in terms of the coordinate ring $\mathbb{C}[X] := Sym(W)/\mathcal{I}$ of the variety (scheme) $X \subset W^*$ that is the zero set of $\mathcal{I}$, namely

$$\dim \mathbb{C}[X]_{d+\tau} \leq \binom{a_d+\tau}{d+\tau} + \binom{a_{d-1}+\tau}{d+\tau-1} + \cdots + \binom{a_\delta+\tau}{\delta+\tau}.$$

Corollary 7.7.0.4. *Let $\mathcal{I}$ be an ideal such that $\dim \mathcal{I}_d \geq \dim S^{d-q}\mathbb{C}^N = \binom{N+d-q-1}{d-q}$ for some $q < d$. Then $\dim \mathcal{I}_{d+\tau} \geq \dim S^{d-q+\tau}\mathbb{C}^N = \binom{N+\tau+d-q-1}{\tau+d-q}$.*

Proof. First use the identity

$$\binom{a+b}{b} = \sum_{j=1}^{q} \binom{a+b-j}{b-j+1} + \binom{a+b-q}{b-q} \tag{7.7.3}$$

with $a = N-1$, $b = d$. Write this as

$$\binom{N-1+d}{d} = Q_d + \binom{N-1+d-q}{d-q}.$$

Set

$$Q_{d+\tau} := \sum_{j=1}^{q} \binom{N-1+d+\tau-j}{d+\tau-j+1}.$$

By Macaulay's theorem, any ideal $\mathcal{I}$ with

$$\dim \mathcal{I}_d \geq \binom{N-1+d-q}{d-q}$$

must satisfy

$$\dim \mathcal{I}_{d+\tau} \geq \binom{N-1+d+\tau}{d+\tau} - Q_{d+\tau} = \binom{N-1+d-q+\tau}{d-q+\tau}.$$

□

We will use Corollary 7.7.0.4 with $N = n^2$, $d = n-k$, and $d - q = m$.

7.8. Case C1

Our assumption is $(m+1)(n-k)\leq n$. It will be sufficient to show that some $R \in \mathrm{End}(W)\cdot \det_n$ satisfies $\mathrm{rank}((\ell^{n-m}\,\mathrm{perm}_m)_{(k,n-k)[\tau]}) < \mathrm{rank}(R_{k,n-k[\tau]})$. Block the matrix $x = (x^s_u) \in \mathbb{C}^{n^2}$, with $1 \leq s,u \leq n$, as a union of $n-k$ blocks of size $m \times m$ in the upper-left corner plus the remainder, which by our assumption includes at least $n-k$ elements on the diagonal. Set each diagonal block to the matrix (y^i_j), with $1 \leq i,j \leq n$, (there are $n-k$ such blocks), fill the remainder of the diagonal with ℓ (there are at least $n-k$ such terms), and fill the remainder of the matrix with zeros. Let R be the restriction of the determinant to this subspace. Then the space of partials of R of degree $n-k$, $R_{k,n-k}(S^k\mathbb{C}^{n^2*}) \subset S^{n-k}\mathbb{C}^{n^2}$ contains a space isomorphic to $S^{n-k}\mathbb{C}^{m^2+1}$, and $\mathcal{I}^{\ell^{n-m}\,\mathrm{perm}_m,k}_{n-k} \subset S^{n-k}\mathbb{C}^{m^2+1}$ so we conclude.

Example 7.8.0.1. Let $m=2$, $n=6$, $k=4$. The matrix is

$$\begin{pmatrix} y^1_1 & y^1_2 & & & & \\ y^2_1 & y^2_2 & & & & \\ & & y^1_1 & y^1_2 & & \\ & & y^2_1 & y^2_2 & & \\ & & & & \ell & \\ & & & & & \ell \end{pmatrix}.$$

The polynomial $(y^1_1)^2$ is the image of $\frac{\partial^4}{\partial x^2_2 \partial x^4_4 \partial x^5_5 \partial x^6_6}$ and the polynomial $y^1_2 y^2_2$ is the image of $\frac{\partial^4}{\partial x^2_1 \partial x^3_3 \partial x^5_5 \partial x^6_6}$.

7.9. Case C2

As long as $k < n-m$, $\mathcal{I}^{\ell^{n-m}\,\mathrm{perm}_m,k}_{n-k} \subset \ell^{n-m-k}\cdot S^mW$, so

$$\dim \mathcal{I}^{\ell^{n-m}\,\mathrm{perm}_m,k}_{n-k+\tau} \leq \binom{n^2+m+\tau-1}{m+\tau}. \tag{7.9.1}$$

By Corollary 7.7.0.4, it will be sufficient to show that

$$\dim \mathcal{I}^{\det_n,k}_{n-k} = \binom{n}{k}^2 \geq \dim S^mW = \binom{n^2+m-1}{m}. \tag{7.9.2}$$

In the range $2m \leq k \leq n-2m$, the quantity $\binom{n}{k}$ is minimized at $k=2m$ and $k=n-2m$, so it is enough to show that

$$\binom{n}{2m}^2 \geq \binom{n^2+m-1}{m}. \tag{7.9.3}$$

Using (**??**)

$$\begin{aligned}
\ln \binom{n}{2m}^2 &= 2[n\ln(n) - 2m\ln(2m) - (n-2m)\ln(n-2m)] - \Theta(\ln(n)) \\
&= 2[n\ln(\frac{n}{n-2m}) + 2m\ln(\frac{n-2m}{2m})] - \Theta(\ln(n)) \\
&\leq 4m + m\ln[(\frac{n}{2m}-1)^4] - \Theta(\ln(n)),
\end{aligned}$$

where to obtain the last line we used $(1-\frac{2m}{n})^n > e^{-2m}e^{\Theta(\frac{m^2}{n})}$, and

$$\begin{aligned}
\ln \binom{n^2+m-1}{m} &= (n^2+m-1)\ln(n^2+m-1) - m\ln(m) - (n^2-1)\ln(n^2-1) - \Theta(\ln(n)) \\
&= (n^2-1)\ln(\frac{n^2+m-1}{n^2-1}) + m\ln(\frac{n^2+m-1}{m}) - \Theta(\ln(n)) \\
&= m\ln(\frac{n^2}{m} - \frac{m-1}{m}) + m - \Theta(\ln(n)).
\end{aligned}$$

So (7.9.3) will hold when $(\frac{n}{2m}-1)^4 > (\frac{n^2}{m} - \frac{m-1}{m})$ which holds for all sufficiently large m when $n > m^2$.

7.10. Case C3

Here we simply degenerate $\det_n$ to $R = \ell_1^n + \ell_2^n$ by e.g., setting all diagonal elements to ℓ_1, all the sub-diagonal elements to ℓ_2 as well as the $(1,n)$-entry, and setting all other elements of the matrix to zero. Then $\mathcal{I}^{R,k}_{n-k} = \operatorname{span}\{\ell_1^{n-k}, \ell_2^{n-k}\}$. In degree $n-k+\tau$, this ideal consists of all polynomials of the form $\ell_1^{n-k}Q_1 + \ell_2^{n-k}Q_2$ with $Q_1, Q_2 \in S^\tau\mathbb{C}^{n^2}$, which has dimension $2\dim S^\tau\mathbb{C}^{n^2} - \dim S^{\tau-(n-k)}\mathbb{C}^{n^2}$ because the polynomials of the form $\ell_1^{n-k}\ell_2^{n-k}Q_3$ with $Q_3 \in S^{\tau-(n-k)}\mathbb{C}^{n^2}$ appear in both terms. By this discussion, or simply because this is a complete intersection ideal, we have

$$\dim \mathcal{I}^{R,k}_{n-k+\tau} = 2\binom{n^2+\tau-1}{\tau} - \binom{n^2+\tau-(n-k)-1}{\tau-(n-k)}. \tag{7.10.1}$$

We again use the estimate (7.9.1) from Case C2, so we need to show

$$2\binom{n^2+\tau-1}{\tau} - \binom{n^2+\tau+m-1}{\tau+m} - \binom{n^2+\tau-(n-k)-1}{\tau-(n-k)} > 0.$$

Divide by $\binom{n^2+\tau-1}{\tau}$. We need

$$2 > \Pi_{j=1}^{m} \frac{n^2+\tau+m-j}{\tau+m-j} + \Pi_{j=1}^{n-k} \frac{\tau-j}{n^2+\tau-j} \tag{7.10.2}$$

$$= \Pi_{j=1}^{m}(1+\frac{n^2}{\tau+m-j}) + \Pi_{j=1}^{n-k}(1-\frac{n^2}{n^2+\tau-j}) \tag{7.10.3}$$

The second line is less than

$$(1+\frac{n^2}{\tau})^m + (1-\frac{n^2}{n^2+\tau-1})^{n-k}. \tag{7.10.4}$$

We analyze (7.10.4) as a function of τ. Write $\tau = n^2 m\delta$, for some constant δ. Then (7.10.4) is bounded above by

$$e^{\frac{1}{\delta}} + e^{\frac{2}{\delta}-\frac{n}{m\delta}}.$$

The second term goes to zero for large m, so we just need the first term to be less than 2, so we take, e.g. $\delta = \frac{3}{2}$.

7.11. Case C4

We use a lower bound on $\mathcal{I}_{n-k+\tau}^{\det_n,k}$ from [**GKKS13a**]: Given a polynomial f given in coordinates, its *leading monomial* in some monomial order, is the monomial in its expression that is highest in the order. If an ideal is generated by $f_1, \ldots, f_q$ in degree $n-k$, then in degree $n-k+\tau$, its dimension is at least the number of monomials in degree $n-k+\tau$ that contain a leading monomial from one of the f_j.

If we order the variables in $\mathbb{C}^{n^2}$ by $x_1^1 > x_2^1 > \cdots > x_n^1 > x_1^2 > \cdots > x_n^n$, then the leading monomial of any minor is the product of the elements on the principal diagonal. Even estimating just these monomials is difficult, so in [**GKKS13a**] they restrict further to only look at leading monomials of size $(n-k)$ minors among the variables on the diagonal and super diagonal: $\{x_1^1, \ldots, x_n^n, x_2^1, x_3^2, \ldots, x_n^{n-1}\}$. Among these, they compute that the number of leading monomials of degree $n-k$ is $\binom{n+k}{2k}$. Then then show that in degree $n-k+\tau$ the dimension of this ideal is bounded below by $\binom{n+k}{2k}\binom{n^2+\tau-2k}{\tau}$ so we conclude

$$\dim \mathcal{I}_{n-k+\tau}^{\det_n,k} \geq \binom{n+k}{2k}\binom{n^2+\tau-2k}{\tau}. \tag{7.11.1}$$

We compare this with the very crude estimate

$$\dim \mathcal{I}_{n-k+\tau}^{\ell^{n-m}\operatorname{perm}_m,k} \leq \sum_{j=0}^{k} \binom{m}{j}^2 \binom{n^2+\tau-1}{\tau},$$

where $\sum_{j=0}^{k}\binom{m}{j}^2$ is the dimension of the space of partials of order k of $\ell^{n-m}\,\mathrm{perm}_m$, and the $\binom{n^2+\tau-1}{\tau}$ is what one would have if there were no syzygies (relations among the products).

We have

(7.11.2)
$$\begin{aligned}\ln\binom{n+k}{2k} &= n\ln\frac{n+k}{n-k}+k\ln\frac{n^2-k^2}{4k^2}+\Theta(\ln(n))\\ &= k\ln\frac{n^2-k^2}{4k^2}+\Theta(\ln(n))\end{aligned}$$

(7.11.3)
$$\begin{aligned}\ln\frac{\binom{n^2+\tau-2k}{\tau}}{\binom{n^2+\tau-1}{\tau}} &= n^2\ln\frac{(n^2+\tau-2k)(n^2-1)}{(n^2-2k)(n^2+\tau-1)}+\tau\ln\frac{n^2+\tau-2k}{n^2+\tau-1}+2k\ln\frac{n^2-2k}{n^2+\tau-2k}+\Theta(\ln(n))\\ &= -2k\ln(\frac{\tau}{n^2}+1)+\Theta(\ln(n)),\end{aligned}$$

where the second lines of expressions (7.11.2),(7.11.3) hold because $k<2m$. We split this into two sub-cases: $k\geq\frac{m}{2}$ and $k<\frac{m}{2}$.

7.11.1. Subcase $k\geq\frac{m}{2}$. In this case we have $\sum_{j=0}^{k}\binom{m}{j}^2<\binom{2m}{m}$. We show the ratio

$$\frac{\binom{n+k}{2k}\binom{n^2+\tau-2k}{\tau}}{\binom{2m}{m}\binom{n^2+\tau-1}{\tau}} \tag{7.11.4}$$

is greater than one. Now

$$\begin{aligned}\ln\binom{2m}{m} &= m\ln 4+\Theta(\ln(m)).\\ &= k\ln(4^{\frac{m}{k}})+\Theta(\ln(m)).\end{aligned} \tag{7.11.5}$$

If

$$k\ln\left(\frac{n^2-k^2}{4k^2}\frac{1}{(\frac{\tau}{n^2}+1)^2}\frac{1}{4^{\frac{m}{k}}}\right)\pm\Theta(\ln(n))$$

is positive, then (7.11.4) is greater than one. This will occur if

$$\frac{n^2-k^2}{4k^2}\frac{1}{(\frac{\tau}{n^2}+1)^2}\frac{1}{4^{\frac{m}{k}}}>1$$

i.e., if

$$\tau<n^2(\frac{\sqrt{n^2-k^2}}{2k4^{\frac{m}{2k}}}-1).$$

Write this as

$$\tau<n^2(\frac{n}{2\epsilon m4^{\frac{1}{2\epsilon}}}-1). \tag{7.11.6}$$

The worst case is $\epsilon = 2$ where it suffices to take $\tau < \frac{n^3}{6m}$.

7.11.2. Subcase $k < \frac{m}{2}$. Here we use that $\sum_{j=0}^{k} \binom{m}{j}^2 < k\binom{m}{k}^2$ and a similar argument gives that it suffices to have

$$\tau < n^2\left(\frac{\sqrt{n^2-k^2}}{2k}\frac{1}{\sqrt{\frac{m}{k}}-1} - 1\right).$$

The smallest upper bound for τ occurs when $k = \frac{m}{2}$, where the estimate easily holds when $\tau < \frac{n^3}{6m}$.

7.12. Polynomial identity testing, hitting sets and explicit Noether normalization

This section to be written*

7.13. Raz's theorem on tensor rank and arithmetic formulas

This section to be written*

Chapter 8

Representation theory and its uses in complexity theory

In this chapter I derive the representation theory of the general linear group and give numerous applications to complexity theory. In order to get to the applications as soon as possible, I summarize basic facts about representations of the general linear group $GL(V)$ in §8.1. The first application, in §8.2, explains the theory of *Young flattenings* underlying the equations that led to the $\mathbf{2n^2 - n}$ lower bound for the border rank of matrix multiplication (Theorem 2.6.3.6). I also explain how the method of shifted partial derivatives may be viewed as a special case of Young flattenings. Next, in §8.3, I briefly discuss how representation theory has been used to find equations for secant varieties of Segre varieties (and other varieties). In §8.4, I describe severe restrictions on (modules of) polynomials to be useful for the permanent v. determinant problem. In §8.5, I give the proofs of several statements about $\mathcal{D}et_n$ from Chapter 7. In §8.6, I begin to develop representation theory. There are several paths to obtaining the representation theory of the general linear group. I use the path via the double commutant theorem, the algebraic Peter-Weyl theorem and Schur-Weyl duality. The reason for this choice is that the (finite) Peter-Weyl theorem is the starting point of the Cohn-Umans program of §3.5 and the algebraic Peter-Weyl theorem was the starting point of the program of [**MS01, MS08**] described in §8.8. The representations of the general linear group are then derived in §8.7. In §8.8 I begin a discussion of the program of [**MS01, MS08**], as refined in [**BLMW11**], to separate the permanent from the determinant via representation theory. This is continued in §8.9, where detailed information about the coordinate ring of the orbit is given, §8.10, which contains a general discussion of plethysm coefficients, and §8.11, which presents results of [**IP15**] and [**BIP16**] that show this program cannot work as stated. I then outline the proof of Theorems 7.4.7.1 and 7.4.7.6 giving the exponential separation of the permanent from the determinant in the restricted model of equivariant determinantal expressions. I conclude, in §8.13 with a description of the symmetry groups of other polynomials, which will be useful for future work.

8.1. Representation theory of the general linear group

Irreducible representations of $GL(V)$ in $V^{\otimes d}$ are indexed by partitions of d with length at most $\mathbf{v}$, as we will prove in Theorem 8.7.1.2. Let $S_\pi V$ denote the isomorphism class of the irreducible representation associated to the partition π, and let $S_{\overline{\pi}}V$ denote some particular realization of $S_\pi V$. For a partition $\pi = (p_1, \dots, p_k)$, write $|\pi| = p_1 + \cdots + p_k$ and $\ell(\pi) = k$. If a number is repeated I sometimes use superscripts to record its multiplicity, for example $(2,2,1,1,1) = (2^2, 1^3)$.

To visualize π, define a *Young diagram* associated to a partition π to be a collection of left-aligned boxes with p_j boxes in the the j-th row, as in Figure 8.1.1.

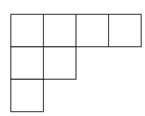


Figure 8.1.1. Young diagram for $\pi = (4,2,1)$

Define the *conjugate partition* π' to π to be the partition whose Young diagram is the reflection of the Young diagram of π in the north-west to south-east diagonal.

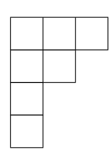


Figure 8.1.2. Young diagram for $\pi' = (3,2,1,1)$, the conjugate partition to $\pi = (4,2,1)$.

8.1.1. Lie algebras. Associated to any Lie group G is a *Lie algebra* $\mathfrak{g}$, which is a vector space that may be identified with $T_{\mathrm{Id}}G$. For basic information on Lie algebras associated to a Lie group, see any of [**Spi79, IL03, Pro07**].

When $G = GL(V)$, then $\mathfrak{g} = \mathfrak{gl}(V) := V^*{\otimes}V$. If $G \subseteq GL(V)$, so that G acts on $V^{\otimes d}$, there is an induced action of $\mathfrak{g} \subseteq \mathfrak{gl}(V)$ given by, for $X \in \mathfrak{g}$,

$$\begin{aligned}&X.(v_1{\otimes}v_2{\otimes}\cdots\otimes v_d)\\ &\quad= (X.v_1){\otimes}v_2{\otimes}\cdots\otimes v_d + v_1{\otimes}(X.v_2){\otimes}\cdots\otimes v_d + \cdots + v_1{\otimes}v_2{\otimes}\cdots\otimes v_{d-1}{\otimes}(X.v_d).\end{aligned}$$

To see why this is a natural induced action, consider a curve $g(t) \subset G$ with $g(0) = \mathrm{Id}$ and $X = g'(0)$ and take

$$\frac{d}{dt}\,|_{t=0}\; g\cdot(v_1{\otimes}\cdots\otimes v_d) = \frac{d}{dt}\,|_{t=0}\;(g\cdot v_1){\otimes}\cdots\otimes(g\cdot v_d).$$

One concludes by applying the Leibniz rule.

8.1.2. Weights. Fix a basis $e_1, \ldots, e_{\mathbf{v}}$ of V, let $T \subset GL(V)$ denote the subgroup of diagonal matrices, called a *maximal torus*, let $B \subset GL(V)$ be the subgroup of upper triangular matrices, called a *Borel subgroup*, and let $N \subset B$ be the upper triangular matrices with 1's along the diagonal. The Lie algebra $\mathfrak{n}$ of N consists of nilpotent matrices. Call $z \in V^{\otimes d}$ a *weight vector* if $T[z] = [z]$. If

$$\begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_{\mathbf{v}}\end{pmatrix} z = (x_1)^{p_1}\cdots(x_{\mathbf{v}})^{p_{\mathbf{v}}} z$$

we say z has *weight* $(p_1, \ldots, p_{\mathbf{v}}) \in \mathbb{Z}^{\mathbf{v}}$.

Call z a *highest weight vector* if $B[z] = [z]$, i.e., if $Nz = z$. If M is an irreducible $GL(V)$-module and $z \in M$ is a highest weight vector, call the weight of z the *highest weight* of M. A necessary condition for two irreducible $GL(V)$-modules to be isomorphic is that they have the same highest weight (because they must also be isomorphic T-modules). The condition is also sufficient, see §8.7.

Exercise 8.1.2.1: (1) Show that z is a highest weight vector if and only if $\mathfrak{n}.z = 0$.

The elements of $\mathfrak{n}$ are often called *raising operators*.

Exercise 8.1.2.2: (1) Show that if $z \in V^{\otimes d}$ is a highest weight vector of weight $(p_1, \ldots, p_{\mathbf{v}})$, then $(p_1, \ldots, p_{\mathbf{v}})$ is a partition of d. ⊚

When $G = GL(A_1) \times \cdots \times GL(A_n)$, the maximal torus in G is the product of the maximal tori in the $GL(A_j)$, and similarly for the Borel. A weight is then defined to be an n-tuple of weights etc...

Because of the relation with weights, it will often be convenient to add a string of zeros to a partition to make it a string of $\mathbf{v}$ integers.

Exercise 8.1.2.3: (1) Show that the space $S^2(S^2\mathbb{C}^2)$ contains a copy of $S_{22}\mathbb{C}^2$ by showing that $(x_1^2)(x_2^2) - (x_1x_2)(x_1x_2) \in S^2(S^2\mathbb{C}^2)$ is a highest weight vector.

Exercise 8.1.2.4: (1!) Find highest weight vectors in $V, S^2V, \Lambda^2V, S^3V, \Lambda^3V$ and the kernels of the symmetrization and skew-symmetrization maps $V\otimes S^2V \to S^3V$ and $V\otimes\Lambda^2V \to \Lambda^3V$. Note that both of the last two modules have highest weight $(2,1)$, i.e., they are realizations of $S_{21}V$.

Exercise 8.1.2.5: (2) More generally, find a highest weight vector for the kernel of the symmetrization map $V\otimes S^{d-1}V \to S^dV$ and of the kernel of the "exterior derivative" (or "Koszul") map

$$S^kV\otimes\Lambda^tV \to S^{k-1}V\otimes\Lambda^{t+1}V \tag{8.1.1}$$

$$x_1\cdots x_k\otimes y_1\wedge\cdots\wedge y_t \mapsto \sum_{j=1}^{k} x_1\cdots\hat{x}_j\cdots x_k\otimes x_j\wedge y_1\wedge\cdots\wedge y_t.$$

Exercise 8.1.2.6: (1!) Let $\pi = (p_1,\dots,p_\ell)$ be a partition with at most $\mathbf{v}$ parts and let $\pi' = (q_1,\dots,q_{p_1})$ denote the conjugate partition. Show that

$$z_\pi := (e_1\wedge\cdots\wedge e_{q_1})\otimes(e_1\wedge\cdots\wedge e_{q_2})\otimes\cdots\otimes(e_1\wedge\cdots\wedge e_{q_{p_1}}) \in V^{\otimes|\pi|} \tag{8.1.2}$$

is a highest weight vector of weight π.

Exercise 8.1.2.7: (2) Show that a basis of the highest weight space of $[2,1]\otimes S_{21}V \subset V^{\otimes 3}$ is $v_1 = e_1\wedge e_2\otimes e_1$ and $v_2 = e_1\otimes e_1\wedge e_2$. Let $\mathbb{Z}_3 \subset \mathfrak{S}_3$ be the cylic permutation of the three factors in $V^{\otimes 3}$ and show that $\omega v_1 \pm \omega^2 v_2$ are eigenvectors for this action with eigenvalues ω, ω^2, where $\omega = e^{\frac{2\pi i}{3}}$.

The Lie algebra of $SL(V)$, denoted $\mathfrak{sl}(V)$, is the set of traceless endomorphisms. One can defined weights for the Lie algebra of the torus, which are essentially the logs of the corresponding torus in the group. In particular, vectors of $\mathfrak{sl}$-weight zero have $GL(V)$-weight $(d,\dots,d) = (d^{\mathbf{v}})$ for some d.

8.1.3. The Pieri rule.

I describe the decomposition of $S_\pi V\otimes V$ as a $GL(V)$-module. Write $\pi' = (q_1,\dots,q_{p_1})$ and recall z_π from (8.1.2). Consider the vectors:

$$(e_1\wedge\cdots\wedge e_{q_1}\wedge e_{q_1+1})\otimes(e_1\wedge\cdots\wedge e_{q_2})\otimes\cdots\otimes(e_1\wedge\cdots\wedge e_{q_{p_1}})$$

$$\vdots$$

$$(e_1\wedge\cdots\wedge e_{q_1})\otimes(e_1\wedge\cdots\wedge e_{q_2})\otimes\cdots\otimes(e_1\wedge\cdots\wedge e_{q_{p_1}}\wedge e_{q_{p_1}+1})$$

$$(e_1\wedge\cdots\wedge e_{q_1})\otimes(e_1\wedge\cdots\wedge e_{q_2})\otimes\cdots\otimes(e_1\wedge\cdots\wedge e_{q_{p_1}})\otimes e_1.$$

These are all highest weight vectors obtained by tensoring z_π with a vector in V and skew-symmetrizing appropriately, so the associated modules are contained in $S_\pi V \otimes V$. With a little more work, one can show these are highest weight vectors of all the modules that occur in $S_\pi V \otimes V$. If $q_j = q_{j+1}$ one gets the same module if one inserts e_{q_j+1} into either slot, but its multiplicity in $S_\pi V \otimes V$ is one. More generally one obtains:

Theorem 8.1.3.1 (The Pieri formula)**.** *The decomposition of $S_\pi V \otimes S^d V$ is multiplicity free. The partitions corresponding to modules $S_\mu V$ that occur are those obtained from the Young diagram of π by adding d boxes to the diagram of π, with no two boxes added to the same column.*

Definition 8.1.3.2. Let π, μ be partitions with $\ell(\mu) < \ell(\pi)$ One says μ *interlaces* π if $p_1 \geq m_1 \geq p_2 \geq m_2 \geq \cdots \geq m_{\ell(\pi)-1} \geq p_{\ell(\pi)}$.

Exercise 8.1.3.3: (1) Show that $S_\pi V \otimes S_{(d)} V$ consists of all the $S_\mu V$ such that $|\mu| = |\pi| + d$ and π interlaces μ.

Exercise 8.1.3.4: (1) Show that a necessary condition for $S_\pi V$ to appear in $S^d(S^n V)$ is that $\ell(\pi) \leq d$.

Although a pictorial proof is possible, the standard proof of the Pieri formula uses a *character* (see §8.6.7) calculation, computing $\chi_\pi \chi_{(d)}$ as a sum of χ_μ's. See, e.g., [**Mac95**, §I.9]. A different proof, using Schur-Weyl duality is in [**GW09**, §9.2]. There is an algorithm to compute arbitrary tensor product decompositions called the *Littlewood Richardson Rule*. See, e.g., [**Mac95**, §I.9] for details.

Similar considerations give:

Theorem 8.1.3.5. *[The skew-Pieri formula] The decomposition of $S_\pi V \otimes \Lambda^k V$ is multiplicity free. The partitions corresponding to modules $S_\mu V$ that occur are those obtained from the Young diagram of π by adding k boxes to the diagram of π, with no two boxes added to the same row.*

8.1.4. The $GL(V)$-modules not appearing in the tensor algebra of V. The $GL(V)$-module V^* does not appear in the tensor algebra of V. Nor do the one-dimensional representations $\det^{-k} : GL(V) \to GL(\mathbb{C}^1)$ given by, for $v \in \mathbb{C}^1$, $\det^{-k}(g)v := \det(g)^{-k} v$.

Exercise 8.1.4.1: (1) Show that if $\pi = (p_1, \ldots, p_{\mathbf{v}})$ with $p_{\mathbf{v}} > 0$, then $\det^{-1} \otimes S_\pi V = S_{(p_1 - 1, \ldots, p_{\mathbf{v}} - 1)} V$. ⊚

Exercise 8.1.4.2: (1) Show that as a $GL(V)$-module, $V^* = \Lambda^{\mathbf{v}-1} V \otimes \det^{-1} = S_{1^{\mathbf{v}-1}} V \otimes \det^{-1}$. ⊚

Every irreducible $GL(V)$-module is of the form $S_\pi V \otimes \det^{-k}$ for some $k \geq 0$. Thus they may be indexed by non-increasing sequences of integers

$(p_1,\dots,p_{\mathbf{v}})$ where $p_1\geq p_2\geq\cdots\geq p_{\mathbf{v}}$. Such a module is isomorphic to $S_{(p_1-p_{\mathbf{v}},\dots,p_{\mathbf{v}-1}-p_{\mathbf{v}},0)}V{\otimes}\det^{p_{\mathbf{v}}}$.

Using

$$S_\pi V{\otimes}V^*=S_\pi V{\otimes}\Lambda^{\mathbf{v}-1}V{\otimes}\det{}^{-1},$$

we may compute the decomposition of $S_\pi V{\otimes}V^*$ using the skew-symmetric version of the Pieri rule.

Example 8.1.4.3. Let $\mathbf{w}=3$, then

$$\begin{aligned}S_{(32)}W{\otimes}W^* &= S_{(43)}W{\otimes}\det{}^{-1}\oplus S_{(331)}W{\otimes}\det{}^{-1}\oplus S_{(421)}W{\otimes}\det{}^{-1}\\ &= S_{(43)}W{\otimes}\det{}^{-1}\oplus S_{(22)}W\oplus S_{(31)}W.\end{aligned}$$

The first module does not occur in the tensor algebra but the rest do.

8.1.5. $SL(V)$**-modules in** $V^{\otimes d}$**.** Every $SL(V)$-module is the restriction to $SL(V)$ of some $GL(V)$-module. However distinct $GL(V)$-modules, when restricted to $SL(V)$ can become isomorphic, such as the trivial representation and $\Lambda^{\mathbf{v}}V$.

Proposition 8.1.5.1. *Let $\pi=(p_1,\dots,p_{\mathbf{v}})$ be a partition. The $SL(V)$-modules in the tensor algebra $V^{\otimes}$ that are isomorphic to $S_\pi V$ are $S_\mu V$ with $\mu=(p_1+j,p_2+j,\dots,p_{\mathbf{v}}+j)$ for $-p_{\mathbf{v}}\leq j<\infty$.*

Exercise 8.1.5.2: (2) Prove Proposition 8.1.5.1. ⊚

For example, for SL_2-modules, $S_{p_1,p_2}\mathbb{C}^2\simeq S^{p_1-p_2}\mathbb{C}^2$. We conclude:

Corollary 8.1.5.3. *A complete set of the finite dimensional irreducible representations of SL_2 are the $S^d\mathbb{C}^2$ with $d\geq 0$.*

The $GL(V)$-modules that are $SL(V)$-equivalent to $S_\pi V$ may be visualized as being obtained by erasing or adding columns of size $\mathbf{v}$ from the Young diagram of π, as in Figure 8.1.5.

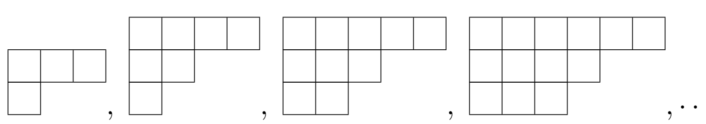


Figure 8.1.3. Young diagrams for SL_3-modules equivalent to $S_{421}\mathbb{C}^3$

Exercise 8.1.5.4: (1!) Let $T^{SL}\subset SL(V)$ be the diagonal matrices with determinant one. Show that $(V^{\otimes d})^{T^{SL}}$ is zero unless $d=\delta\mathbf{v}$ for some natural number δ and in this case it consists of all vectors of weight $(\delta^{\mathbf{v}})$.

8.2. Young flattenings

Most known equations for border rank of tensors, i.e., polynomials in the ideal of the variety $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ and symmetric border rank of polynomials, i.e., polynomials in the ideal of the variety $\sigma_r(v_d(\mathbb{P}V))$, are obtained by taking minors of some auxiliary matrix constructed from the tensor (polynomial). What follows is a general way to use representation theory to find such matrices.

8.2.1. The case of polynomials. Let $P \in S^dV$. Recall the flattenings from §6.2: $P_{k,d-k} : S^kV^* \to S^{d-k}V$. We may think of this as a consequence of the fact that $S^dV \subset S^kV \otimes S^{d-k}V$. The generalization is similar: we want to find linear inclusions $S^dV \subset U \otimes W$, i.e., as a space of linear maps from U^* to W. If the rank of the linear map associated to ℓ^d is r_0 and the rank of the linear map associated to P is r, then $\underline{\mathbf{R}}_S(P) \geq \frac{r}{r_0}$.

Exercise 8.2.1.1: (1!) Prove the last assertion. ⊚

This method works best when r_0 is small. For example in the classical flattening case $r_0 = 1$.

Representation theory comes in because we will take U, W to be $GL(V)$-modules and the linear inclusion a $GL(V)$-module map. It turns out that *we know all such maps.* The Pieri rule §8.1.3 says they are all of the form $S^dV \subset S_\pi V^* \otimes S_\mu V$ where the Young diagram of μ is obtained from the Young diagram of π by adding d boxes, with no two boxes added to the same column. To make this completely correct, we need to consider sequences with negative integers, where e.g., the Young diagram of $(-d)$ should be thought of as $-d$ boxes in a row. Alternatively, one can work with $SL(V)$-modules, as then $S_{(-d)}V = S_{(d^{\mathbf{v}-1})}V$ as $SL(V)$-modules. For every such pair there is exactly one $GL(V)$-inclusion. Call the resulting linear map a *Young-flattening.*

The classical case is $\pi = (-k)$ and $\mu = (d-k)$, or in terms of $SL(V)$-modules, $\pi = (k^{\mathbf{v}-1})$ and $\mu = (k^{\mathbf{v}}, d-k)$. The main example in [**LO13**], called a *Koszul flattening* was constructed as follows: take the classical flattening $P_{k,d-k} : S^kV^* \to S^{d-k}V$ and tensor it with Id_{Λ^pV} for some p, to get a map $S^kV^* \otimes \Lambda^pV \to S^{d-k}V \otimes \Lambda^pV$. Now include $S^{d-k}V \subset S^{d-k-1}V \otimes V$, so we have a map $S^kV^* \otimes \Lambda^pV \to S^{d-k-1}V \otimes V \otimes \Lambda^pV$ and finally skew-symmetrize the last two factors to obtain a map

$$P^{\wedge p}_{k,d-k} : S^kV^* \otimes \Lambda^pV \to S^{d-k-1}V \otimes \Lambda^{p+1}V. \tag{8.2.1}$$

If one views this as a map $S^dV \otimes (S^kV^* \otimes \Lambda^pV) \to S^{d-k-1}V \otimes \Lambda^{p+1}V$, it is a $GL(V)$-module map. By the Pieri rule,

$$(S^kV^* \otimes \Lambda^pV)^* = S_{k,1^{\mathbf{v}-p}}V \otimes \det{}^{-1} \oplus S_{k+1,1^{\mathbf{v}-p-1}}V \otimes \det{}^{-1}$$

and

$$S^{d-k-1}V\otimes\Lambda^{p+1}V = S_{d-k-1,1^{p+1}}V \oplus S_{d-k,1^p}V.$$

Although in practice one usually works with the map (8.2.1), the map is zero except restricted to the map between irreducible modules:

$$[S_{k,1^{\mathbf{v}-p}}V^*\otimes \det^{-1}]^* \to S_{d-k,1^p}V.$$

The method of shifted partial derivatives §7.5 is a type of Young flattening which I will call a *Hilbert flattening*, because it is via Hilbert functions of Jacobian ideals. It is the symmetric cousin of the Koszul flattening: take the classical flattening $P_{k,d-k} : S^kV^* \to S^{d-k}V$ and tensor it with Id_{S^pV} for some p, to get a map $S^kV^*\otimes S^pV \to S^{d-k}V\otimes S^pV$. Now simply take the projection $S^{d-k}V\otimes S^pV \to S^{d-k+p}V$, to obtain a map

$$S^kV^*\otimes S^pV \to S^{d-k+p}V. \tag{8.2.2}$$

The target is an irreducible $GL(V)$-module, so the pruning is easier here.

8.2.2. The case of $A\otimes B\otimes C$. Young flattenings can also be defined for tensors. For tensors in $A\otimes B\otimes C$, the Koszul flattenings $T_A^{\wedge p} : \Lambda^pA\otimes B^* \to \Lambda^{p+1}A\otimes C$ used in §2.6 appear to be the only useful cases.

In principle there are numerous inclusions

$$A\otimes B\otimes C \subset (S_\pi A\otimes S_\mu B\otimes S_\nu C)^*\otimes(S_{\tilde\pi}A\otimes S_{\tilde\mu}B\otimes S_{\tilde\nu}C),$$

where the Young diagram of $\tilde\pi$ is obtained from the Young diagram of π by adding a box (and similarly for μ,ν), and the case of Koszul flattenings is where (up to permuting the three factors) $\pi = (1^p)$, $\mu = (1^{\mathbf{b}-1})$ (so $S_\mu B \simeq B^*$) and $\nu = \emptyset$.

Exercise 2.5.0.1 already indicates why symmetrization is not useful, and an easy generalization of it proves this to be the case. But perhaps additional skew-symmetrization could be useful: Let $T \in A\otimes B\otimes C$ and consider $T\otimes \mathrm{Id}_{\Lambda^pA}\otimes \mathrm{Id}_{\Lambda^qB}\otimes \mathrm{Id}_{\Lambda^sC}$ as a linear map $B^*\otimes\Lambda^qB^*\otimes\Lambda^pA\otimes\Lambda^sC \to \Lambda^qB^*\otimes\Lambda^pA\otimes A\otimes\Lambda^sC\otimes C$. Now quotient to the exterior powers to get a map:

$$T_{p,q,s} : \Lambda^{q+1}B^*\otimes\Lambda^pA\otimes\Lambda^sC \to \Lambda^qB^*\otimes\Lambda^{p+1}A\otimes\Lambda^{s+1}C.$$

This generalizes the map $T_A^{\wedge p}$ which is the case $q = s = 0$. Claim: this generalization does not give better lower bounds for border rank than Koszul flattenings when $\mathbf{a} = \mathbf{b} = \mathbf{c}$. (Although it is possible it could give better lower bounds for some particular tensor.) If T has rank one, say $T = a\otimes b\otimes c$, the image of $T_{p,q,s}$ is

$$\Lambda^q(b^\perp)\otimes(a\wedge\Lambda^pA)\otimes(c\wedge\Lambda^sC).$$

Here $b^\perp := \{\beta \in B^* \mid \beta(b) = 0\}$. The image of $(a\otimes b\otimes c)_{p,q,s}$ has dimension

$$d_{p,q,s} := \binom{\mathbf{b}-1}{q}\binom{\mathbf{a}-1}{p}\binom{\mathbf{c}-1}{s}.$$

Thus the size $rd_{p,q,s}+1$ minors of $T_{p,q,s}$ potentially give equations for the variety of tensors of border rank at most r. We have nontrivial minors as long as

$$rd_{p,q,s}+1 \leq \min\{\dim(\Lambda^q B^*{\otimes}\Lambda^{p+1}A{\otimes}\Lambda^{s+1}), \dim(\Lambda^{q+1}B^*{\otimes}\Lambda^p A{\otimes}\Lambda^s C)\},$$

i.e., as long as

$$r < \min\{\frac{\binom{\mathbf{b}}{q}\binom{\mathbf{a}}{p+1}\binom{\mathbf{c}}{s+1}}{\binom{\mathbf{b}-1}{q}\binom{\mathbf{a}-1}{p}\binom{\mathbf{c}-1}{s}}, \frac{\binom{\mathbf{b}}{q+1}\binom{\mathbf{a}}{p}\binom{\mathbf{c}}{s}}{\binom{\mathbf{b}-1}{q}\binom{\mathbf{a}-1}{p}\binom{\mathbf{c}-1}{s}}\},$$

i.e.

$$r < \min\{\frac{\mathbf{abc}}{(\mathbf{b}-q)(p+1)(s+1)}, \frac{\mathbf{abc}}{(q+1)(\mathbf{a}-p)(\mathbf{c}-s)}\}.$$

Consider the case $q=0$, so we need

$$r < \min\{\frac{\mathbf{ac}}{(p+1)(s+1)}, \frac{\mathbf{abc}}{(\mathbf{a}-p)(\mathbf{c}-s)}\}.$$

Let's specialize to $\mathbf{a}=\mathbf{c}$, $p=q$, so we need

$$r < \min\{\frac{\mathbf{a}^2}{(p+1)^2}, \frac{\mathbf{a}^2\mathbf{b}}{(\mathbf{a}-p)^2}\}.$$

Consider the case $\mathbf{a}=mp$ for some m. Then if m is large, the first term is large, but the second is very close to $\mathbf{b}$. So unless the dimensions are unbalanced, one is unlikely to get any interesting equations out of these Young flattenings.

8.2.3. General perspective. Let $X \subset \mathbb{P}V$ be a G-variety for some reductive group G, where $V=V_\lambda$ is an irreducible G-module. The goal is to find irreducible G-modules V_μ, V_ν such that $V_\lambda \subset V_\mu{\otimes}V_\nu$. Then given $v \in V$, we obtain a linear map $v_{\mu,\nu}: V_\mu^* \to V_\nu$. Say the maximum rank of such a linear map for $x \in X$ is q, then the size $(qr+1)$-minors of $v_{\mu,\nu}$ test membership $\sigma_r(X)$.

8.3. Additional uses of representation theory to find modules of equations

In this section, I briefly cover additional techniques for finding modules of polynomials in ideals of G-varietieis. I am brief because either the methods are not used in this book or they are described at length in [**Lan12**].

8.3.1. A naïve algorithm. Let $X \subset \mathbb{P}W$ be a G-variety. We are primarily interested in the cases $X = \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)) \subset \mathbb{P}(A \otimes B \otimes C)$, where $G = GL(A) \times GL(B) \times GL(C)$ and $X = \mathcal{D}et_n \subset \mathbb{P}S^n\mathbb{C}^{n^2}$, where $G = GL_{n^2}$. Since the ideal of X will be a G-module, we can look for irreducible modules in the ideal of X by testing highest weight vectors. If $U \subset S^dW^*$ is an irreducible G-module with highest weight vector u, then $U \subset I(X)$ if and only if $u \in I(X)$ because if $u \in I(X)$ then $g(u) \in I(X)$ for all $g \in G$ and such vectors span U. Thus in each degree d, we can in principle determine $I_d(X)$ by a finite calculation. In practice we test each highest weight vector u on a "random" point $[x] \in X$. If $u(x) \neq 0$, we know for sure that $U \not\subset I_d(X)$. If $u(x) = 0$, then with extremely high probability (probability one if the point is truly randomly chosen), we have $U \subset I(X)$. After testing several such points, we have high confidence in the result. Once one has a candidate module by such tests, one can often prove it is in the ideal by different methods.

More precisely, if S^dW^* is multiplicity free, there are a finite number of highest weight vectors to check. If a given module has multiplicity m, then we need to take a basis $u_1, \dots, u_m$ of the highest weight space, test on say $x_1, \dots, x_q$ with $q \geq m$ if $\sum_j y_j u_j(x_s) = 0$ for some constants $y_1, \dots, y_m$ and all $1 \leq s \leq q$.

To carry out this procedure in our two cases we would respectively need

- A method to decompose $S^d(A \otimes B \otimes C)^*$ (resp. $S^d(S^n\mathbb{C}^{n^2})$) into irreducible submodules.

- A method to explicitly write down highest weight vectors.

There are several systematic techniques for accomplishing both these tasks that work well in small cases, but as cases get larger one needs to introduce additional methods to be able to carry out the calculations in practice. The first task amounts to the well-studied problem of computing *Kronecker coefficients* defined in §8.9.2. I briefly discuss the second task in §8.7.2.

8.3.2. Enhanced search using numerical methods. Rather than discuss the general theory, I outline the method used in [**HIL13**] to find equations for $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$. First fix a "random" linear space $L \subset \mathbb{P}^{63}$ of dimension 4 (i.e., codim$\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$) and consider the finite set $Z := \sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)) \cap L$. The first objective is to compute points in Z, with a goal of computing every point in Z. To this end, we first computed one point in Z as follows. One first picks a random point $x^* \in \sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$, which is easy since an open dense subset of $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ is parameterizable. Let $\tilde{L}$ be a system of 59 linear forms so that L is the zero locus of $\tilde{L}$ and let L_{t,x^*} be the zero locus of

$L(x) - t \cdot L(x^*)$. Since $x^* \in \sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)) \cap L_{1,x^*}$, a point in Z is the endpoint of the path defined by $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)) \cap L_{t,x^*}$ at $t = 0$ starting from x^* at $t = 1$.

Even though the above process could be repeated for different x^* to compute points in Z, we instead used *monodromy loops* [**SVW01**] for generating more points in Z. After performing 21 loops, the number of points in Z that we computed stabilized at 15,456. The *trace test* [**SVW02**] shows that 15,456 is indeed the degree of $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ thereby showing we had indeed computed Z.

From Z, we performed two computations. The first was the *membership test* of [**HS13**] for deciding if $M_{\langle 2\rangle} \in \sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$, which requires tracking 15,456 homotopy paths that start at the points of Z and end on a $\mathbb{P}^4$ containing $M_{\langle 2\rangle}$. In this case, each of these 15,456 paths converged to points in $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ distinct from $M_{\langle 2\rangle}$ providing a numerical proof that $M_{\langle 2\rangle} \notin \sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$. The second was to compute the minimal degree of nonzero polynomials vanishing on $Z \subset L$. This sequence of polynomial interpolation problems showed that no nonconstant polynomials of degree ≤ 18 vanished on Z and hence $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$. The 15456×8855 matrix resulting from polynomial interpolation of homogeneous forms of degree 19 in 5 variables using the approach of [**GHPS14**] has a 64-dimensional null space. Thus, the minimal degree of nonzero polynomials vanishing on $Z \subset L$ is 19, showing $\dim I_{19}(\sigma_6) \leq 64$.

The next objective was to verify that the minimal degree of nonzero polynomials vanishing on the curve $C := \sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)) \cap K \subset K$ for a fixed "random" linear space $K \subset \mathbb{P}^{63}$ of dimension 5 was also 19. We used 50,000 points on C and the 50000×42504 matrix resulting from polynomial interpolation of homogeneous forms of degree 19 in 6 variables using the approach of [**GHPS14**] also has a 64-dimensional null space. With this agreement, we decomposed $S^6(\mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \mathbb{C}^4)$ and looked for a 64-dimensional submodule. The only reasonable candidate was to take a copy of $S_{5554}\mathbb{C}^4 \otimes S_{5554}\mathbb{C}^4 \otimes S_{5554}\mathbb{C}^4$. We found a particular copy that was indeed in the ideal and then proved that $M_{\langle 2\rangle}$ is not contained in $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ by showing a polynomial in this module did not vanish on it. The evaluation was numerical, so the result was:

Theorem 8.3.2.1. [**HIL13**] *With extremely high probability, the ideal of $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ is generated in degree 19 by the module $S_{5554}\mathbb{C}^4 \otimes S_{5554}\mathbb{C}^4 \otimes S_{5554}\mathbb{C}^4$. This module does not vanish on $M_{\langle 2\rangle}$.*

In the same paper, a copy of the trivial degree twenty module $S_{5555}\mathbb{C}^4 \otimes S_{5555}\mathbb{C}^4 \otimes S_{5555}\mathbb{C}^4$ is shown to be in the ideal of $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ by symbolic methods, giving a new proof that:

Theorem 8.3.2.2. [**Lan06, HIL13**] $\underline{\mathbf{R}}(M_{\langle 2\rangle}) = 7$.

The same methods have shown $I_{45}(\sigma_{15}(Seg(\mathbb{P}^3 \times \mathbb{P}^7 \times \mathbb{P}^8)) = 0$ and that $I_{186,999}(\sigma_{18}(Seg(\mathbb{P}^6 \times \mathbb{P}^6 \times \mathbb{P}^6)) = 0$ (this variety is a hypersurface), both of which are relevant for determining the border rank of $M_{\langle 3\rangle}$, see [**HIL13**].

8.3.3. Inheritance. Inheritance is a general technique for studying equations of G-varieties that come in series. It is discussed extensively in [**Lan12**, §7.4,16.4].

If $V \subset W$ then $S_{\overline{\pi}}V \subset V^{\otimes d}$ induces a module $S_{\overline{\pi}}W \subset W^{\otimes d}$ by, e.g., choosing a basis of W whose first $\mathbf{v}$ vectors are a basis of V. Then the two modules have the same highest weight vector and one obtains the $GL(W)$-module the span of the $GL(W)$-orbit of the highest weight vector.

Because the realizations of $S_\pi V$ in $V^{\otimes d}$ do not depend on the dimension of V, one can reduce the study of $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ to that of $\sigma_r(Seg(\mathbb{P}^{r-1} \times \mathbb{P}^{r-1} \times \mathbb{P}^{r-1}))$. As discussed in §3.3.1 this latter variety is an orbit closure, namely the orbit closure of the so called unit tensor $M_{\langle 1\rangle}^{\oplus r}$.

Proposition 8.3.3.1. [**LM04**, Prop. 4.4] *For all vector spaces B_j with $\dim B_j = \mathbf{b}_j \geq \dim A_j = \mathbf{a}_j \geq r$, a module $S_{\overline{\mu}_1}B_1 \otimes \cdots \otimes S_{\overline{\mu}_n}B_n$ such that $\ell(\mu_j) \leq \mathbf{a}_j$ for all j, is in $I_d(\sigma_r(Seg(\mathbb{P}B_1^* \times \cdots \times \mathbb{P}B_n^*)))$ if and only if $S_{\overline{\mu}_1}A_1 \otimes \cdots \otimes S_{\overline{\mu}_n}A_n$ is in $I_d(\sigma_r(Seg(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_n^*)))$.*

Corollary 8.3.3.2. [**LM04, AR03**] *Let $\dim A_j \geq r$, $1 \leq j \leq n$. The ideal of $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ is generated by the modules inherited from the ideal of $\sigma_r(Seg(\mathbb{P}^{r-1} \times \cdots \times \mathbb{P}^{r-1}))$ and the modules generating the ideal of $Sub_{r,\dots,r}$. The analogous scheme and set-theoretic results hold as well.*

8.3.4. Prolongation. Prolongation (and multi-prolongation) provides a systematic method to find equations for secant varieties that is particularly effective for secant varieties of homogeneous varieties. For a general discussion and proofs see [**Lan12**, §7.5]. For our purposes, we will need the following:

Proposition 8.3.4.1. *Given $X \subset \mathbb{P}V^*$, $I_{r+1}(\sigma_r(X)) = (I_2(X) \otimes S^{r-1}V) \cap S^{r+1}V$.*

Proposition 8.3.4.2. *Let $X \subset \mathbb{P}V$ be a variety with $I_{d-1}(X) = 0$. Then for all $\delta < (d-1)r$, $I_\delta(\sigma_r(X)) = 0$.*

Corollary 8.3.4.3. $I_d(\sigma_d(v_n(\mathbb{P}V)) = 0$.

8.4. Necessary conditions for modules of polynomials to be useful for GCT

The polynomial $\ell^{n-m}\operatorname{perm}_m \in S^n\mathbb{C}^{n^2}$ has two properties that can be studied individually: it is *padded*: i.e., it is divisible by a large power of a linear form, and its zero set is a *cone* with a $n^2 - m^2 - 1$ dimensional vertex, that is, it only uses $m^2 + 1$ of the n^2 variables in an expression in good coordinates. Both of these properties restrict the types of polynomials we should look for. Equipped with the language of representation theory we can give precise descriptions of the modules we should restrict our attention to, which I call *GCT useful*.

I begin with the study of cones, as this is a classical topic.

8.4.1. Cones. Recall the subspace variety $Sub_k(S^dV) \subset \mathbb{P}S^dV$ from §6.2.2, the polynomials whose associated hypersurfaces are cones with a $\mathbf{v} - k$ dimensional vertex.

Proposition 8.4.1.1. *$I_\delta(Sub_k(S^dV))$ consists of the isotypic components of the modules $S_\pi V^*$ appearing in $S^\delta(S^dV^*)$ such that $\ell(\pi) > k$.*

Exercise 8.4.1.2: (2!) Prove Proposition 8.4.1.1. ⊚

With just a little more effort, one can prove these degree $k+1$ equations generate the ideal:

Theorem 8.4.1.3. [**Wey03**, Cor. 7.2.3] *The ideal of $Sub_k(S^dV)$ is generated by the image of $\Lambda^{k+1}V^* \otimes \Lambda^{k+1}S^{d-1}V^* \subset S^{k+1}(V^* \otimes S^{d-1}V^*)$ in $S^{k+1}(S^dV^*)$, the size $k+1$ minors of the $(k, d-k)$-flattening.*

Aside 8.4.1.4. Here is further information about the variety $Sub_k(S^dV)$: First, it is a good example of a variety admitting a *Kempf-Weyman desingularization*, a type of desingularization that G-varieties often admit. Rather than discuss the general theory here (see [**Wey03**] for a full exposition or [**Lan12**, Chap. 17] for an elementary introduction), I just explain this example, which gives a proof of Theorem 8.4.1.3, although more elementary proofs are possible. The Grassmannian $G(k, V)$ has a tautological vector bundle $\pi : \mathcal{S} \to G(k, V)$, where the fiber over a k-plane E is just the k-plane itself. The whole bundle is a sub-bundle of the trivial bundle $\underline{V}$ with fiber V. Consider the bundle $S^d\mathcal{S} \subset S^d\underline{V}$. We have a projection map $p : S^d\underline{V} \to S^dV$. The image of $S^d\mathcal{S}$ under p is $\hat{S}ub_k(S^dV)$. Moreover, the map is a *desingularization*, that is $S^d\mathcal{S}$ is smooth, and the map to $\hat{S}ub_k(S^dV)$ is generically one to one. In particular, this implies $\dim \hat{S}ub_k(S^dV) = \dim(S^d\mathcal{S}) = \binom{k+d-1}{d} + d(\mathbf{v} - k)$. One obtains the entire minimal free resolution of $Sub_k(S^dV)$ by "pushing down" a tautological resolution "upstairs".

8.4.2. The variety of padded polynomials. Define the *variety of padded polynomials*

$$\mathrm{Pad}_{n-m}(S^nW) :=$$
$$\mathbb{P}\{P \in S^nW \mid \ P = \ell^{n-m}h, \text{for some } \ell \in W, h \in S^mW\} \subset \mathbb{P}S^nW.$$

Proposition 8.4.2.1. [**KL14**] *Let $\pi = (p_1, \ldots, p_{\mathbf{w}})$ be a partition of dn. If $p_1 < d(n-m)$, then the isotypic component of $S_\pi W^*$ in $S^d(S^nW^*)$ is contained in $I_d(\mathrm{Pad}_{n-m}(S^nW))$.*

Proof. Fix a (weight) basis $e_1, \ldots, e_{\mathbf{w}}$ of W with dual basis $x_1, \ldots, x_{\mathbf{w}}$ of W^*. Note any element $\ell^{n-m}h \in \mathrm{Pad}_{n-m}(S^nW)$ is in the $GL(W)$-orbit of $(e_1)^{n-m}\tilde{h}$ for some $\tilde{h}$, so it will be sufficient to show that the ideal in degree d contains the modules vanishing on the orbits of elements of the form $(e_1)^{n-m}h$. The highest weight vector of any copy of $S_{(p_1,\ldots,p_{\mathbf{w}})}W^*$ in $S^d(S^nW^*)$ will be a linear combination of vectors of the form $m_I := (x_1^{i_1^1} \cdots x_{\mathbf{w}}^{i_{\mathbf{w}}^1}) \cdots (x_1^{i_1^d} \cdots x_{\mathbf{w}}^{i_{\mathbf{w}}^d})$, where $i_j^1 + \cdots + i_j^d = p_j$ for all $1 \leq j \leq \mathbf{w}$ and $i_1^k + \cdots + i_{\mathbf{w}}^k = n$ for all $1 \leq k \leq d$ as these are all the vectors of weight π in $S^d(S^nW)$. Each m_I vanishes on any $(e_1)^{n-m}h$ unless $p_1 \geq d(n-m)$. (For a coordinate-free proof, see [**KL14**].) □

What we really need to study is the variety $\mathrm{Pad}_{n-m}(Sub_k(S^dW))$ of padded cones.

Proposition 8.4.2.2. [**KL14**] *$I_d(\mathrm{Pad}_{n-m}(Sub_k(S^nW^*)))$ consists of all modules $S_{\overline{\pi}}W$ such that $S_{\overline{\pi}}\mathbb{C}^k$ is in the ideal of $\mathrm{Pad}_{n-m}(S^n\mathbb{C}^{k*})$ and all modules whose associated partition has length at least $k+1$.*

Exercise 8.4.2.3: (2) Prove Proposition 8.4.2.2.

In summary:

Proposition 8.4.2.4. *In order for a module $S_{(p_1,\ldots,p_\ell)}W^*$, where $(p_1, \ldots, p_\ell)$ is a partition of dn to be GCT-useful for showing $\ell^{n-m}\,\mathrm{perm}_m \notin \overline{GL_{n^2} \cdot \det_n}$ we must have*

- *$\ell \leq m^2 + 1$, and*
- *$p_1 > d(n-m)$.*

8.5. Proofs of results stated earlier regarding $\mathcal{D}et_n$

8.5.1. Proof of Proposition 6.7.1.2. Recall $P_\Lambda(M) = \overline{\det}_n(M_\Lambda, \ldots, M_\Lambda, M_S)$ from §6.7.1 where $M = M_\Lambda + M_S$ is the decomposition of the matrix M into its skew-symmetric and symmetric components. We need to show $\overline{GL_{n^2} \cdot [P_\Lambda]}$ has codimension one in $\mathcal{D}et_n$ and is not contained in $\mathrm{End}(\mathbb{C}^{n^2}) \cdot [\det_n]$. We compute the stabilizer of P_Λ inside $GL(E \otimes E)$, where $E = \mathbb{C}^n$. The action of $GL(E)$ on $E \otimes E$ by $M \mapsto gMg^T$ preserves P_Λ up to scale, and the Lie algebra of the stabilizer of $[P_\Lambda]$ is a $GL(E)$ submodule of $\mathrm{End}(E\otimes E)$. Note that $\mathfrak{sl}(E) = S_{21^{n-2}}E$ and $\mathfrak{gl}(E) = \mathfrak{sl}(E) \oplus \mathbb{C}$. Decompose $\mathrm{End}(E\otimes E)$ as a $GL(E)$-module:

$$\begin{aligned}\mathrm{End}(E\otimes E) &= \mathrm{End}(\Lambda^2 E) \oplus \mathrm{End}(S^2E) \oplus \mathrm{Hom}(\Lambda^2E, S^2E) \oplus \mathrm{Hom}(S^2E, \Lambda^2E)\\ &= \Lambda^2E\otimes\Lambda^2E^* \oplus S^2E\otimes S^2E^* \oplus \Lambda^2E^*\otimes S^2E \oplus S^2E^*\otimes\Lambda^2E\\ &= (\mathfrak{gl}(E) \oplus S_{2^2,1^{n-2}}E) \oplus (\mathfrak{gl}(E) \oplus S_{4,2^{n-1}}E) \oplus (\mathfrak{sl}(E) \oplus S_{3,1^{n-2}}E) \oplus (\mathfrak{sl}(E) \oplus S_{3^2,2^{n-2}}E)\end{aligned} \tag{8.5.1}$$

By testing highest weight vectors, one concludes the Lie algebra of G_{P_Λ} is isomorphic to $\mathfrak{gl}(E) \oplus \mathfrak{gl}(E)$, which has dimension $2n^2 = \dim G_{\det_n} + 1$, implying $\overline{GL(W)\cdot P_\Lambda}$ has codimension one in $\overline{GL(W)\cdot [\det_n]}$. Since it is not contained in the orbit of the determinant, it must be an irreducible component of its boundary. Since the zero set is not a cone, P_Λ cannot be in $\mathrm{End}(W)\cdot \det_n$ which consists of $GL(W)\cdot\det_n$ plus cones, as any element of $\mathrm{End}(W)$ either has a kernel or is invertible.

Exercise 8.5.1.1: (3) Verify by testing on highest weight vectors that the only summands in (8.5.1) annihilating P_Λ are those in $\mathfrak{gl}(E) \oplus \mathfrak{gl}(E)$. Note that as a $\mathfrak{gl}(E)$-module, $\mathfrak{gl}(E) = \mathfrak{sl}(E) \oplus \mathbb{C}$ so one must test the highest weight vector of $\mathfrak{sl}(E)$ and $\mathbb{C}$.

8.5.2. The module structure of the equations for hypersurfaces with degenerate duals. Recall the equations for $\mathcal{D}_{k,d,N} \subset \mathbb{P}(S^d\mathbb{C}^{N*})$ that we found in §6.5.1 that enabled the lower bound $\overline{\mathrm{dc}}(\mathrm{perm}_m) \geq \frac{m^2}{2}$. In this subsection I describe the module structure of the equations, in particular I verify that they are GCT-useful.

Write $P = \sum_J \tilde{P}_J x^J$ with the sum over $|J| = d$. The weight of a monomial $\tilde{P}_{J_0}x^{J_0}$ is $J_0 = (j_1, \dots, j_n)$. Adopt the notation $[i] = (0, \dots, 0, 1, 0, \dots, 0)$ where the 1 is in the i-th slot and similarly for $[i,j]$ where there are two 1's. The entries of $P_{d-2,2}$ are, for $i \neq j$, $(P_{d-2,2})_{i,j} = P_{I+[i,j]}x^I$, and for $i = j$, $P_{I+2[i]}x^I$, where $|I| = d-2$, and P_J is $\tilde{P}_J$ with the coefficient adjusted, e.g., $P_{(d,0,\dots,0)} = d(d-1)\tilde{P}_{(d,0,\dots,0)}$ etc.. (This won't matter because we are only concerned with the weights of the coefficients, not their values.) To determine the highest weight vector, take $L = \mathrm{span}\{e_1, e_2\}$, $F = \mathrm{span}\{e_1, \dots, e_{k+3}\}$. The highest weight term of

$$(x_1^{e-d}P|_L) \wedge (x_1^{e-d-1}x_2P|_L) \wedge \cdots \wedge (x_2^{e-d}P|_L) \wedge (\det_{k+3}(P_{d-2,2}\,|_F))|_L$$

is the coefficient of $x_1^e \wedge x_1^{e-1}x_2 \wedge \cdots \wedge x_1^{e-(e-d+2)}x_2^{e-d+2}$. It will not matter how we distribute these for the weight, so take the coefficient of x_1^e in $(\det_{k+3}(P_{d-2,2}\,|_F))|_L$. It has leading term

$$P_{(d,0,\dots,0)}P_{(d-2,2,0,\dots,0)}P_{(d-2,0,2,0,\dots,0)}\cdots P_{(d-2,0,\dots,0,2,0,\dots,0)}$$

which is of weight $(d+(k+2)(d-2), 2^{k+2})$. For each $(x_1^{e-d-s}x_2^sP|_L)$ take the coefficient of $x_1^{e-s-1}x_2^{s+1}$ which has the coefficient of $P_{(d-1,1,0,\dots,0)}$ each time, to get a total weight contribution of $((e-d+1)(d-1),(e-d+1),0,\dots,0)$ from these terms. Adding the weights together, and recalling that $e = (k+3)(d-2)$ the highest weight is

$$(d^2k + 2d^2 - 2dk - 4d + 1, dk + 2d - 2k - 3, 2^{k+1}),$$

which may be written as

$$((k+2)(d^2 - 2d) + 1, (k+2)(d-2) + 1, 2^{k+1}).$$

In summary:

Theorem 8.5.2.1. [**LMR13**] *The ideal of the variety $\mathcal{D}_{k,d,N} \subset \mathbb{P}(S^d\mathbb{C}^{N*})$ contains a copy of the GL_N-module $S_{\pi(k,d)}\mathbb{C}^N$, where*

$$\pi(k,d) = ((k+2)(d^2 - 2d) + 1,\ d(k+2) - 2k - 3,\ 2^{k+1}).$$

Since $|\pi| = d(k+2)(d-1)$, these equations have degree $(k+2)(d-1)$.

Observe that the module $\pi(2n-2,n)$ indeed satisfies the requirements to be $(m, \frac{m^2}{2})$-GCT useful, as $p_1 = 2n^3 - 2n^2 + 1 > n(n-m)$ and $\ell(\pi(2n-2,n)) = 2n+1$.

Recall that $Dual_{k,d,N} \subset \mathbb{P}S^d\mathbb{C}^{N*}$ is the Zariski closure of the irreducible polynomials whose hypersurfaces have k-dimensional dual varieties. The following more refined information may be useful for studying permanent v. determinant:

Proposition 8.5.2.2. [**LMR13**] *When restricted to the open subset of irreducible hypersurfaces in $S^d\mathbb{C}^{N^*}$, $Dual_{k,d,N} = \mathcal{D}_{k,d,N}$ as sets.*

Proof. Let $P \in \mathcal{D}_{k,d,N}$ be irreducible. For each $(L,F) \in G(2,F) \times G(k+3,V)$ one obtains set-theoretic equations for the condition that $P|_L$ divides $Q|_L$, where $Q = \det(P_{d-2,2}|_F)$. But P divides Q if and only if restricted to each plane P divides Q, so these conditions imply that the dual variety of the irreducible hypersurface $Z(P)$ has dimension at most k. $\square$

Theorem 8.5.2.3. [**LMR13**] *$\mathcal{D}et_n$ is an irreducible component of $\mathcal{D}_{2n-2,n,n^2}$*

The proof of Theorem 8.5.2.3 requires familiarity with Zariski tangent spaces to schemes. Here is an outline: Given two schemes, X, Y with X irreducible and $X \subseteq Y$, an equality of Zariski tangent spaces, $T_xX = T_xY$ for some $x \in X_{smooth}$, implies that X is an irreducible component of Y (and in particular, if Y is irreducible, that $X = Y$). The following theorem is a more precise version:

Theorem 8.5.2.4. [**LMR13**] *The scheme $\mathcal{D}_{2n-2,n,n^2}$ is smooth at $[\det_n]$, and $\mathcal{D}et_n$ is an irreducible component of $\mathcal{D}_{2n-2,n,n^2}$.*

The idea of the proof is as follows: We need to show $T_{[det_n]}\mathcal{D}_{n,2n-2,n^2} = T_{[det_n]}\mathcal{D}et_n$. We already know $T_{[det_n]}\mathcal{D}et_n \subseteq T_{[det_n]}\mathcal{D}_{n,2n-2,n^2}$. Both of these vector spaces are $G_{\det_n}$-submodules of $S^n(E\otimes F)$. In 8.7.1.3 you will prove the *Cauchy formula* that $S^n(E\otimes F) = \oplus_{|\pi|=n} S_\pi E\otimes S_\pi F$.

Exercise 8.5.2.5: (2) Show that $[\det_n] = S_{1^n}E\otimes S_{1^n}F$ and $\hat{T}_{\det_n}\mathcal{D}et_n = S_{1^n}E\otimes S_{1^nF} \oplus S_{2,1^{n-1}}E\otimes S_{2,1^{n-1}}F$. ⊚

So as a $GL(E)\times GL(F)$-module, $T_{[det_n]}\mathcal{D}et_n = S_{2,1^{n-2}}E\otimes S_{2,1^{n-2}}F$. The problem now becomes to show that none of the other modules in $S^n(E\otimes F)$ are in $T_{[det_n]}\mathcal{D}_{n,2n-2,n^2}$. To do this, it suffices to check a single point in each module. A first guess would be to check highest weight vectors, but these are not so easy to write down in any uniform manner. Fortunately in this case there is another choice, namely the *immanants* IM_π defined by Littlewood [**Lit06**], the unique trivial representation of the diagonal $\mathfrak{S}_n$ in the weight $((1^n),(1^n))$ subspace of $S_\pi E\otimes S_\pi F$, and the proof in [**LMR13**] proceeds by checking that none of these other than $IM_{2,1^{n-2}}$ are contained in $T_{[det_n]}\mathcal{D}_{n,2n-2,n^2}$.

Theorem 8.5.2.4 implies that the $GL(W)$-module of highest weight $\pi(2n-2,n)$ given by Theorem 8.5.2.1 gives local equations at $[\det_n]$ of $\mathcal{D}et_n$, of degree $2n(n-1)$. Since $Sub_k(S^n\mathbb{C}^N) \subset Dual_{k,n,N}$, the zero set of the equations is strictly larger than $\mathcal{D}et_n$. Recall that $\dim Sub_k(S^n\mathbb{C}^{n^2}) = \binom{k+n+1}{n} + (k+2)(N-k-2) - 1$. For $k = 2n-2, N = n^2$, this is larger than the dimension of the orbit of $[\det_n]$, and therefore $Dual_{2n-2,n,n^2}$ is not irreducible.

8.6. Double-Commutant and algebraic Peter-Weyl Theorems

I now present the theory that will enable proofs of the statments in §8.1 and §3.5.

8.6.1. Algebras and their modules. For an algebra $\mathcal{A}$, and $a \in \mathcal{A}$ the space $\mathcal{A}a$ is a left ideal and a (left) $\mathcal{A}$-module.

Let G be a finite group. Recall from §3.5.1 the notation $\mathbb{C}[G]$ for the space of functions on G, and $\delta_g \in \mathbb{C}[G]$ for the function such that $\delta_g(h) = 0$ for $h \neq g$ and $\delta_g(g) = 1$. Define a representation $L : G \to GL(\mathbb{C}[G])$ by $L(g)\delta_h = \delta_{gh}$ and extending the action linearly. Define a second representation $R : G \to GL(\mathbb{C}[G])$ by $R(g)\delta_h = \delta_{hg^{-1}}$. Thus $\mathbb{C}[G]$ is a $G\times G$-module under the representation (L,R), and for all $c \in \mathbb{C}[G]$, the ideal $\mathbb{C}[G]c$ is a G-module under the action L.

A representation $\rho : G \to GL(V)$ induces an algebra homomorphism $\mathbb{C}[G] \to \mathrm{End}(V)$, and it is equivalent that V is a G-module or a left $\mathbb{C}[G]$-module.

A module M (for a group, ring, or algebra) is *simple* if it has no proper submodules. The module M is *semi-simple* if it may be written as the direct sum of simple modules. An algebra is *completely reducible* if all its modules are semi-simple. For groups alone I will continue to use the terminology *irreducible* for a simple module, *completely reducible* for a semi-simple module, and *reductive* for a group such that all its modules can be decomposed into a direct sum of irreducible modules.

Exercise 8.6.1.1: (2) Show that if $\mathcal{A}$ is completely reducible, V is an $\mathcal{A}$-module with an $\mathcal{A}$-submodule $U \subset V$, then there exists an $\mathcal{A}$-invariant complement to U in V and a projection map $\pi : V \to U$ that is an $\mathcal{A}$-module map. ⊚

8.6.2. The double-commutant theorem. Our sought-after decomposition of $V^{\otimes d}$ as a $GL(V)$-module will be obtained by exploiting the fact that the actions of $GL(V)$ and $\mathfrak{S}_d$ on $V^{\otimes d}$ commute. In this subsection we study commuting actions in general, as this is also the basis of the generalized DFT used in the Cohn-Umans method §3.5, and the starting point of the program of [**MS01, MS08**]. References for this section are [**Pro07**, Chap. 6] and [**GW09**, §4.1.5]. Let $S \subset \mathrm{End}(V)$ be any subset. Define the *centralizer* or *commutator* of S to be

$$S' := \{X \in \mathrm{End}(V) \mid Xs = sX \ \forall s \in S\}$$

Proposition 8.6.2.1.

(1) $S' \subset \mathrm{End}(V)$ *is a sub-algebra.*

(2) $S \subset (S')'$.

Exercise 8.6.2.2: (1!) Prove Proposition 8.6.2.1.

Theorem 8.6.2.3. *[Double-Commutant Theorem] Let $\mathcal{A} \subset \mathrm{End}(V)$ be a completely reducible associative algebra. Then $\mathcal{A}'' = \mathcal{A}$.*

There is an ambiguity in the notation S' as it makes no reference to V, so instead introduce the notation $\mathrm{End}_S(V) := S'$.

Proof. By Proposition 8.6.2.1, $\mathcal{A} \subseteq \mathcal{A}''$. To show the reverse inclusion, say $T \in \mathcal{A}''$. Fix a basis $v_1, \dots, v_{\mathbf{v}}$ of V. Since the action of T is determined by its action on a basis, we need to find $a \in \mathcal{A}$ such that $av_j = Tv_j$ for $j = 1, \dots, \mathbf{v}$. Let $w := v_1 \oplus \cdots \oplus v_{\mathbf{v}} \in V^{\oplus \mathbf{v}}$ and consider the submodule $\mathcal{A}w \subseteq V^{\oplus \mathbf{v}}$. By Exercise 8.6.1.1, there exists an $\mathcal{A}$-invariant complement to this submodule and an $\mathcal{A}$-equivariant projection $\pi : V^{\oplus \mathbf{v}} \to \mathcal{A}w \subset V^{\oplus \mathbf{v}}$, that is, a projection π that commutes with the action of $\mathcal{A}$, i.e., $\pi \in \mathrm{End}_{\mathcal{A}}(V^{\oplus \mathbf{v}})$.

Since $T \in \mathrm{End}_{\mathcal{A}}(V)$ and the action on $V^{\oplus \mathbf{v}}$ is diagonal, $T \in \mathrm{End}_{\mathcal{A}}(V^{\oplus \mathbf{v}})$. We have $\pi(Tw) = T(\pi(w))$ but $T(\pi(w)) = T(w) = Tv_1 \oplus \cdots \oplus Tv_{\mathbf{v}}$. But since $\pi(Tw) \in \mathcal{A}w$, there must be some $a \in \mathcal{A}$ such that $aw = T(w)$, i.e., $av_1 \oplus \cdots \oplus av_{\mathbf{v}} = Tv_1 \oplus \cdots \oplus Tv_{\mathbf{v}}$, i.e., $av_j = Tv_j$ for $j = 1, \ldots, \mathbf{v}$. $\square$

Burnside's theorem, stated in §3.5, has a similar proof:

Theorem 8.6.2.4. *[Burnside] Let $\mathcal{A} \subseteq \mathrm{End}(V)$ be a finite dimensional simple sub-algebra of $\mathrm{End}(V)$ (over $\mathbb{C}$) acting irreducibly on a finite-dimensional vector space V. Then $\mathcal{A} = \mathrm{End}(V)$. More generally, a finite dimensional semi-simple associative algebra $\mathcal{A}$ over $\mathbb{C}$ is isomorphic to a direct sum of matrix algebras:*

$$\mathcal{A} \simeq Mat_{d_1 \times d_1}(\mathbb{C}) \oplus \cdots \oplus Mat_{d_q \times d_q}(\mathbb{C})$$

for some $d_1, \ldots, d_q$.

Proof. For the first assertion, we need to show that given $X \in \mathrm{End}(V)$, there exists $a \in \mathcal{A}$ such that $av_j = Xv_j$ for $v_1, \ldots, v_{\mathbf{v}}$ a basis of V. Now just imitate the proof of Theorem 8.6.2.3. For the second assertion, note that $\mathcal{A}$ is a direct sum of simple algebras. $\square$

Remark 8.6.2.5. A pessimist could look at this theorem as a disappointment: all kinds of interesting looking algebras over $\mathbb{C}$, such as the group algebra of a finite group, are actually just plain old matrix algebras in disguise. An optimist could view this theorem as stating there is a rich structure hidden in matrix algebras. We will determine the matrix algebra structure explicitly for the group algebra of a finite group.

8.6.3. Consequences for reductive groups. Let S be a group or algebra and let V, W be S-modules, adopt the notation $\mathrm{Hom}_S(V, W)$ for the space of S-module maps $V \to W$, i.e.,

$$\begin{aligned}\mathrm{Hom}_S(V, W) :&= \{f \in \mathrm{Hom}(V, W) \mid s(f(v)) = f(s(v)) \ \forall\ s \in S,\ v \in V\} \\ &= (V^* \otimes W)^S.\end{aligned}$$

Theorem 8.6.3.1. *Let G be a reductive group and let V be a G-module. Then*

(1) *The commutator $\mathrm{End}_G(V)$ is a semi-simple algebra.*
(2) *The isotypic components of G and $\mathrm{End}_G(V)$ in V coincide.*
(3) *Let U be one such isotypic component, say for irreducible representations A of G and B of $\mathrm{End}_G(V)$. Then, as a $G \times \mathrm{End}_G(V)$-module,*

$$U = A \otimes B,$$

as an $\mathrm{End}_G(V)$-module

$$B = \mathrm{Hom}_G(A, U),$$

and as a G-module

$$A = \mathrm{Hom}_{\mathrm{End}_G(V)}(B, U).$$

In particular, $\mathrm{mult}(A, V) = \dim B$ *and* $\mathrm{mult}(B, V) = \dim A$.

Example 8.6.3.2. Below we will see that $\mathrm{End}_{GL(V)}(V^{\otimes d}) = \mathbb{C}[\mathfrak{S}_d]$. Recall from Equation (4.5.1) that $V^{\otimes 3} = S^3V \oplus (S_{21}V)^{\oplus 2} \oplus \Lambda^3 V$ as a $GL(V)$ module. As an $\mathfrak{S}_3 \times GL(V)$-module, we have the decomposition $V^{\otimes 3} = ([3] \otimes S^3V) \oplus ([2,1] \otimes S_{21}V) \oplus ([1,1,1] \otimes \Lambda^3 V)$ which illustrates Theorem 8.6.3.1.

To prove the theorem, we will need the following lemma:

Lemma 8.6.3.3. *For $W \subset V$ a G-submodule and $f \in \mathrm{Hom}_G(W, V)$, there exists $a \in \mathrm{End}_G(V)$ such that $a|_W = f$.*

Proof. Consider the diagram

$$\begin{array}{ccc} \mathrm{End}(V) & \longrightarrow & \mathrm{Hom}(W, V) \\ \downarrow & & \downarrow \\ \mathrm{End}_G(V) & \longrightarrow & \mathrm{Hom}_G(W, V) \end{array}$$

The vertical arrows are G-equivariant projections, and the horizontal arrows are restriction of domain of a linear map. The diagram is commutative. Since the vertical arrows and upper horizontal arrow are surjective, we conclude the lower horizontal arrow is surjective as well. $\square$

Proof of Theorem 8.6.3.1. I first prove (3): The space $\mathrm{Hom}_G(A, V)$ is an $\mathrm{End}_G(V)$-module because for $s \in \mathrm{Hom}_G(A, V)$ and $a \in \mathrm{End}_G(V)$, the composition $as : A \to V$ is still a G-module map. We need to show (i) that $\mathrm{Hom}_G(A, V)$ is an irreducible $\mathrm{End}_G(V)$-module and (ii) that the isotypic component of A in V is $A \otimes \mathrm{Hom}_G(A, V)$.

To show (i), it is sufficient to show that for all nonzero $s, t \in \mathrm{Hom}_G(A, V)$, there exists $a \in \mathrm{End}_G(V)$ such that $at = s$. Since tA and sA are isomorphic G-modules, by Lemma 8.6.3.3, there exists $a \in \mathrm{End}_G(V)$ extending an isomorphism between them, so $a(tA) = sA$, i.e., $at : A \to sA$ is an isomorphism. Consider the isomorphism $S : A \to sA$, given by $a \mapsto sa$, so $S^{-1}at$ is a nonzero scalar c times the identity. Then $\tilde{a} := \frac{1}{c}a$ has the property that $\tilde{a}t = s$.

To see (ii), let U be the isotypic component of A, so $U = A \otimes B$ for some vector space B. Let $b \in B$ and define a map $\tilde{b} : A \to V$ by $a \mapsto a \otimes b$, which is a G-module map where the action of G on the target is just the action on the first factor. Thus $B \subseteq \mathrm{Hom}_G(A, V)$. Any G-module map $A \to V$ by definition has image in U, so equality holds.

(3) implies (2).

To see (1), note that $\mathrm{End}_G(V)$ is semi-simple because if the irreducible $G \times \mathrm{End}_G(V)$-components of V are U_i, then $\mathrm{End}_G(V) = \oplus_i \mathrm{End}_G(U_i) = \oplus_i \mathrm{End}_G(A_i \otimes B_i) = \oplus_i \mathrm{End}(B_i)$. □

8.6.4. Matrix coefficients. For affine algebraic reductive groups, one can obtain all their (finite dimensional) irreducible representations from the ring of regular functions on G, denoted $\mathbb{C}[G]$. Here G is an *affine algebraic variety*, i.e., a subvariety of $\mathbb{C}^N$ for some N, so $\mathbb{C}[G] = \mathbb{C}[x_1, \dots, x_N]/I(G)$.

Exercise 8.6.4.1: (1!) Show that GL_n is an affine algebraic subvariety of $\mathbb{C}^{n^2+1}$ with coordinates (x^i_j, z) by considering the polynomial $z\,\mathrm{det}_n(x^i_j) - 1$.

Thus $\mathbb{C}[GL(W)]$ may be defined to be the restriction of polynomial functions on $\mathbb{C}^{n^2+1}$ to the subvariety isomorphic to $GL(W)$. (For a finite group, all complex-valued functions on G are algebraic, so this is consistent with our earlier notation.) If $G \subset GL(W)$ is defined by algebraic equations, this also enables us to define $\mathbb{C}[G]$ because $G \subset GL(W)$ is a subvariety. In this section and the next, we study the structure of $\mathbb{C}[G]$ as a G-module.

Let G be an affine algebraic group. Let $\rho : G \to GL(V)$ be a finite dimensional representation of G. Define a map $i_V : V^* \otimes V \to \mathbb{C}[G]$ by $i_V(\alpha \otimes v)(g) := \alpha(\rho(g)v)$. The space of functions $i_V(V^* \otimes V)$ is called the *space of matrix coefficients of* V.

Exercise 8.6.4.2: (1)

i) Show i_V is a $G \times G$-module map.
ii) Show that if V is irreducible, i_V is injective. ⊚
iii) If we choose a basis $v_1, \dots, v_{\mathbf{v}}$ of V with dual basis $\alpha^1, \dots, \alpha^{\mathbf{v}}$, then $i_V(\alpha^i \otimes v_j)(g)$ is the (i, j)-th entry of the matrix representing $\rho(g)$ in this basis (which explains the name "matrix coefficients").
iv) Compute the matrix coefficient basis of the three irreducible representations of $\mathfrak{S}_3$ in terms of the standard basis δ_σ, $\sigma \in \mathfrak{S}_3$.
v) Let $G = GL_2\mathbb{C}$, write $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$, and compute the matrix coefficient basis as functions of a, b, c, d when $V = S^2\mathbb{C}^2, S^3\mathbb{C}^2$ and $\Lambda^2\mathbb{C}^2$.

Theorem 8.6.4.3. *Let G be an affine algebraic group and V an irreducible G-module. Then $i_V(V^* \otimes V)$ equals the isotypic component of type V in $\mathbb{C}[G]$ under the action L and the isotypic component of V^* in $\mathbb{C}[G]$ under the action R.*

Proof. It suffices to prove one of the assertions, consider the action L. Let $j : V \to \mathbb{C}[G]$ be a G-module map under the action L. We need to show

$j(V) \subset i_V(V^*\otimes V)$. Define $\alpha \in V^*$ by $\alpha(v) := j(v)(\mathrm{Id})$. Then $j(v) = i_V(\alpha\otimes v)$, as $j(v)g = j(v)(g \cdot \mathrm{Id}) = j(gv)(\mathrm{Id}) = \alpha(gv) = i_V(\alpha\otimes v)g$. □

8.6.5. Application to representations of finite groups. Theorem 8.6.4.3 implies:

Theorem 8.6.5.1. *Let G be a finite group, then as a $G \times G$-module under the action (L, R) and as an algebra,*

$$\mathbb{C}[G] = \bigoplus_i V_i\otimes V_i^* \tag{8.6.1}$$

where the sum is over all the distinct irreducible representations of G.

Exercise 8.6.5.2: (1!) Let G be a finite group and H a subgroup. Show that $\mathbb{C}[G/H] = \oplus_i V_i^*\otimes(V_i)^H$ as a G-module under the action L.

8.6.6. The algebraic Peter-Weyl Theorem. Theorem 8.6.5.1 has generalizations to several different classes of groups. The most important generalization for our purposes is to reductive algebraic groups. The proof is unchanged, except that one has an infinite sum:

Theorem 8.6.6.1. *Let G be a reductive algebraic group. Then there are only countably many non-isomorphic finite dimensional irreducible G-modules. Let Λ_G^+ denote a set indexing the irreducible G-modules, and for $\lambda \in \Lambda_G^+$, let V_λ denote the irreducible module associated to λ. Then, as a $G\times G$-module*

$$\mathbb{C}[G] = \bigoplus_{\lambda\in\Lambda_G^+} V_\lambda\otimes V_\lambda^*.$$

Corollary 8.6.6.2. *Let $H \subset G$ be a closed subgroup. Then, as a G-module,*

$$\mathbb{C}[G/H] = \mathbb{C}[G]^H = \bigoplus_{\lambda\in\Lambda_G^+} V_\lambda\otimes(V_\lambda^*)^H = \bigoplus_{\lambda\in\Lambda_G^+} V_\lambda^{\oplus \dim(V_\lambda^*)^H}. \tag{8.6.2}$$

Here G acts on the V_λ and $(V_\lambda^*)^H$ is just a vector space whose dimension records the multiplicity of V_λ in $\mathbb{C}[G/H]$.

Exercise 8.6.6.3: (2!) Use Corollary 8.6.6.2 to determine $\mathbb{C}[v_d(\mathbb{P}V)]$ (even if you already know it by a different method).

8.6.7. Characters and representations of finite groups. Let $\rho : G \to GL(V)$ be a representation. Define a function $\chi_\rho : G \to \mathbb{C}$ by $\chi_\rho(g) = \mathrm{trace}(\rho(g))$. The function χ_ρ is called the *character* of ρ.

Exercise 8.6.7.1: (1) Show that χ_ρ is constant on conjugacy classes of G.

A function $f : G \to \mathbb{C}$ such that $f(hgh^{-1}) = f(g)$ for all $g, h \in G$ is called a *class function*.

Exercise 8.6.7.2: (1) Show that for representations $\rho_j : G \to GL(V_j)$, that $\chi_{\rho_1 \oplus \rho_2} = \chi_{\rho_1} + \chi_{\rho_2}$.

Exercise 8.6.7.3: (1) Given $\rho_j : G \to GL(V_j)$ for $j = 1, 2$, define $\rho_1 {\otimes} \rho_2 : G \to GL(V_1 {\otimes} V_2)$ by $\rho_1 {\otimes} \rho_2(g)(v_1 {\otimes} v_2) = \rho_1(g)v_1 {\otimes} \rho_2(g)v_2$. Show that $\chi_{\rho_1 \otimes \rho_2} = \chi_{\rho_1} \chi_{\rho_2}$.

Theorem 8.6.5.1 is not yet useful, as we do not yet know what the V_i are. Let $\mu_i : G \to GL(V_i)$ denote the representation. It is not difficult to show that the functions χ_{μ_i} are linearly independent in $\mathbb{C}[G]$. (One uses a G-invariant Hermitian inner-product $\langle \chi_V, \chi_W \rangle := \frac{1}{|G|} \sum_{g \in G} \chi_V(g) \overline{\chi_W(g)}$ and shows that they are orthogonal with respect to this inner-product, see, e.g., [**FH91**, §2.2].) On the other hand, we have a natural basis of the class functions, namely the δ-functions on each conjugacy class. Let C_j be a conjugacy class of G and define $\delta_{C_j} := \sum_{g \in C_j} \delta_g$. It is straight-forward to see, via the DFT (§3.5.1), that the span of the δ_{C_j}'s equals the span of the χ_{μ_i}'s, that is *the number of distinct irreducible representations of* G *equals the number of conjugacy classes* (see, e.g, [**FH91**, §2.2] for the standard proof using the Hermitian inner-product on class functions and [**GW09**, §4.4] for a DFT proof).

Remark 8.6.7.4. The classical Heisenberg uncertainty principle from physics, in the language of mathematics, is that it is not possible to localize both a function and its Fourier transform. A discrete analog of this uncertainty principle holds, in that the transforms of the delta functions have large support in terms of matrix coefficients and vice versa. Another manifestation of the uncertainty principle is that the relation between these two bases can be complicated.

8.6.8. Representations of $\mathfrak{S}_d$. When $G = \mathfrak{S}_d$, we get lucky: we will associate irreducible representations directly to conjugacy classes.

The conjugacy class of a permutation is determined by its decomposition into a product of disjoint cycles. The conjugacy classes of $\mathfrak{S}_d$ are in 1-1 correspondence with the set of partitions of d: to a partition $\pi = (p_1, \ldots, p_r)$ one associates the conjugacy class of an element with disjoint cycles of lengths $p_1, \ldots, p_r$. Let $[\pi]$ denote the isomorphism class of the irreducible $\mathfrak{S}_d$-module associated to the partition π. In summary:

Proposition 8.6.8.1. *The irreducible representations of* $\mathfrak{S}_d$ *are indexed by partitions of* d.

Thus as an $\mathfrak{S}_d \times \mathfrak{S}_d$ module under the (L, R)-action:

$$\mathbb{C}[\mathfrak{S}_d] = \bigoplus_{|\pi|=d} [\pi]^*_L {\otimes} [\pi]_R. \tag{8.6.3}$$

We can say even more: as $\mathfrak{S}_d$ modules, $[\pi]$ is isomorphic to $[\pi]^*$. This is usually proved by first noting that for any finite group G, and any irreducible representation μ, $\chi_{\mu^*} = \overline{\chi_\mu}$ where the overline denotes complex conjugate and then observing that the characters of $\mathfrak{S}_d$ are all real-valued functions. Thus we may rewrite (8.6.3) as

$$\mathbb{C}[\mathfrak{S}_d] = \bigoplus_{|\pi|=d} [\pi]_L \otimes [\pi]_R. \tag{8.6.4}$$

Exercise 8.6.8.2: (2) Show $[d] \subset [\pi]\otimes[\mu]$ if and only if $\pi = \mu$. ⊚

8.7. Representations of $\mathfrak{S}_d$ and $GL(V)$

In this section we finally obtain our goal of the decomposition of $V^{\otimes d}$ as a $GL(V)$-module. Representations of $GL(V)$ occurring in $V^{\otimes}$ may also be indexed by partitions, which is explained in §8.7.1 where Schur-Weyl duality is stated and proved.

8.7.1. Schur-Weyl duality. We have already seen that the actions of $GL(V)$ and $\mathfrak{S}_d$ on $V^{\otimes d}$ commute.

Proposition 8.7.1.1. $\mathrm{End}_{GL(V)}(V^{\otimes d}) = \mathbb{C}[\mathfrak{S}_d]$.

Proof. We will show that $\mathrm{End}_{\mathbb{C}[\mathfrak{S}_d]}(V^{\otimes d})$ is the algebra generated by $GL(V)$ and conclude by the double commutant theorem. Since

$$\begin{aligned}\mathrm{End}(V^{\otimes d}) &= V^{\otimes d}\otimes(V^{\otimes d})^* \\ &\simeq (V\otimes V^*)^{\otimes d}\end{aligned}$$

under the re-ordering isomorphism, $\mathrm{End}(V^{\otimes d})$ is spanned by elements of the form $X_1\otimes\cdots\otimes X_d$ with $X_j \in \mathrm{End}(V)$, i.e., elements of $\hat{S}eg(\mathbb{P}(\mathrm{End}(V)) \times \cdots \times \mathbb{P}(\mathrm{End}(V)))$. The action of $X_1\otimes\cdots\otimes X_d$ on $v_1\otimes\cdots\otimes v_d$ induced from the $GL(V)^{\times d}$-action is $v_1\otimes\cdots\otimes v_d \mapsto (X_1v_1)\otimes\cdots\otimes(X_dv_d)$. Since $g \in GL(V)$ acts by $g\cdot(v_1\otimes\cdots\otimes v_d) = gv_1\otimes\cdots\otimes gv_d$, the image of $GL(V)$ in $(V\otimes V^*)^{\otimes d}$ lies in $S^d(V\otimes V^*)$, in fact it is a Zariski open subset of $\hat{v}_d(\mathbb{P}(V\otimes V^*))$ which spans $S^d(V\otimes V^*)$. In other words, the algebra generated by $GL(V)$ is $S^d(V\otimes V^*) \subset \mathrm{End}(V^{\otimes d})$. But by definition $S^d(V\otimes V^*) = [(V\otimes V^*)^{\otimes d}]^{\mathfrak{S}_d}$ and we conclude. □

Applying Theorem 8.6.3.1 we obtain:

Theorem 8.7.1.2. *[Schur-Weyl duality] The irreducible decomposition of $V^{\otimes d}$ as a $GL(V)\times\mathbb{C}[\mathfrak{S}_d]$-module (equivalently, as a $GL(V)\times\mathfrak{S}_d$-module) is*

$$V^{\otimes d} = \bigoplus_{|\pi|=d} S_\pi V\otimes[\pi], \tag{8.7.1}$$

where $S_\pi V := \mathrm{Hom}_{\mathfrak{S}_d}([\pi], V^{\otimes d})$ *is an irreducible* $GL(V)$*-module.*

Note that as far as we know, $S_\pi V$ could be zero. (It will be zero whenever $\ell(\pi) \geq \dim V$.)

Exercise 8.7.1.3: (2) Show that as a $GL(E)\times GL(F)$-module, $S^d(E\otimes F) = \oplus_{|\pi|=d} S_\pi E\otimes S_\pi F$. This is called the *Cauchy formula.* ⊚

8.7.2. Explicit realizations of representations of $\mathfrak{S}_d$ and $GL(V)$. By Theorem 8.6.5.1 we may explicitly realize each irreducible $\mathfrak{S}_d$-module via some projection from $\mathbb{C}[\mathfrak{S}_d]$. The question is, which projections?

Given π we would like to find elements $c_{\overline{\pi}} \in \mathbb{C}[\mathfrak{S}_d]$ such that $\mathbb{C}[\mathfrak{S}_d]c_{\overline{\pi}}$ is isomorphic to $[\pi]$. I write $\overline{\pi}$ instead of just π because the elements are far from unique; there is a vector space of dimension $\dim[\pi]$ of such projection operators by Theorem 8.6.5.1, and the overline signifies a specific realization. In other words, the $\mathfrak{S}_d$-module map $RM_{c_{\overline{\pi}}} : \mathbb{C}[\mathfrak{S}_d] \to \mathbb{C}[\mathfrak{S}_d]$, $f \mapsto fc_{\overline{\pi}}$ should kill all $\mathfrak{S}_d^R$-modules not isomorphic to $[\pi]_R$, and the image should be $[\pi]_L\otimes v$ for some $v \in [\pi]_R$. If this works, as a bonus, the map $c_{\overline{\pi}} : V^{\otimes d} \to V^{\otimes d}$ induced from the $\mathfrak{S}_d$-action will have image $S_{\overline{\pi}}V\otimes v$ for the same reason, where $S_{\overline{\pi}}V$ is some realization of $S_\pi V$ and $v \in [\pi]$.

Here are projection operators for the two representations we understand well:

When $\pi = (d)$, there is a unique up to scale $c_{\overline{(d)}}$ and it is easy to see it must be $c_{\overline{(d)}} := \sum_{\sigma\in\mathfrak{S}_d} \delta_\sigma$, as the image of $RM_{c_{\overline{(d)}}}$ is clearly the line through $c_{\overline{(d)}}$ on which $\mathfrak{S}_d$ acts trivially. Note further that $c_{\overline{(d)}}(V^{\otimes d}) = S^dV$ as desired.

When $\pi = (1^d)$, again we have a unique up to scale projection, and its clear we should take $c_{\overline{(1^d)}} = \sum_{\sigma\in\mathfrak{S}_d} \mathrm{sgn}(\sigma)\delta_\sigma$ as the image of any δ_τ will be $\mathrm{sgn}(\tau)c_{\overline{(1^d)}}$, and $c_{\overline{(1^d)}}(V^{\otimes d}) = \Lambda^dV$.

The only other representation of $\mathfrak{S}_d$ that we have a reasonable understanding of is the standard representation $\pi = (d-1,1)$ which corresponds to the complement of the trivial representation in the permutation action on $\mathbb{C}^d$. A basis of this space could be given by $e_1 - e_d, e_2 - e_d, \ldots, e_{d-1} - e_d$. Note that the roles of $1,\ldots,d-1$ in this basis are the "same" in that if one permutes them, one gets the same basis, and that the role of d with respect to any of the other e_j is "skew" in some sense. To capture this behavior, consider

$$c_{\overline{(d-1,1)}} := (\delta_{\mathrm{Id}} - \delta_{(1,d)})(\sum_{\sigma\in\mathfrak{S}_{d-1}[d-1]} \delta_\sigma)$$

where $\mathfrak{S}_{d-1}[d-1] \subset \mathfrak{S}_d$ is the subgroup permuting the elements $\{1, \ldots, d-1\}$. Note that $c_{\overline{(d-1,1)}}\delta_\tau = c_{\overline{(d-1,1)}}$ for any $\tau \in \mathfrak{S}_{d-1}[d-1]$ so the image is of dimension at most $d = \dim(\mathbb{C}[\mathfrak{S}_d]/\mathbb{C}[\mathfrak{S}_{d-1}])$.

Exercise 8.7.2.1: (2) Show that the image is $d-1$ dimensional.

Now consider $RM_{c_{\overline{(d-1,1)}}}(V^{\otimes d})$: after re-orderings, it is the image of the composition of the maps

$$V^{\otimes d} \to V^{\otimes d-2} \otimes \Lambda^2 V \to S^{d-1}V \otimes V.$$

In particular, in the case $d = 3$, it is the image of

$$V \otimes \Lambda^2 V \to S^2 V \otimes V,$$

which is isomorphic to $S_{21}V$, as was mentioned in in §4.5.

Here is the general recipe to construct an $\mathfrak{S}_d$-module isomorphic to $[\pi]$: fill the Young diagram of a partition π of d with integers $1, \ldots, d$ from top to bottom and left to right. For example let $\pi = (4, 2, 1)$ and write:

$$\begin{array}{|c|c|c|c|} \hline 1 & 4 & 6 & 7 \\ \hline 2 & 5 & \multicolumn{2}{c}{} \\ \cline{1-2} 3 & \multicolumn{3}{c}{} \\ \cline{1-1} \end{array} \tag{8.7.2}$$

Define $\mathfrak{S}_{\overline{\pi}'} \simeq \mathfrak{S}_{q_1} \times \cdots \times \mathfrak{S}_{q_{p_1}} \subset \mathfrak{S}_d$ to be the subgroup that preserves the subsets of elements in the columns and $\mathfrak{S}_{\overline{\pi}}$ is the subgroup of $\mathfrak{S}_d$ permuting the elements in the rows.

Explicitly, writing $\pi = (p_1, \ldots, p_{q_1})$ and $\pi' = (q_1, \ldots, q_{p_1})$, $\mathfrak{S}_{q_1}$ permutes the elements of $\{1, \ldots, q_1\}$, $\mathfrak{S}_{q_2}$ permutes the elements of $\{q_1+1, \ldots, q_1+q_2\}$ etc.. Similarly, $\mathfrak{S}_{\overline{\pi}} \simeq \mathfrak{S}_{p_1} \times \cdots \times \mathfrak{S}_{p_\ell} \subset \mathfrak{S}_d$ is the subgroup where $\mathfrak{S}_{p_1}$ permutes the elements $\{1, q_1+1, q_1+q_2+1, \ldots, q_1+\cdots+q_{p_1-1}+1\}$, $\mathfrak{S}_{p_2}$ permutes the elements $\{2, q_1+2, q_1+q_2+2, \ldots, q_1+\cdots+q_{p_1-1}+2\}$ etc..

Define two elements of $\mathbb{C}[\mathfrak{S}_d]$: $s_{\overline{\pi}} := \sum_{\sigma\in\mathfrak{S}_{\overline{\pi}}} \delta_\sigma$ and $a_{\overline{\pi}} := \sum_{\sigma\in\mathfrak{S}_{\overline{\pi}'}} \mathrm{sgn}(\sigma)\delta_\sigma$. Fact: Then $[\pi]$ is the isomorphism class of the $\mathfrak{S}_d$-module $\mathbb{C}[\mathfrak{S}_d]a_{\overline{\pi}}s_{\overline{\pi}}$. (It is also the isomorphism class of $\mathbb{C}[\mathfrak{S}_d]s_{\overline{\pi}}a_{\overline{\pi}}$, although these two realizations are generally distinct.)

Exercise 8.7.2.2: (1) Show that $[\pi'] = [\pi]\otimes[1^d]$ as $\mathfrak{S}_d$-modules. ⊚

The space $V^{\otimes d}c_{\overline{\pi}}$ will be a copy of the module $S_\pi V$ because $c_{\overline{\pi}}$ kills all modules not isomorphic to π and maps $[\pi]$ to a one dimensional vector space. The action on $V^{\otimes d}$ is first to map it to $\Lambda^{q_1}V \otimes \cdots \otimes \Lambda^{q_{p_1}}V$, and then the module $S_\pi V$ is realized as the image of a map from this space to $S^{p_1}V \otimes \cdots \otimes S^{p_{q_1}}V$. So despite their original indirect definition, we may realize the modules $S_\pi V$ explicitly simply be skew-symmetrizations and symmetrizations.

Other realizations of $S_\pi V$ (resp. highest weight vectors for $S_\pi V$, in fact a basis of them) can be obtained by letting $\mathfrak{S}_d$ act on $V^{\otimes d}c_{\overline{\pi}}$ (resp. the highest weight vector of $V^{\otimes d}c_{\overline{\pi}}$).

Example 8.7.2.3. Consider $c_{\overline{(2,2)}}$, associated to

$$\begin{array}{|c|c|}\hline 1 & 3 \\ \hline 2 & 4 \\ \hline \end{array} \tag{8.7.3}$$

which realizes a copy of $S_{(2,2)}V \subset V^{\otimes 4}$. It first maps $V^{\otimes 4}$ to $\Lambda^2 V \otimes \Lambda^2 V$ and then maps that to $S^2V \otimes S^2 V$. Explicitly, the maps are

$$\begin{aligned} a\otimes b\otimes c\otimes c &\mapsto (a\otimes b - b\otimes a)\otimes(c\otimes d - d\otimes c) = a\otimes b\otimes c\otimes d - a\otimes b\otimes d\otimes c - b\otimes a\otimes c\otimes d + b\otimes a\otimes d\otimes c \\ &\mapsto (a\otimes b\otimes c\otimes d + c\otimes b\otimes a\otimes d + a\otimes d\otimes c\otimes b + c\otimes d\otimes a\otimes b) \\ &\quad - (a\otimes b\otimes d\otimes c + d\otimes b\otimes a\otimes c + a\otimes c\otimes d\otimes b + d\otimes c\otimes a\otimes b) \\ &\quad - (b\otimes a\otimes c\otimes d + c\otimes a\otimes b\otimes d + b\otimes d\otimes c\otimes a + c\otimes d\otimes b\otimes a) \\ &\quad + (b\otimes a\otimes d\otimes c + d\otimes a\otimes b\otimes c + b\otimes c\otimes d\otimes a + d\otimes c\otimes b\otimes a) \end{aligned}$$

Exercise 8.7.2.4: Let $v_1 = (e_1 \wedge e_2)\otimes e_1$ and $v_2 = e_1 \otimes (e_1 \wedge e_2)$ denote a basis of the highest weight space for $S_{21}V \otimes [2,1] \subset V^{\otimes 3}$. Compute the action of $\mathbb{Z}_3$ on these vectors and find a new basis consisting of eigenvectors for the $\mathbb{Z}_3$-action. What are the eigenvalues? ⊚

8.8. The program of [MS01, MS08]

Algebraic geometry was used successfully in [**Mul99**] to prove lower bounds in the "PRAM model without bit operations" (the model is defined in [**Mul99**]), and the proof indicated that algebraic geometry, more precisely invariant theory, could be used to resolve the **P** v. **NC** problem (a cousin of permanent v. determinant). This was investigated further in [**MS01, MS08**] and numerous sequels. In this section I present the program outlined in [**MS08**], as refined in [**BLMW11**], as well as an outline of the proof [**IP15, BIP16**] that this program cannot work as originally proposed or even the refinement discussed in [**BLMW11**]. Despite this negative news, the program has opened several promising directions, and inspired perspectives that have led to concrete advances such as [**LR15**] as described in §7.4.7. As explained below, it is conceiveably possible to carry out a variant of the program.

Independent of its viability, I expect the ingredients that went into the program of [**MS01, MS08**] will play a role in future investigations regarding Valiant's conjecture and thus are still worth studying.

8.8.1. Preliminaries. Let $W = \mathbb{C}^{\nu^2}$. Recall $\mathbb{C}[\hat{\mathcal{D}et}_n] := Sym(S^nW^*)/I(\mathcal{D}et_n)$, the homogeneous coordinate ring of the (cone over) $\mathcal{D}et_n$. This is the space of polynomial functions on $\hat{\mathcal{D}et}_n$ inherited from polynomials on the ambient space.

Since $I(\mathcal{D}et_n) \subset Sym(S^nW^*)$ is a $GL(W)$-submodule, and since $GL(W)$ is reductive, we obtain the following splitting as a $GL(W)$-module:

$$Sym(S^nW^*) = I(\mathcal{D}et_n) \oplus \mathbb{C}[\hat{\mathcal{D}et}_n].$$

In particular, if a module $S_\pi W$ appears in $Sym(S^nW^*)$ and it does not appear in $\mathbb{C}[\mathcal{D}et_n]$, it must appear in $I(\mathcal{D}et_n)$.

Now consider

$$\mathbb{C}[GL(W) \cdot \det_n] = \mathbb{C}[GL(W)/G_{\det_n}] = \mathbb{C}[GL(W)]^{G_{\det_n}}.$$

There is an injective map

$$\mathbb{C}[\hat{\mathcal{D}et}_n] \to \mathbb{C}[GL(W) \cdot \det_n]$$

given by restriction of functions. The map is an injection because any function identically zero on a Zariski open subset of an irreducible variety is identically zero on the variety.

Corollary 8.6.6.2 indicates the following plan:

Plan : *Find a module $S_\pi W^*$ not appearing in $\mathbb{C}[GL(W)/G_{\det_n}]$ that does appear in $Sym(S^nW^*)$.*

By the above discussion such a module must appear in $I(\mathcal{D}et_n)$.

Definition 8.8.1.1. An irreducible $GL(W)$-module $S_\pi W^*$ appearing in $Sym(S^nW^*)$ and not appearing in $\mathbb{C}[GL(W)/G_{\det_n}]$ is called an *orbit occurrence obstruction*.

The precise condition a module must satisfy in order to not occur in $\mathbb{C}[GL(W)/G_{\det_n}]$ is explained in Proposition 8.9.2.2. The discussion in §8.4 shows that in order to be useful, π must have a large first part and few parts.

One might object that the coordinate rings of different orbits could coincide, or at least be very close. Indeed this is the case for generic polynomials, but in GCT one generally restricts to polynomials whose symmetry groups *characterize* the orbit as follows:

Definition 8.8.1.2. Let V be a G-module. A point $P \in V$ is *characterized by* its stabilizer G_P if any $Q \in V$ with $G_Q \supseteq G_P$ is of the form $Q = cP$ for some constant c.

We have seen in §6.6 that both the determinant and permanent polynomials are characterized by their stabilizers.

Corollary 8.6.6.2 motivates the study of polynomials characterized by their stabilizers: if $P \in V$ is characterized by its stabilizer, then $G \cdot P$ is the unique orbit in V with coordinate ring isomorphic to $\mathbb{C}[G\cdot P]$ as a G-module. Thus one can think of polynomial sequences that are complete for their complexity classes and are characterized by their stabilizers as "best" representatives of their class.

Remark 8.8.1.3. All $GL(W)$-modules $S_{(p_1,\dots,p_{\mathbf{w}})}W$ may be graded using $p_1+\cdots+p_{\mathbf{w}}$ as the grading. One does not have such a grading for $SL(W)$-modules, which makes their use in GCT more difficult. In [**MS01, MS08**], it was proposed to use the $SL(W)$-module structure because it had the advantage that the SL-orbit of $\det_n$ is already closed. The disadvantage from the lack of a grading appears to outweigh this advantage.

8.9. $\mathbb{C}[GL(W)\cdot \det_n]$

Before determining the module structure of $\mathbb{C}[GL(W)\cdot \det_n]$, I start with the coordinate ring of a generic polynomial for comparison. The calculations of this section follow [**BLMW11**].

8.9.1. Generic polynomials. Let $P \in S^dV$ be generic. If $d, n > 3$, then $G_P = \{\lambda\,\mathrm{Id} : \lambda^d = 1\} \simeq \mathbb{Z}_d$, hence $GL(V)\cdot P \simeq GL(V)/\mathbb{Z}_d$, where $\mathbb{Z}_d$ acts as multiplication by the d-th roots of unity, see [**Pop75**]. (If $P \in S^dV$ is any element, then $\mathbb{Z}_d \subset G_P$.)

We need to determine the $\mathbb{Z}_d$-invariants in $GL(V)$-modules. Since $S_\pi V$ is a submodule of $V^{\otimes|\pi|}$, $\omega \in \mathbb{Z}_d$ acts on $S_\pi V\otimes(\det V)^{-s}$ by the scalar $\omega^{|\pi|-ns}$. By Theorem 8.6.6.1, we conclude the following equality of $GL(V)$-modules:

$$\mathbb{C}[GL(V)\cdot P] = \bigoplus_{(\pi,s)\ \mid\ d||\pi|-ns} (S_\pi V^*)^{\oplus \dim S_\pi V}\otimes(\det V^*)^{-s}.$$

When we pass to $\mathbb{C}[\overline{GL(V)\cdot P}] = \bigoplus_\delta S^\delta(S^dV^*)/I_\delta(\overline{GL(V)\cdot P})$ we loose all terms with $s>0$. Since degree is respected, we may write:

$$\mathbb{C}[\overline{GL(V)\cdot P}]_\delta \subseteq \bigoplus_{\pi\ \mid\ |\pi|=\delta d} (S_\pi V^*)^{\oplus \dim S_\pi V}. \tag{8.9.1}$$

Note that $\mathbb{C}[\overline{GL(V)\cdot P}]_\delta \subset S^\delta(S^dV)$, and there are far fewer modules and multiplicities in $S^\delta(S^dV)$ than on the right hand side of (8.9.1).

8.9.2. The coordinate ring of $GL_{n^2}\cdot \det_n$. Write $\mathbb{C}^{n^2} = E\otimes F$, with $E, F = \mathbb{C}^n$. We first compute the $SL(E)\times SL(F)$-invariants in $S_\pi(E\otimes F)$ where $|\pi| = d$. Recall from §8.7.1 that by definition, $S_\pi W = \mathrm{Hom}_{\mathfrak{S}_d}([\pi], W^{\otimes d})$.

Thus

$$\begin{aligned}S_\pi(E\otimes F) &= \mathrm{Hom}_{\mathfrak{S}_d}([\pi], E^{\otimes d}\otimes F^{\otimes d})\\ &= \mathrm{Hom}_{\mathfrak{S}_d}([\pi], (\bigoplus_{|\mu|=d}[\mu]\otimes S_\mu E)\otimes(\bigoplus_{|\nu|=d}[\nu]\otimes S_\nu F)\\ &= \bigoplus_{|\mu|=|\nu|=d}\mathrm{Hom}_{\mathfrak{S}_d}([\pi],[\mu]\otimes[\nu])\otimes S_\mu E\otimes S_\nu F\end{aligned}$$

The vector space $\mathrm{Hom}_{\mathfrak{S}_d}([\pi],[\mu]\otimes[\nu])$ simply records the multiplicity of $S_\mu E\otimes S_\nu F$ in $S_\pi(E\otimes F)$. The numbers $k_{\pi,\mu,\nu} = \dim \mathrm{Hom}_{\mathfrak{S}_d}([\pi],[\mu]\otimes[\nu])$ are called *Kronecker coefficients*.

Exercise 8.9.2.1: (2) Show that

$$k_{\pi,\mu,\nu} = \mathrm{Hom}_{\mathfrak{S}_d}([d],[\pi]\otimes[\mu]\otimes[\nu]) = \mathrm{mult}(S_\pi A\otimes S_\mu B\otimes S_\nu C, S^d(A\otimes B\otimes C)).$$

In particular, $k_{\pi,\mu,\nu} = k_{\mu,\pi,\nu}$.

Recall from §8.1.5 that $S_\mu E$ is a trivial $SL(E)$ module if and only if $\mu = (\delta^n)$ for some $\delta\in\mathbb{Z}$. Thus so far, we are reduced to studying the Kronecker coefficients $k_{\pi,\delta^n,\delta^n}$. Now take the $\mathbb{Z}_2$ action given by exchanging E and F into account. Write $[\mu]\otimes[\mu] = S^2[\mu]\oplus\Lambda^2[\mu]$. The first module will be invariant under $\mathbb{Z}_2 = \mathfrak{S}_2$, and the second will transform its sign under the transposition. So define the *symmetric Kronecker coefficients* $sk^\pi_{\mu,\mu} := \dim(\mathrm{Hom}_{\mathfrak{S}_d}([\pi], S^2[\mu]))$.

We conclude:

Proposition 8.9.2.2. [**BLMW11**] *Let $W = \mathbb{C}^{n^2}$. The coordinate ring of the $GL(W)$-orbit of $\det_n$ is*

$$\mathbb{C}[GL(W)\cdot \det_n] = \bigoplus_{d\in\mathbb{Z}}\bigoplus_{\pi\,|\,|\pi|=nd}(S_\pi W^*)^{\oplus sk^\pi_{d^n d^n}}.$$

While Kronecker coefficients were studied classically (if not the symmetric version), unfortunately very little is known about them. See, e.g., [**Man15a**] for recent progress and a brief history regarding Kronecker coefficients.

8.10. Plethysm coefficients

I now discuss the decomposition of $S^d(S^n V)$.

8.10.1. Asymptotics. Kronecker coefficients and the plethysm coefficients $\mathrm{mult}(S_\pi W, S^d(S^n W))$ have been well-studied in both the geometry and combinatorics literature. I briefly discuss a geometric method of L. Manivel and J. Wahl [**Wah91, Man97, Man98, Man15b, Man15a**] based on the

Bott-Borel-Weil theorem that realizes modules as spaces of sections of vector bundles on homogeneous varieties. Advantages of the method are: (i) the vector bundles come with filtrations that allow one to organize information, (ii) the sections of the associated graded bundles can be computed explicitly, giving one bounds for the coefficients, and (iii) Serre's theorem on the vanishing of sheaf cohomology tells one that the bounds are achieved eventually, and gives an upper bound for when stabilization occurs.

A basic, if not *the* basic problem in representation theory is: given a group G, an irreducible G-module U, and a subgroup $H \subset G$, decompose U as an H-module. The determination of Kronecker coefficients can be phrased this way with $G = GL(V \otimes W)$, $U = S_\lambda(V \otimes W)$ and $H = GL(V) \times GL(W)$. The determination of plethysm coefficients may be phrased as the case $G = GL(S^n V)$, $U = S^d(S^n V)$ and $H = GL(V)$.

I focus on plethysm coefficients. We want to decompose $S^d(S^n V)$ as a $GL(V)$-module, or more precisely, to obtain qualitative asymptotic information about this decomposition. Note that $S^{dn}V \subset S^d(S^n V)$ with multiplicity one. Beyond that the decomposition gets complicated. Let $x_1, \dots, x_{\mathbf{v}}$ be a basis of V, so $((x_1)^n)^d$ is the highest highest weight vector in $S^d(S^n V)$.

Define two maps (their names come from [**BIP16**]):

Define a map $\mathfrak{m}_{x_1} = \mathfrak{m}_{x_1}^{d,m,n} : S^d(S^m V) \to S^d(S^n V)$ on basis elements by

$$
\begin{aligned}
(8.10.1)\qquad & ((x_1)^{i_1^1}(x_2)^{i_2^1} \cdots (x_d)^{i_d^1}) \cdots ((x_1)^{i_1^d} \cdots (x_d)^{i_d^d}) \\
& \mapsto ((x_1)^{i_1^1+(n-m)}(x_2)^{i_2^1} \cdots (x_d)^{i_d^1}) \cdots ((x_1)^{i_1^d+(n-m)} \cdots (x_d)^{i_d^d})
\end{aligned}
$$

and extend linearly. Call $\mathfrak{m}_{x_1}$ the *inner degree lifting map.* A vector of weight $\mu = (q_1, q_2, \dots, q_d)$ is mapped under $\mathfrak{m}_{x_1}$ to a vector of weight $\pi = (p_1, \dots, p_d) := \mu + (d(n-m)) = (q_1 + d(n-m), q_2, \dots, q_d)$ in $S^d(S^n V)$.

Define a map $\mathfrak{o}_{x_1} = \mathfrak{o}_{x_1}^{\delta,d,n} : S^\delta(S^n V) \to S^d(S^n V)$ on basis elements by
(8.10.2)

$$
(x_{i_{1,1}} \cdots x_{i_{1,n}}) \cdots (x_{i_{\delta,1}} \cdots x_{i_{\delta,n}}) \mapsto (x_{i_{1,1}} \cdots x_{i_{1,n}}) \cdots (x_{i_{\delta,1}} \cdots x_{i_{\delta,n}})(x_1^n) \cdots (x_1^n)
$$

and extend linearly. Call $\mathfrak{o}_{x_1}$ the *outer degree lifting map.* A vector of weight $\mu = (q_1, q_2, \dots, q_d)$ is mapped under $\mathfrak{o}_{x_1}$ to a vector of weight $\pi = (p_1, \dots, p_d) := \mu + ((d-\delta)n) = (q_1 + (d-\delta)n, q_2, \dots, q_d)$ in $S^d(S^n V)$.

Both $\mathfrak{m}_{x_1}$ and $\mathfrak{o}_{x_1}$ take highest weight vectors to highest weight vectors, as Lie algebra raising operators annihilate x_1.

This already shows qualitative behavior if we allow the first part of a partition to grow. More generally, one has:

Theorem 8.10.1.1. [**Man97**] *Let μ be a fixed partition. Then* $\mathrm{mult}(S_{(dn-|\mu|,\mu)}, S^d(S^nV))$ *is a non-decreasing function of both d and n that is constant as soon as $d \geq |\mu|$ or $n \geq \ell(\mu)$.*

More precisely, the innner and outer degree lifting maps $\mathfrak{m}_{x_1}$ and $\mathfrak{o}_{x_1}$ are both injective and eventually isomorphisms on highest weight vectors of isotypic components of partitions $(p_1, \dots, p_{\mathbf{v}})$ with $(p_2, \dots, p_{\mathbf{v}})$ fixed and p_1 growing.

There are several proofs of the stability, the precise stabilization is proved by computing the space of sections of homogeneous vector bundles on $\mathbb{P}V$ via an elementary application of Bott's theorem (see, e.g., [**Wey03**, §4.1] for an exposition).

One way to view what we just did was to write $V = x_1 \oplus T$, so

$$S^n(x_1 \oplus T) = \bigoplus_{j=0}^{n} x_1^{n-j} \otimes S^jT. \tag{8.10.3}$$

Then decompose the d-th symmetric power of $S^n(x_1 \oplus T)$ and examine the stable behavior as we increase d and n. One could think of the decomposition (8.10.3) as the osculating sequence of the n-th Veronese embedding of $\mathbb{P}V$ at $[x_1^n]$ and the further decomposition as the *osculating sequence* (see, e.g., [**IL03**, Chap. 4]) of the d-th Veronese re-embedding of the ambient space refined by (8.10.3).

For Kronecker coefficients and more general decomposition problems the situation is more complicated in that the ambient space is no longer projective space, but a homogeneous variety, and instead of an osculating sequence, one examines *jets* of sections of a *vector bundle*.

We can now prove a partial converse to Proposition 8.4.2.1:

Proposition 8.10.1.2. [**KL14**] *Let $\pi = (p_1, \dots, p_{\mathbf{w}})$ be a partition of dn. If $p_1 \geq \min\{d(n-1), dn-m\}$, then $I_d(\mathrm{Pad}_{n-m}(S^nW))$ does not contain a copy of $S_\pi W^*$.*

Proof. The image of the space of highest weight vectors for the isotypic component of $S_\mu W^*$ in $S^d(S^mW^*)$ under $\mathfrak{m}_{x_1}$ will be in $\mathbb{C}[\mathrm{Pad}_{n-m}(S^nW)]$ because, for example, such a polynomial will not vanish on $(e_1)^{n-m}[(e_1)^{i_1^1} \cdots (e_d)^{i_d^1} + \cdots + (e_1)^{i_1^d} \cdots (e_d)^{i_d^d}]$, but if $p_1 \geq d(n-1)$ we are in the stability range.

For the sufficiency of $p_1 \geq dn-m$, note that if $p_1 \geq (d-1)n + (n-m) = dn-m$, then in an element of weight π, each of the exponents $i_1^1, \dots, i_1^d$ of x_1 must be at least $n-m$. So there again exists an element of $\mathrm{Pad}_{n-m}(S^nW)$ such that a vector of weight π does not vanish on it. $\square$

8.11. Orbit occurrence obstructions can't work: the padding problem

8.11.1. Even occurence obstructions can't work. The program of [**MS01, MS08**] proposes to use orbit occurrence obstructions to prove Valiant's conjecture. In [**IP15**] they show that this cannot work. Furthermore, in [**BIP16**] they prove that one cannot even use the following relaxation of orbit occurrence obstructions:

Definition 8.11.1.1. An irreducible $GL(W)$-module $S_\lambda W^*$ appearing in $Sym(S^n W^*)$ and not appearing in $\mathbb{C}[\hat{\mathcal{D}et}_n]$ is called an *occurrence obstruction.*

Throughout this subsection, set $W = \mathbb{C}^{n^2}$.

The extension is all the more remarkable because they essentially prove that occurrence obstructions cannot even be used to separate any degree m polynomial padded by ℓ^{n-m} in m^2 variables from

$$MJ(v_{n-k}(\mathbb{P}W), \sigma_r(v_k(\mathbb{P}W))) = \overline{GL(W) \cdot [\ell^{n-k}(x_1^k + \cdots + x_r^k)]} \tag{8.11.1}$$

for certain k, r with $kr \le n$. Here MJ is the multiplicative join of §7.2.2.

First we show that the variety (8.11.1) is contained in $\mathcal{D}et_n$. We recall the classical result:

Theorem 8.11.1.2. *[Valiant* [**Val79b**], *Liu-Regan* [**LR06**]*] Every $f \in \mathbb{C}[x_1, \dots, x_n]$ of formula size u is a projection of $\det_{u+1}$. In other words $f \in \mathrm{End}(\mathbb{C}^{(u+1)^2}) \cdot \det_{u+1}$.*

Note that the formula size of $x_1^d + \cdots + x_r^d$ is at most rd.

Corollary 8.11.1.3. [**BIP16**] *If $rs < n$ then $[\ell^{n-s}(x_1^s + \cdots + x_r^s)] \in \mathcal{D}et_n$ and thus $MJ(v_{n-k}(\mathbb{P}W), \sigma_r(v_k(\mathbb{P}W))) = \overline{GL(W) \cdot [\ell^{n-k}(x_1^k + \cdots + x_r^k)]} \subset \mathcal{D}et_n$.*

The main theorem is

Theorem 8.11.1.4. [**BIP16**] *Let $n > m^{25}$. Let $\pi = (p_1, \dots, p_\ell)$ be a partition of dn such that $\ell \le m^2 + 1$ and $p_1 \ge d(n-m)$. If a copy of $S_\pi W^*$ occurs in $S^d(S^n W^*)$ then a copy also occurs in some $\mathbb{C}[\overline{GL(W) \cdot [\ell^{n-k}(x_1^k + \cdots + x_r^k)]}]$ for some r, k with $rk < n$.*

By the above discussion, this implies occurance obstructions cannot be used to separate the permanent from the determinant.

The proof is done by splitting the problem into three cases:

(1) $d \le \sqrt{\frac{n}{m}}$
(2) $d > \sqrt{\frac{n}{m}}$ and $p_1 > dn - m^{10}$
(3) $d > \sqrt{\frac{n}{m}}$ and $p_1 \le dn - m^{10}$.

The first case is an immediate consequence of the prolongation property §8.3.4: take $r = d$ and $k = m$.

The second reduces to the first by two applications of Manivel's stability theorem:

Proposition 8.11.1.5. [**BIP16**, Prop. 5.2] *Let $|\pi| = dn$, $\ell(\pi) \leq m^2 + 1$, $p_2 \leq k$, $m^2k^2 \leq n$ and $m^2k \leq d$. If a copy of $S_\pi W$ occurs in $S^d(S^nW)$, then a copy also occurs in $\mathbb{C}[\overline{GL(W) \cdot [\ell^{n-k}(x_1^k + \cdots + x_{m^2k}^k)]}]$.*

Proof. First note that the inner degree lifting map (8.10.1) $\mathfrak{m}_\ell^{d,k,n} : S^d(S^kW^*) \to S^d(S^nW^*)$ is an isomorphism on highest weight vectors in this range because d is sufficiently large, so there exists μ with $|\mu| = dk$ and $\overline{\pi} = \overline{\mu}$. Moreover, if v_μ is a highest weight vector of weight μ, then $\mathfrak{m}_\ell^{d,k,n}(v_\mu)$ is a highest weight vector of weight π. Since m^2k is sufficently large, there exists ν with $|\mu| = m^2k^2 = (m^2k)k$, with $\overline{\nu} = \overline{\mu}$ such $v_\mu = \mathfrak{o}_{x_1}(w_\nu)$, where w_ν is a highest weight vector of weight ν in $S^{m^2k}(S^kW^*)$. Since $I_{m^2k}(\sigma_{m^2k}(v_k(\mathbb{P}W))) = 0$, we conclude that a copy of $S_\nu W^*$ is in $\mathbb{C}[\sigma_{m^2k}(v_k(\mathbb{P}W))]$ and then by the discussion above the modules corresponding to μ and π are respectively in the coordinate rings of $MJ([\ell^{d-m^2k}], \sigma_{m^2k}(v_k(\mathbb{P}W)))$ and $MJ([\ell^{n-k}], \sigma_{m^2k}(v_k(\mathbb{P}W)))$. Since $(m^2k)k \leq n$, the result follows by prolongation. $\square$

The third case relies on a building block construction made possible by the following exercise:

Exercise 8.11.1.6: (1) Show that if V is a $GL(W)$-module and $Q \in S_\lambda W \subset S^dV$ and $R \in S_\mu W \subset S^\delta V$ are both highest weight vectors, then $QR \in S_{\lambda+\mu}W \subset S^{d+\delta}V$ is also a highest weight vector.

Exercise 8.11.1.6, combined with the fact that if $Q, R \in \mathbb{C}[X]$, then $QR \in \mathbb{C}[X]$ enables the building block construction. I will show (Corollary 9.2.1.2) that for n even, there exists a copy of $S_{n^d}W$ in $\mathbb{C}[\sigma_d(v_n(\mathbb{P}W))]$, providing one of the building blocks. The difficulty in their proof lies in establishing the other base building block cases. See [**BIP16**] for the details.

Remark 8.11.1.7. In [**IP15**] the outline of the proof is similar, except there is an interesting argument by contradiction: they show that if in a certain range of n and m, if an orbit occurrence obstruction exists, then the same is true for larger values of n with the same m. But this contradicts Valiant's result (see §6.6.3) that if $n = 4^m$, then $\ell^{n-m}\,\mathrm{perm}_m \in \mathcal{D}et_n$.

It is conceivably possible to carry out the program either taking into account information about multiplicities, or with the degree m iterated matrix multiplication polynomial IMM_n^m in place of the determinant, as the latter can be compared to the permanent without padding.

8.12. Proofs of equivariant complexity bounds

remind reader what proving

While it is not formally necessary for the proof, the guide for the proof of Theorem 7.4.7.6 is the *Howe-Young duality functor* : The involution on the space of symmetric functions (see [**Mac95**, §I.2]), that exchanges elementary symmetric functions with complete symmetric functions, extends to modules of the general linear group. This functor exchanges symmetrization and skew-symmetrization. I expect it will be useful for future work regarding permanent v. determinant. The idea is that one first proves the theorem (with a supplementary hypothesis) for the determinant, which is easy, and then the functor provides a guide as to how to write the proof for the permanent. We will see this functor again in §10.3.

The supplementary hypothesis is *regularity*. By von-Zur Gathen's regularity theorem 6.3.3.1, for any determinantal expression for the permanent, we may assume the constant part of $\tilde{A}$ is the identity except for a zero in the $(1,1)$-slot, which I will denote Λ_{n-1}. Add this as a hypothesis for the determinant, and call such a *regular determinantal expression* for the determinant. The determinantal expressions of the determinant we saw in §7.3 are regular.

8.12.1. Malcev's theorem. **condense this para*** Let G be an affine complex algebraic group. The group G is *unipotent* if it is isomorphic to a subgroup of the group U_n of upper triangular matrices with 1's on the diagonal.

Given a complex algebraic group G, there exists a maximal normal unipotent subgroup $R^u(G)$, called the *unipotent radical*. The quotient $G/R^u(G)$ is reductive. Moreover there exists subgroups L in G such that $G = R^u(G)L$. In particular such L are reductive. Such a subgroup L is not unique, but any two such are conjugate in G (in fact by an element of $R^u(G)$). Such a subgroup L is called a *Levi factor of* G. A good reference is [**OV90**, Thm. 4. Chap. 6].

Malcev's theorem (see, e.g., [**OV90**, Thm. 5. Chap. 6]) states that fixing a Levi subgroup $L \subset G$ and given any reductive subgroup H of G, there exists $g \in R^u(G)$ such that $gHg^{-1} \subseteq L$.

For example, when G is a parabolic subgroup, e.g. $G = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$, we have $L = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}$ and $R^u(G) = \begin{pmatrix} \mathrm{Id}_a & * \\ 0 & \mathrm{Id}_b \end{pmatrix}$.

A more important example for us is $R^u(\mathbb{G}_{\det_n,\Lambda_{n-1}}) = (\mathbb{H} \oplus \mathbb{H}^* \otimes \ell_2)$ and a Levi subgroup is $L = (GL(\ell_2) \times GL(\mathbb{H})) \rtimes \mathbb{Z}_2$.

8.12.2. Outline of the proofs of lower bounds. ***clean*** Let $P \in S^m V^*$ be perm_m or $\det_m$. Say a regular representation $\tilde{A}$ is equivariant with respect to some $G \subseteq G_P$.

The matrix Λ_{n-1} induces a splitting of $\mathbb{C}^n = \ell \oplus \mathbb{H}$, where ℓ is a line, and similarly for $\mathbb{C}^{n*}$. Write

$$\mathcal{M}_n(\mathbb{C}) = \begin{pmatrix} \ell_1^* {\otimes} \ell_2 & \mathbb{H}^* {\otimes} \ell_2 \\ \ell_1^* {\otimes} \mathbb{H} & \mathbb{H}^* {\otimes} \mathbb{H} \end{pmatrix}, \quad \Lambda_{n-1} = \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{Id}_{\mathbb{H}} \end{pmatrix}.$$

The first step consists in lifting G to the symmetry group of $\tilde{A}$, G_A from Definition 7.4.7.2. More precisely, in each case we construct a reductive subgroup $\tilde{G}$ of G_A such that $\bar{\rho}_A : \tilde{G} \longrightarrow G$ is finite and surjective. In a first reading, it is relatively harmless to assume that $\tilde{G} \simeq G$. Then, using Malcev's theorem, after possibly conjugating $\tilde{A}$, we may assume that $\tilde{G}$ is contained in $(GL(\ell_2) \times GL(\mathbb{H})) \rtimes \mathbb{Z}_2$. Up to considering an index two subgroup of $\tilde{G}$ if necessary, we assume that $\tilde{G}$ is contained in $GL(\ell_2) \times GL(\mathbb{H})$.

Now, both $Mat_n(\mathbb{C}) = (\ell_1 {\oplus} \mathbb{H})^* {\otimes} (\ell_2 {\oplus} \mathbb{H})$ and V (via $\bar{\rho}_A$) are $\tilde{G}$-modules. Moreover, A is an equivariant embedding of V in $Mat_n(\mathbb{C})$. This turns out to be a very restrictive condition.

If $m \geq 2$ the $\ell_1^* {\otimes} \ell_2$ coefficient of $\tilde{A}$ has to be zero. Then, since $P \neq 0$, the projection of $A(V)$ on $\ell_1^* {\otimes} \mathbb{H} \simeq \mathbb{H}$ has to be non-zero. We thus have a G-submodule $\mathbb{H}_1 \subset \mathbb{H}$ isomorphic to an irreducible submodule of V. A similar argument shows that there must be another irreducible G-submodule $\mathbb{H}_2 \subset \mathbb{H}$ such that an irreducible submodule of V appears in $\mathbb{H}_1^* {\otimes} \mathbb{H}_2$.

In each case, we can construct a sequence of irreducible sub-$\tilde{G}$-modules $\mathbb{H}_k$ of $\mathbb{H}$ satisfying very restrictive conditions. This allows us to get our lower bounds.

8.12.3. Regular determinantal representations of the determinant. In this subsection $E, F \simeq \mathbb{C}^m$.

Proposition 8.12.3.1. [**LR15**] *The following is a regular determinantal representation of* $\det_m$ *that respects* $GL(E)$*. Let* $\mathbb{C}^n = \bigoplus_{j=0}^{m-1} \Lambda^j E$*, so* $n = 2^m - 1$ *and* $\mathrm{End}(\mathbb{C}^n) = \oplus_{0 \leq i,j \leq m-1} \mathrm{Hom}(\Lambda^j E, \Lambda^i E)$*. Fix an identification* $\Lambda^m E \simeq \Lambda^0 E$*. Set*

$$\Lambda_0 = \sum_{k=1}^{m-1} \mathrm{Id}_{\Lambda^k E},$$

and

$$\tilde{A} = \Lambda_0 + \sum_{k=0}^{m-1} e x_k {\otimes} f_{k+1}. \tag{8.12.1}$$

Then $\det_m = \det_n \circ \tilde{A}$ *if* $m \equiv 1, 2 \bmod 4$ *and* $\det_m = -\det_n \circ \tilde{A}$ *if* $m \equiv 0, 3 \bmod 4$. *In bases respecting the direct sum, the linear part, other than the last term which lies in the upper right block, lies just below the diagonal blocks, and all blocks other than the upper right, the diagonal and sub-diagonal are zero.*

A linear map $u : F \to E$ induces linear maps

$$\begin{aligned} u^{\wedge k} &: \Lambda^k F \to \Lambda^k E \\ v_1 \wedge \cdots \wedge v_k &\mapsto u(v_1) \wedge \cdots \wedge u(v_k). \end{aligned} \tag{8.12.2}$$

In the case $k = m$, $u^{\wedge m}$ is called the determinant of u and we denote it $\mathcal{D}et(u) \in \Lambda^m F^* \otimes \Lambda^m E$. The map

$$\begin{aligned} E \otimes F^* = \mathrm{Hom}(F, E) &\longrightarrow \Lambda^m F^* \otimes \Lambda^m E \\ u &\longmapsto \mathcal{D}et(u) \end{aligned}$$

is polynomial, homogeneous of degree m, and equivariant for the natural action of $GL(E) \times GL(F)$.

The transpose of u is

$$\begin{aligned} u^T : E^* &\longrightarrow F^*, \\ \varphi &\longmapsto \varphi \circ u. \end{aligned}$$

Hence $u^T \in F^* \otimes E$ is obtained from u by switching E and F^*, and $\mathcal{D}et(u^T) \in \Lambda^m E \otimes \Lambda^m F^*$. Moreover, $\mathcal{D}et(u^T) = \mathcal{D}et(u)^T$.

Proof of Proposition 8.12.3.1. Set $P = \det_n \circ \tilde{A}$. To analyze the action of $GL(E)$ on $\tilde{A}$, reinterpret $\mathbb{C}^{n*} \otimes \mathbb{C}^n$ without the identification $\Lambda^0 E \simeq \Lambda^m E$ as $(\oplus_{j=0}^{m-1} \Lambda^j E)^* \otimes (\oplus_{i=1}^m \Lambda^i E)$.

For each $u \in E \otimes F^*$, associate to $\tilde{A}(u)$ a linear map $\tilde{a}(u) : \oplus_{j=0}^{m-1} \Lambda^j E \to \oplus_{i=1}^m \Lambda^i E$. Then $\mathcal{D}et(\tilde{a}(u)) \in \Lambda^n(\oplus_{j=0}^{m-1} \Lambda^j E^*) \otimes \Lambda^n(\oplus_{i=1}^m \Lambda^i E)$. This space may be canonically identified as a $GL(E)$-module with $\Lambda^0 E^* \otimes \Lambda^m E \simeq \Lambda^m E$. (The identification $\Lambda^0 E \simeq \Lambda^m E$ allows one to identify this space with $\mathbb{C}$.) Using the maps (8.12.2), we get a $GL(E)$-equivariant map $\mathcal{D}et \circ \tilde{a} : E \otimes F^* \to \Lambda^m E$.

Hence for all $u \in E \otimes F^*$ and all $g \in GL(E)$,

$$\begin{aligned} \mathcal{D}et(\tilde{a}(g^{-1}u)) &= (g \cdot \mathcal{D}et)(\tilde{a}(u)) \\ &= \det(g)^{-1} \mathcal{D}et(\tilde{a}(u)). \end{aligned} \tag{8.12.3}$$

Equation (8.12.3) shows that $GL(E)$ is contained in the image of $\bar{\rho}_A$.

Equation (8.12.3) also proves that P is a scalar (possibly zero) multiple of the determinant. Consider $P(\mathrm{Id}_m) = \det_n(\tilde{A}(\mathrm{Id}_m))$. Perform a Laplace expansion of this large determinant: there is only one non-zero expansion term, so P is the determinant up to a sign.

See [**LR15**] for the verification that the sign is correct as stated. $\square$

Theorem 8.12.3.2. [**LR15**] *The smallest size regular equivariant determinantal expression for* $\det_m$ *is* $\binom{2m}{m} - 1 \sim 4^m$.

As in the case of the permanent, we can get an exponential lower bound using only about half the symmetries of the determinant.

Theorem 8.12.3.3. [**LR15**] *Let* $\tilde{A}_m : \mathcal{M}_m(\mathbb{C}) \longrightarrow \mathcal{M}_n(\mathbb{C})$ *be a regular determinantal representation of* $\det_m$ *that respects* $GL(E)$*. Then* $n \geq 2^m - 1$.

Moreover, there exists a regular determinantal representation of $\det_m$ *equivariant with respect to* $GL(E)$ *of size* $2^m - 1$.

I give the proof of Theorem 8.12.3.3, the proof of Theorem 8.12.3.2 is similar.

Proof of Theorem 8.12.3.3. Let $\tilde{A} = \Lambda_{n-1} + A : Mat_m(\mathbb{C}) \to Mat_n(\mathbb{C})$ be a regular determinantal representation of $\det_m$ that is equivariant with respect to $GL(E)$. It remains to prove that $n \geq 2^m - 1$.

After possibly conjugating $\tilde{A}$, we construct a connected reductive subgroup L of $GL(\ell_2) \times GL(\mathbb{H})$ mapping onto $GL(E)$ by $\bar{\rho}_A$.

We have an action of L on $Mat_n(\mathbb{C})$, but we would like to work with $GL(E)$. Towards this end, there exists a finite cover $\tau : \tilde{L} \longrightarrow L$ that is isomorphic to the product of a torus and a product of simple simply connected groups. In particular there exists a subgroup of $\tilde{L}$ isomorphic to $\mathbb{C}^* \times SL(E)$ such that $\bar{\rho}_A \circ \tau(\mathbb{C}^* \times SL(E)) = GL(E)$. The group $\mathbb{C}^* \times SL(E)$ acts trivially on ℓ_1, on ℓ_2 (by some character) and on $\mathbb{H}$. It acts on $Mat_n(\mathbb{C}) = (\ell_1^* \oplus \mathbb{H}) \otimes (\ell_2 \oplus \mathbb{H})$ accordingly.

The $\mathbb{C}^* \times SL(E)$-module $A(V)$ is isomorphic to the sum of m copies of E, and E is an irreducible $\mathbb{C}^* \times SL(E)$-module. In particular its equivariant projection on $\ell_1^* \otimes \ell_2$ is zero, which implies that the $(1,1)$ entry of the matrix of $\tilde{A}$ (in adapted bases) is zero. Consider the equivariant projection of $A(V)$ on $\ell_1^* \otimes \mathbb{H}$. This projection in bases goes to the remainder of the first column. It must be non-zero or $\det_n \circ \tilde{A}$ will be identically zero. Since it is equivariant, $\ell_1^* \otimes \mathbb{H} \simeq \mathbb{H}$ must contain E as a $\mathbb{C}^* \times SL(E)$-module. Similarly, examining the first row, $\mathbb{H}^* \otimes \ell_2$ has to contain E as a $\mathbb{C}^* \times SL(E)$-module.

If $m = 2$, it is possible that $\mathbb{H} \simeq E$ and $\mathbb{H} \otimes \ell_2 \simeq E^*$. In this case, $\det_2$ is a quadratic form, and we obtain a determinantal representation of size 3.

Assume now that $m \geq 3$, in particular that E and E^* are not isomorphic as $SL(E)$-modules. We just proved that $\mathbb{H}$ must contain a subspace isomorphic to E, say $\mathbb{H}_1$. Since $\mathbb{H}_1^*\otimes\ell_2$ is an irreducible $SL(E)$-module and not isomorphic to E, the projection of $A(V)$ on this factor is zero.

Choose a $\mathbb{C}^*\times SL(E)$-stable complement S_1 to $\mathbb{H}_1$ in $\mathbb{H}$. If the projection of $A(V)$ to the block $\mathbb{H}_1^*\otimes S_1$ is zero, by expanding the columns corresponding to $\mathbb{H}_1^*$, one sees that $\det_m$ is equal to the determinant in $(\ell_1\oplus S_1)^*\otimes(\ell_2\otimes S_1)$, and we can restart the proof with S_1 in place of $\mathbb{H}$.

So assume that the projection of $A(V)$ onto the block $\mathbb{H}_1^*\otimes S_1$ is non-zero. Then there must be some irreducible $(\mathbb{C}^*\times SL(E))$-submodule $\mathbb{H}_2$ such that $\mathbb{H}_1^*\otimes\mathbb{H}_2$ contains E as a submodule. Continuing, we get a sequence of simple $(\mathbb{C}^*\times SL(E))$-submodules $\mathbb{H}_1,\dots,\mathbb{H}_k$ of $\mathbb{H}$ such that E is a submodule of $\mathbb{H}_i^*\otimes\mathbb{H}_{i+1}$ and of $\ell_2\otimes\mathbb{H}_k^*$.

The situation is easy to visualize with Young diagrams. As an $SL(E)$-module, E (resp. E^*) corresponds the class of to a single box (resp. a column of $m-1$ boxes). (As a $GL(E)$-module, E^* corresponds to a diagram with -1 boxes.) The Pieri formula 8.1.3.1 implies that $E\subset S_\pi E^*\otimes S_\mu E$ if and only if the diagram of μ is obtained from the diagram of π by adding a box. Thus the sequence of Young diagrams associated to the irreducible $SL(E)$ modules $\mathbb{H}_i$ start with one box, and increases by one box at each step. Thus we must have $\mathbb{H}_k$ associated to $\pi=(c^{m-1},c-1)$ for some c. To have the proper $\mathbb{C}^*$-action, we choose the action on ℓ_2 to cancel the $(c-1)\times m$ box inside the Young diagram of π. We deduce that $(\mathbb{H}_1,\cdots,\mathbb{H}_k)=(\Lambda^1E,\Lambda^2E,\dots,\Lambda^{m-1}E\simeq E^*)$ is the unique minimal sequence of modules. In particular the dimension of $\mathbb{H}$ is at least $\sum_{k=1}^{m-1}\binom{m}{k}=2^m-2$. ☐

8.12.4. Proofs of results on determinantal representations of perm_m**.** Recall Theorem 7.4.7.1 giving a lower bound for Γ_m^E-equivariant representations for perm_m and the notation $(S^kE)_{reg}$ from §6.6.3.

Proof of Proposition 6.6.3.3. The maps $s_k(v):(S^kE)_{reg}\to(S^{k+1}E)_{reg}$ are related to the maps $ex_k(v):\Lambda^kE\to\Lambda^{k+1}E$ as follows. The sources of both maps have bases indexed by multi-indices $I=(i_1,\dots,i_k)$ with $1\le i_1<\cdots<i_k\le m$, and similarly for the targets. The maps are the same on these basis vectors except for with $s_k(v)$ all the coefficients are positive whereas with $ex_k(v)$ there are signs. Thus the polynomial computed by (6.6.7) is the same as the polynomial computed by (8.12.1) except all the y^i_j appear positively. Reviewing the sign calculation, we get the result. ☐

Outline of proof of Theorem 7.4.7.1. Write $E,F=\mathbb{C}^m$. Let $\tilde{A}$ be a determinantal representation of perm_m such that $\tilde{A}(0)=\Lambda_{n-1}$. Embed Γ_m^E in $GL(\mathrm{Hom}(F,E))$ by $g\longmapsto\{M\mapsto gM\}$. We assume the image of $\bar{\rho}_A$ contains Γ_m^E. Set $T=T^{GL(E)}$.

As in the proof of Theorem 8.12.3.3, we get a reductive subgroup L of $(GL(\ell_2) \times GL(\mathbb{H})) \rtimes \mathbb{Z}_2$ mapping onto Γ_m^E by $\bar{\rho}_A$. In the determinant case, at this point we dealt with the universal cover of the connected reductive group $GL(E)$. Here the situation is more complicated for two reasons. First, there is no "finite universal cover" of $\mathfrak{S}_m$ (see e.g. [**Ste89, Józ89**]). Second, since our group is not connected, we will have to deal with the factor $\mathbb{Z}_2$ coming from transposition, which will force us to work with a subgroup of Γ_m^E. Fortunately this will be enough for our purposes.

We first deal with the $\mathbb{Z}_2$: Since $L/(L \cap G^\circ_{\det_n,\Lambda_{n-1}})$ embeds in $G_{\det_n,\Lambda_{n-1}}/G^\circ_{\det_n,\Lambda_{n-1}} \simeq \mathbb{Z}_2$, the subgroup $L \cap G^\circ_{\det_n,\Lambda_{n-1}}$ has index 1 or 2 in L. Since the alternating group $\mathcal{A}_m$ is the only index 2 subgroup of $\mathfrak{S}_m$, $\bar{\rho}_A(L \cap G^\circ_{\det_n,\Lambda_{n-1}})$ contains $T \rtimes \mathcal{A}_m \subset \Gamma_m^E$. In any case, there exists a reductive subgroup L' of L such that $\bar{\rho}_A(L') = T \rtimes \mathcal{A}_m \subset \Gamma_m^E$.

To get around the lack of a lift, one proves that irreducible L' modules can be labeled only using labels from $\bar{\rho}_A(L') = T^{GL(E)} \rtimes \mathcal{A}_m$, see [**LR15**]. One than argues as in the proof of Theorem 8.12.3.3. The difference is that each $\mathbb{H}_s$ must be a $\mathcal{A}_m$-module that contains an irreducible $\mathcal{A}_m$-module acting transitively on size s subsets of $[m]$, so $\dim \mathbb{H}_s \geq \binom{m}{s}$, and one concludes as before. $\square$

8.13. Symmetries of other polynomials relevant for complexity theory

A central insight from GCT is that polynomials that are determined by their symmetry groups should be considered preferred representatives of their complexity classes. Although the motivation for this statement is the program discussed in §8.8, which is not viable, the idea has already guided several positive results: the symmetries of the matrix multiplication tensor have given deep insight into its decompositions, and were critical for proving its border rank lower bounds. We have already determined the symmetry groups of the determinant and permanent. In this section I present the symmetry groups of a few additional polynomials. These auxiliary results may be skipped on a first reading.

Throughout this section $G = GL(V)$, $\dim V = n$, and I use index ranges $1 \leq i, j, k \leq n$.

8.13.1. Techniques. One technique for determining G_P is to form auxiliary objects from P which have a symmetry group H that one can compute, and by construction H contains G_P. Usually it is easy to find a group H' that clearly is contained in G_P, so if $H = H'$, we are done.

One can determine the connected component of the stabilizer by a Lie algebra calculation: If we are concerned with $p \in S^dV$, the connected component of the identity of the stabilizer of p in $GL(V)$ is the connected Lie group associated to the Lie subalgebra of $\mathfrak{gl}(V)$ that annihilates p. (The analogous statement holds for tensors.) To see this, let $\mathfrak{h} \subset \mathfrak{gl}(V)$ denote the annihilator of p and let $H = exp(\mathfrak{h}) \subset GL(V)$ the corresponding Lie group. Then it is clear that H is contained in the stabilizer as $h \cdot p = exp(X) \cdot p = (Id + X + \frac{1}{2}XX + ...)p$ the first term preserves p and the remaining terms annihilate it. Similarly, if H is the group preserving p, taking the derivative of any curve in H through Id at $t = 0$ give $\frac{d}{dt}|_{t=0}h(t) \cdot p = 0$.

To recover the full stabilizer from knowledge of the connected component of the identity, we have the following observation, the first part was exploited in [**BGL14**]:

Proposition 8.13.1.1. *Let V be an irreducible $GL(W)$-module. Let G_v^0 be the identity component of the stabilizer G_v of some $v \in V$ in $GL(W)$. Then G_v is contained in the normalizer $N(G_v^0)$ of G_v^0 in $GL(W)$. If G_v^0 is semi-simple and $[v]$ is determined by G_v^0, then equality holds up to scalar multiples of the identity in $GL(W)$.*

Proof. First note that for any group H, the full group H normalizes H^0. (If $h \in H^0$, take a curve h_t with $h_0 = \mathrm{Id}$ and $h_1 = h$, then take any $g \in H$, the curve gh_tg^{-1} connects gh_1g^{-1} to the identity.) So G_v is contained in the normalizer of G_0 in $GL(W)$.

For the second assertion, let $h \in N(G^0)$ be in the normalizer. We have $h^{-1}ghv = g'v = v$ for some $g' \in G^0$, and thus $g(hv) = (hv)$. But since $[v]$ is the unique line preserved by G^0 we conclude $hv = \lambda v$ for some $\lambda \in \mathbb{C}^*$. $\square$

For those familiar with representation theory, we have the following lemma:

Lemma 8.13.1.2. [**BGL14**, Prop. 2.2] *Let G^0 be semi-simple and act irreducibly on V. Then its normalizer $N(G^0)$ is generated by G^0, the scalar matrices, and a finite group constructed as follows: Assume we have chosen a Borel for G^0, and thus have distinguished a set of simple roots Δ and a group homomorphism $Aut(\Delta) \to GL(V)$. Assume $V = V_\lambda$ is the irreducible representation with highest weight λ of G^0 and consider the subgroup $Aut(\Delta, \lambda) \subset Aut(\Delta)$ that fixes λ. Then $N(G^0) = ((\mathbb{C}^* \times G^0)/Z) \rtimes Aut(\Delta, \lambda)$.*

For the proof, see [**BGL14**].

Further techniques come from geometry. Consider the hypersurface $Z(P) := \{[v] \in \mathbb{P}V^* \mid P(v) = 0\} \subset \mathbb{P}V^*$. If all the irreducible components of P are reduced, then $G_{Z(P)} = G_{[P]}$, as a reduced polynomial may be recovered up to scale from its zero set, and in general $G_{Z(P)} \supseteq G_{[P]}$. Consider its singular set $Z(P)_{sing}$ **Ref where defined*** , which may be described as the zero set of the image of $P_{1,d-1}$ (which is essentially the exterior derivative dP). If $P = \sum_I a_I x^I$, where $a_{i_1,\dots,i_d}$ is symmetric in its lower indices, then $Z(P)_{sing} = \{[v] \in \mathbb{P}V^* \mid a_{i_1,i_2,\dots,i_d} x^{i_2}(v) \cdots x^{i_d}(v) = 0 \ \forall i_1\}$. While we could consider the singular locus of the singular locus etc.., it turns out to be easier to work with what I will call the *Jacobian loci*. For an arbitrary variety $X \subset \mathbb{P}V$, define $X_{Jac,1} := \{x \in \mathbb{P}V \mid dP_x = 0 \forall P \in I(X)\}$. If X is a hypersurface, then $X_{Jac,1} = X_{sing}$ but in general they can be different. Define $X_{Jac,k} := (X_{Jac,k-1})_{Jac,1}$. Algebraically, if $X = Z(P)$ for some $P \in S^dV$, then the ideal of $Z(P)_{Jac,k}$ is generated by the image of $P_{k,d-k} : S^kV^* \to S^{d-k}V$. The symmetry groups of these varieties all contain G_P.

8.13.2. The Fermat. This example follows [**CKW10**]. Let $\mathrm{fermat}^d_n := x_1^d + \cdots + x_n^d \in S^d\mathbb{C}^n$. The GL_n-orbit closure of $[\mathrm{fermat}^d_n]$ is the n-th secant variety of the Veronese variety $\sigma_n(v_d(\mathbb{P}^{n-1})) \subset \mathbb{P}S^d\mathbb{C}^n$. It is clear $\mathfrak{S}_n \subset G_{\mathrm{fermat}}$, as well as the diagonal matrices whose entries are d-th roots of unity. We need to see if there is anything else. The first idea, to look at the singular locus, does not work, as the zero set is smooth, so we consider $\mathrm{fermat}_{2,d-2} = x_1^2 \otimes x^{d-2} + \cdots + x_n^2 \otimes x^{d-2}$. Write the further polarization $P_{1,1,d-2}$ as a symmetric matrix whose entries are homogeneous polynomials of degree $d-2$ (the Hessian matrix). We get

$$\begin{pmatrix} x_1^{d-2} & & \\ & \ddots & \\ & & x_n^{d-2} \end{pmatrix}.$$

Were the determinant of this matrix $GL(V)$-invariant, we could proceed as we did with $e_{n,n}$, using unique factorization. Although it is not, it is close enough as follows:

following is out of place - refer back to it Recall that for a linear map $f : W \to V$, where $\dim W = \dim V = n$, we have $f^{\wedge n} \in \Lambda^n W^* \otimes \Lambda^n V$ and an element $(h, g) \in GL(W) \times GL(V)$ acts on $f^{\wedge n}$ by $(h, g) \cdot f^{\wedge n} = (\det(h))^{-1}(\det(g)) f^{\wedge n}$. In our case $W = V^*$ so $P^{\wedge n}_{2,d-2}(x) = \det(g)^2 P^{\wedge n}_{2,d-2}(g \cdot x)$, and the polynomial obtained by the determinant of the Hessian matrix is invariant up to scale.

Arguing as in ****, $\sum_j (g_1^{j_1} x_{j_1})^{d-2} \cdots (g_n^{j_n} x_{j_n})^{d-2} = x_1^{d-2} \cdots x_n^{d-2}$ and we conclude again by unique factorization that g is in $\mathfrak{S}_n \ltimes T_n$. Composing with a permutation matrix to make $g \in T$, we see that, by acting on the Fermat itself, that the entries on the diagonal are d-th roots of unity.

In summary:

Proposition 8.13.2.1. $G_{x_1^d+\cdots+x_n^d} = \mathfrak{S}_n \ltimes (\mathbb{Z}_d)^{\times n}$.

Exercise 8.13.2.2: (2) Show that the Fermat is characterized by its symmetries.

8.13.3. The sum-product polynomial. The following polynomial, called the *sum-product polynomial*, will be important when studying depth-3 circuits:

$$SP_r^n := \sum_{i=1}^{r} \Pi_{j=1}^n x_{ij} \in S^n(\mathbb{C}^{nr}).$$

Its $GL(rn)$-orbit closure is the r-th secant variety of the Chow variety $\sigma_r(Ch_n(\mathbb{C}^{nr}))$.

Exercise 8.13.3.1: (2)Determine $G_{SP_r^n}$ and show that SP_r^n is characterized by its symmetries.

8.13.4. Iterated matrix multiplication. Let $IMM_n^k \in S^n(\mathbb{C}^{k^2n})$ denote the iterated matrix multiplication operator for $k\times k$ matrices, $(X_1,\dots,X_n)\mapsto \operatorname{trace}(X_1\cdots X_n)$. Letting $V_j=\mathbb{C}^k$, invariantly

$$\begin{aligned} IMM_n^k = \mathrm{Id}_{V_1}\otimes\cdots\otimes\ \mathrm{Id}_{V_n} &\in (V_1\otimes V_2^*)\otimes(V_2\otimes V_3^*)\otimes\cdots\otimes(V_{n-1}\otimes V_n^*)\otimes(V_n\otimes V_1^*)\\ &\subset S^n((V_1\otimes V_2^*)\oplus(V_2\otimes V_3^*)\oplus\cdots\oplus(V_{n-1}\otimes V_n^*)\oplus(V_n\otimes V_1^*)), \end{aligned}$$

and the connected component of the identity of $G_{IMM_n^k}\subset GL(\mathbb{C}^{k^2n})$ is clear.

The case of IMM_n^3 is important as this sequence is complete for the complexity class $\mathbf{VP}_e$, of sequences of polynomials admitting small formulas, see [**BOC92**]. Moreover IMM_n^n is complete for the same complexity class as the determinant, namely **VQP**, see [**Blä01**].

add symmetry group in here from Fulvio

Problem 8.13.4.1. Find equations in the ideal of $\overline{GL_{9n}\cdot IMM_n^3}$. Determine lower bounds for the inclusions $\mathcal{P}erm_m\subset\overline{GL_{9n}\cdot IMM_n^3}$ and study common geometric properties (and differences) of $\mathcal{D}et_n$ and $\overline{GL_{9n}\cdot IMM_n^3}$.

8.13.5. The Pascal determinant. Let k be even, and let $A_j=\mathbb{C}^n$. Define the k-factor *Pascal determinant* $PD_{k,n}$ to be the unique up to scale element of $\Lambda^nA_1\otimes\cdots\otimes\Lambda^nA_k\subset S^n(A_1\otimes\cdots\otimes A_k)$. Choose the scale such that if $X=\sum x_{i_1,\dots,i_k}a_{1,i_1}\otimes\cdots\otimes a_{k,i_k}$ with $a_{\alpha,j}$ a basis of A_α, then

$$PD_{k,n}(X)=\sum_{\sigma_2,\dots,\sigma_k\in\mathfrak{S}_n}\operatorname{sgn}(\sigma_2\cdots\sigma_k)x_{1,\sigma_2(1),\dots,\sigma_k(1)}\cdots x_{n,\sigma_2(n),\dots,\sigma_k(n)}$$

By this expression we see, fixing k, that $(PD_{k,n})\in\mathbf{VNP}$.

Proposition 8.13.5.1 (Gurvits)**.** *The sequence* $(PD_{4,n})$ *is* **VNP** *complete.*

Proof. Set $x_{ijkl} = 0$ unless $i = j$ and $k = l$. Then $x_{i,\sigma_2(i),\sigma_3(i),\sigma_4(i)} = 0$ unless $\sigma_2(i) = i$ and $\sigma_3(i) = \sigma_4(i)$ so the only nonzero monomials are those where $\sigma_2 = \mathrm{Id}$ and $\sigma_3 = \sigma_4$, since the sign of σ_3 is squared, the result is the permanent. □

Thus we could just as well work with the sequence $PD_{4,n}$ as the permanent. Since $\Pi SL(A_j) \subset G_{PD_{4,n}}$, it resembles $G_{\det_n} = G_PD_{2,n}$.

It is clear the identity component of the stabilizer includes $SL_n^{\times k}/\mu_{n,k}$ where μ_n is as in §6.6.1, and a straight-forward Lie algebra calculation confirms this is the entire identity component. (Alternatively, one can use Dynkin's classification [**Dyn52**] of maximal subalgebras.) It is also clear that $\mathfrak{S}_k$ preserves $PD_{n,k}$ by permuting the factors.

Theorem 8.13.5.2 (Garibaldi, personal communication)**.** *For all k even*

$$G_{PD_{k,n}} = SL_n^{\times \cdots \times k}/\mu_{n,k} \rtimes \mathfrak{S}_k$$

Note that this includes the case of the determinant, and gives a new proof.

The result will follow from the following Lemma and Proposition 8.13.1.1.

Lemma 8.13.5.3. *[Garibaldi, personal communication] Let* $V = A_1 \otimes \cdots \otimes A_k$. *The normalizer of* $SL_n^{\times k}/\mu_n$ *in* $GL(V)$ *is* $GL_n^{\times k}/Z \rtimes \mathfrak{S}_k$, *where* Z *denotes the kernel of the product map* $(\mathbb{C}^*)^{\times k} \to \mathbb{C}^*$.

Proof of Lemma 8.13.5.3. We use Lemma 8.13.1.2. In our case, the Dynkin diagram for (Δ, λ) is

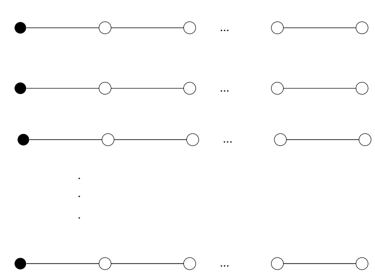


Figure 8.13.1. Marked Dynkin diagram for V

and $Aut(\Delta, \lambda)$ is clearly $\mathfrak{S}_k$. □

The theorem follows.

Chapter 9

The Chow variety of products of linear forms

In the GCT approach to Valiant's conjecture, one wants to understand the GL_{n^2}-module structure of $\mathbb{C}[\overline{GL_{n^2}\cdot \det_n}]$ via $\mathbb{C}[GL_{n^2}\cdot\det_n]$. In this chapter I discuss a "toy" problem that turns out to be deep, subtle and have surprising connections with several different areas of mathematices. Moreover, the orbit and orbit closures in question: $GL_n \cdot x_1\cdots x_n$ and $\overline{GL_n \cdot x_1\cdots x_n} = Ch_n(\mathbb{C}^n)$ are degenerations of the corresponding objects for the determinant, so information about them gives information about the determinant orbit closure.

This subject has a remarkable history beginning over 100 years ago, beginning with Brill, Gordan, Hermite and Hadamard. The history is rife with rediscoveries and errors that only make the subject more intriguing.

overview of chap here*

In this chapter I present two (possibly more) results that require a more advanced background in algebraic geometry. In §9.3 I present M. Brion's proof of the asymptotic surjectivity of the Hermite-Hadamard-Howe map. In §10.1 I present S. Kumar's proof of the non-normality of the determinant orbit closure.

9.1. The coordinate ring

I begin in with the GCT perspective:

9.1.1. Application of the algebraic Peter-Weyl theorem. Let $x_1, \dots, x_n \in V^*$ be a basis. Recall the symmetry group of $x_1 \cdots x_n$ from §4.2 is $\Gamma_n := T^{SL_n} \rtimes \mathfrak{S}_n$. Also recall that for any orbit, G/H, the algebraic Peter-Weyl theorem discussed in §8.6 implies $\mathbb{C}[G/H] = \oplus_{\lambda \in \Lambda_G^+} V_\lambda \otimes (V_\lambda^*)^H$, so we obtain

$$\mathbb{C}[GL(V) \cdot (x_1 \cdots x_n)] = \bigoplus_{\ell(\pi) \leq n} (S_\pi V)^{\oplus \dim (S_\pi V^*)^{\Gamma_n}},$$

where here $\pi = (p_1, \dots, p_n)$ with $p_j \in \mathbb{Z}$ satisfying $p_1 \geq p_2 \geq \cdots \geq p_n$ (i.e., π is not required to be a partition). We break up the determination of $(S_\pi V^*)^{\Gamma_n}$ into two problems: first determine the $T = T^{SL_n}$-invariants. By Exercise 8.1.5.4, these are the weight $(s, \dots, s) = (s^n)$ subspaces, so in particular $|\pi| = sn$ for some $s \in \mathbb{Z}$. Write this as $(S_\pi V^*)_0$, as these are the $\mathfrak{sl}(V)$-weight zero subspaces.

It remains to determine $(S_\pi V^*)_0^{\mathfrak{S}_n}$. This is not known. In the next subsection, I relate it to another quantity we don't know. Remarkably, this will enable us to get a satisfactory answer.

9.1.2. Plethysm and the double commutant theorem. Let $\mathfrak{S}_n \wr \mathfrak{S}_d \subset \mathfrak{S}_{dn}$ denote the *wreath product*, which, by definition, is the normalizer of $\mathfrak{S}_n^{\times d}$ in $\mathfrak{S}_{dn}$. It is the semi-direct product of $\mathfrak{S}_n^{\times d}$ with $\mathfrak{S}_d$, where $\mathfrak{S}_d$ acts by permuting the factors of $\mathfrak{S}_n^{\times d}$, see e.g., [**Mac95**, p 158]. The group $\mathfrak{S}_n \wr \mathfrak{S}_d$ acts on $V^{\otimes dn}$ by considering it as $(V^{\otimes n})^{\otimes d}$, d blocks of n-copies of V, permuting the n copies of V within each block as well as permuting the blocks. Thus $S^d(S^n V) = (V^{\otimes dn})^{\mathfrak{S}_n \wr \mathfrak{S}_d}$.

Since

$$(V^{\otimes dn})^{\mathfrak{S}_n \wr \mathfrak{S}_d} = (\bigoplus_{|\pi| = dn} [\pi] \otimes S_\pi V)^{\mathfrak{S}_n \wr \mathfrak{S}_d} = \bigoplus_{|\pi| = dn} [\pi]^{\mathfrak{S}_n \wr \mathfrak{S}_d} \otimes S_\pi V$$

we see, as long as $\dim V$ is sufficently large,

$$\mathrm{mult}(S_\pi V, S^d(S^n V)) = \dim [\pi]^{\mathfrak{S}_n \wr \mathfrak{S}_d}.$$

Unfortunately the action of $\mathfrak{S}_n \wr \mathfrak{S}_d$ is difficult to analyze.

Theorem 9.1.2.1. [**Gay76**] *Let μ be a partition of $\mathbf{v}\delta$ (so that $(S_\mu V)_0 \neq 0$). Suppose that the decomposition of $(S_\mu V)_0$ into irreducible $\mathcal{W}_V = \mathfrak{S}_{\mathbf{v}}$-modules is*

$$(S_\mu V)_0 = \bigoplus_{|\pi| = \mathbf{v}} [\pi]^{\oplus s_{\mu,\pi}}.$$

Then one has the decomposition of $GL(V)$-modules

$$S_\pi(S^\delta V) = \bigoplus_{|\mu| = \delta \mathbf{v}} (S_\mu V)^{\oplus s_{\mu,\pi}}.$$

In particular, for $\delta = 1$, i.e., $|\mu| = \mathbf{v}$, $(S_\mu V)_0 = [\mu]$.

Corollary 9.1.2.2. *Assume* $\dim V \geq d$. *Then*

$$\mathrm{mult}(S_\pi V, S^d(S^n V)) = \mathrm{mult}([d], (S_\pi \mathbb{C}^d)_0).$$

I prove the Corollary.

Proof of Cor. 9.1.2.2. Without loss of generality, assume $\dim V = d$. The $\mathcal{W}_V$-module decomposition of $S^d(S^nV)_0$ is $S^d(S^nV)_0 = \mathrm{Ind}_{\mathfrak{S}_n \wr \mathfrak{S}_d}^{\mathfrak{S}_{dn}} \mathrm{triv}$, where triv denotes the trivial $\mathfrak{S}_n \wr \mathfrak{S}_d$-module.

We have

$$\begin{aligned}\mathrm{mult}_{GL(V)}(S_\pi V, S^d(S^nV)) &= \mathrm{mult}_{\mathcal{W}}((S_\pi V)_0, (S^d(S^nV))_0)\\ &= \mathrm{mult}_{\mathcal{W}}((S_\pi V)_0, \mathrm{Ind}_{\mathfrak{S}_n \wr \mathfrak{S}_d}^{\mathfrak{S}_{dn}} \mathrm{triv}))\\ &= \dim (S_\pi V)_0^{\mathfrak{S}_n \wr \mathfrak{S}_d}\end{aligned}$$

the last line by *Frobenius reciprocity*: for finite groups $H \subset G$, an H-module W and a G-module U, $\mathrm{Hom}_{\mathbb{C}[H]}(W, U) = \mathrm{Hom}_{\mathbb{C}[G]}(\mathbb{C}[G] \otimes_{\mathbb{C}[H]} W, U)$, i.e., the multiplicity of U in $\mathrm{Ind}_H^G(W)$ is the multiplicity of W in $\mathrm{Res}_H^G(U)$. See, e.g. [**FH91**, §3.3]. **rest of proof???*** □

For a recent example of the state of the art, see [**CIM15**].

9.1.3. Back to the coordinate ring. Now specialize to the case of modules appearing in $Sym(S^nV)$. Corollary 9.1.2.2 says $\dim(S_\pi V)_0^{\mathfrak{S}_n} = \mathrm{mult}(S_\pi V, S^n(S^sV))$. If we consider all the π's together, we conclude

$$\mathbb{C}[GL(V) \cdot (x_1 \cdots x_n)]_{poly} = \bigoplus_s S^n(S^sV^*).$$

In particular, $\bigoplus_s S^n(S^sV^*)$ inherits a ring structure. We'll return to this in §9.3.1.

9.1.4. The Hermite-Hadamard-Howe map and the ideal of the Chow variety. After the modern perspective presented above, I now go back to the classical perspective of the nineteenth century. The two taken together give an interesting picture. The following linear map was first defined when $\dim V = 2$ by Hermite (1854), and in general independently by Hadamard (1897), Howe (1988), and Brion (1993).

Definition 9.1.4.1. The *Hermite-Hadamard-Howe map* $h_{d,n} : S^d(S^nV) \to S^n(S^dV)$ is defined as follows: First include $S^d(S^nV) \subset V^{\otimes nd}$. Next, reorder the copies of V from d blocks of n to n blocks of d and symmetrize the blocks of d to obtain an element of $(S^dV)^{\otimes n}$. Finally, thinking of S^dV as a single vector space, symmetrize the n blocks.

For example, putting subscripts on V to indicate position:

$$\begin{aligned} S^2(S^3V) \subset V^{\otimes 6} &= V_1 \otimes V_2 \otimes V_3 \otimes V_4 \otimes V_5 \otimes V_6 \\ &\to (V_1 \otimes V_4) \otimes (V_2 \otimes V_5) \otimes (V_3 \otimes V_6) \\ &\to S^2V \otimes S^2V \otimes S^2V \\ &\to S^3(S^2V) \end{aligned}$$

Note that $h_{d,n}$ is a $GL(V)$-module map.

Example 9.1.4.2. Here is $h_{2,2}((xy)^2)$:

$$\begin{aligned} (xy)^2 &= \frac{1}{4}[(x \otimes y + y \otimes x) \otimes (x \otimes y + y \otimes x)] \\ &= \frac{1}{4}[x \otimes y \otimes x \otimes y + x \otimes y \otimes y \otimes x + y \otimes x \otimes x \otimes y + y \otimes x \otimes y \otimes x] \\ &\mapsto \frac{1}{4}[x \otimes x \otimes y \otimes y + x \otimes y \otimes y \otimes x + y \otimes x \otimes x \otimes y + y \otimes y \otimes x \otimes x] \\ &\mapsto \frac{1}{4}[2(x^2) \otimes (y^2) + 2(xy) \otimes (xy)] \\ &\mapsto \frac{1}{2}[(x^2)(y^2) + (xy)(xy)]. \end{aligned}$$

Exercise 9.1.4.3: (1!) Show that $h_{d,n}((x_1)^n \cdots (x_d)^n) = (x_1 \cdots x_d)^n$.

Exercise 9.1.4.4: (2) Show that $h_{d,n} : S^d(S^nV) \to S^n(S^dV)$ is "self-dual" in the sense that $h^T_{d,n} = h_{n,d} : S^n(S^dV^*) \to S^d(S^nV^*)$. Conclude that $h_{d,n}$ surjective if and only if $h_{n,d}$ is injective.

Theorem 9.1.4.5 (Hadamard [**Had97**])**.** $\ker h_{d,n} = I_d(Ch_n(V^*))$.

Proof. Let $P \in S^d(S^nV)$. Since $Seg(v_n(\mathbb{P}V) \times \cdots \times v_n(\mathbb{P}V))$ spans $(S^nV)^{\otimes d}$, its projection to $S^d(S^nV)$ also spans, so we may write $P = \sum_j (x_{1j})^n \cdots (x_{dj})^n$ for some $x_{\alpha,j} \in V$. Let $\ell^1, \dots, \ell^n \in V^*$. Recall $\overline{P}$ is P considered as a linear form on $V^{*\otimes dn}$.

$$\begin{aligned} P(\ell^1 \cdots \ell^n) &= \langle \overline{P}, (\ell^1 \cdots \ell^n)^d \rangle \\ &= \sum_j \langle (x_{1j})^n \cdots (x_{dj})^n, (\ell^1 \cdots \ell^n)^d \rangle \\ &= \sum_j \langle (x_{1j})^n, (\ell^1 \cdots \ell^n) \rangle \cdots \langle (x_{dj})^n, (\ell^1 \cdots \ell^n) \rangle \\ &= \sum_j \Pi_{s=1}^n \Pi_{i=1}^d x_{ij}(\ell_s) \\ &= \sum_j \langle x_{1j} \cdots x_{dj}, (\ell^1)^d \rangle \cdots \langle x_{1j} \cdots x_{dj}, (\ell^n)^d \rangle \\ &= \langle \overline{h_{d,n}(P)}, (\ell^1)^d \cdots (\ell^n)^d \rangle \end{aligned}$$

If $h_{d,n}(P)$ is nonzero, there will be some monomial of the form $(\ell^1)^d \cdots (\ell^n)^d$ it will pair with to be nonzero (again, using the spanning property). On the other hand, if $h_{d,n}(P) = 0$, then P annihilates all points of $Ch_n(V^*)$. $\square$

Exercise 9.1.4.6: (1) Show that if $h_{d,n} : S^d(S^n\mathbb{C}^m) \to S^n(S^d\mathbb{C}^m)$ is not surjective, then $h_{d,n} : S^d(S^n\mathbb{C}^k) \to S^n(S^d\mathbb{C}^k)$ is not surjective for all $k > m$, and that the partitions describing the kernel are the same in both cases if $d \leq m$. ⊚

Exercise 9.1.4.7: (1) Show that if $h_{d,n} : S^d(S^n\mathbb{C}^m) \to S^n(S^d\mathbb{C}^m)$ is surjective, then $h_{d,n} : S^d(S^n\mathbb{C}^k) \to S^n(S^d\mathbb{C}^k)$ is surjective for all $k < m$.

Example 9.1.4.8 (The case $\dim V = 2$)**.** When $\dim V = 2$, every polynomial decomposes as a product of linear factors, so the ideal of $Ch_n(\mathbb{C}^2)$ is zero. We recover the following theorem of Hermite:

Theorem 9.1.4.9 (Hermite reciprocity)**.** *The map $h_{d,n} : S^d(S^n\mathbb{C}^2) \to S^n(S^d\mathbb{C}^2)$ is an isomorphism for all d, n. In particular $S^d(S^n\mathbb{C}^2)$ and $S^n(S^d\mathbb{C}^2)$ are isomorphic GL_2-modules.*

Often in modern textbooks (e.g., [**FH91**]) only the "In particular" is stated.

Originally Hadamard thought the maps $h_{d,n}$ were always of maximal rank, but later he realized he did not have a proof. In [**Had99**] he did prove:

Theorem 9.1.4.10 (Hadamard [**Had99**])**.** *The map $h_{3,3} : S^3(S^3V) \to S^3(S^3V)$ is an isomorphism.*

In the same paper, he posed the question:

Question 9.1.4.11. Is $h_{d,n}$ always of maximal rank?

Howe [**How87**] also investigated the map $h_{d,n}$ and wrote "it is reasonable to expect" that $h_{d,n}$ is always of maximal rank.

Proof of Theorem 9.1.4.10. Without loss of generality, assume $\mathbf{w} = 3$ and $x_1, x_2, x_3 \in V^*$ are a basis. Say we had $P \in I_3(Ch_3(V^*))$. Consider P restricted to the line in S^3V^* spanned by $x_1^3 + x_2^3 + x_3^3$ and $x_1x_2x_3$. Write $P(\mu(x_1^3 + x_2^3 + x_3^3) - \lambda x_1x_2x_3)$ as a cubic polynomial on $\mathbb{P}^1$ with coordinates $[\mu, \lambda]$. Note that $P(\mu, \nu)$ vanishes at the four points $[0, 1], [1, 3], [1, 3\omega], [1, 3\omega^2]$ where ω is a primitive third root of unity. A cubic polynomial on $\mathbb{P}^1$ vanishing at four points is identically zero, so the whole line is contained in $Z(P)$. In particular, $P(1, 0) = 0$, i.e., P vanishes on $x_1^3 + x_2^3 + x_3^3$. Hence it must vanish identically on $\sigma_3(v_3(\mathbb{P}^2))$. But $I_3(\sigma_3(v_3(\mathbb{P}^2))) = 0$, see, e.g., Corollary 8.3.4.3, (In fact $\sigma_3(v_3(\mathbb{P}^2)) \subset \mathbb{P}S^3\mathbb{C}^3$ is a hypersurface of degree four.) $\square$

Remark 9.1.4.12. The above proof is due to A. Abdesselam (personal communication). It is a variant of Hadamard's original proof, where instead of $x_1^3+x_2^3+x_3^3$ one uses an arbitrary cubic f, and generalizing $x_1x_2x_3$ one uses the Hessian $H(f)$. Then the curves $f=0$ and $H(f)=0$ intersect in 9 points (the nine flexes of $f=0$) and there are four groups of three lines going through these points, i.e. four places where the polynomial becomes a product of linear forms.

Theorem 9.1.4.13. [**BL89**] *(also see* [**McK08**, Thm. 8.1] *and* [**Ike15**]*) If $h_{d,n}$ is surjective, then $h_{d',n}$ is surjective for all $d'>d$. In other words, if $h_{n,d}$ is injective, then $h_{n,d'}$ is injective for all $d'>d$.*

Outline of proof. I follow the proof in [**Ike15**]. Write $V=E\oplus F$ with $\dim E=d$ and $\dim F=n$. Give E a basis $e_1,\dots,e_d$ and F a basis $f_1,\dots,f_n$ inducing a basis of V ordered $(e_1,e_2,\dots,f_n)$. Write $(V^{\otimes dn})_{(\alpha,\beta)}$ for the $\alpha=(a_1,\dots,a_d),\beta=(b_1,\dots,b_n)$ $GL(E)\times GL(F)$-weight space. Define the lowering map $\phi_{i,j}:(V^{\otimes dn})_{(\alpha,\beta)}\to(V^{\otimes dn})_{(a_1,\dots,a_{i-1},(a_i-1),a_{i+1},\dots,a_d),\beta=(b_1,\dots,(b_j+1),\dots,b_n)}$ induced from the map $V\to V$ that sends e_i to f_j and maps all other basis vectors to themselves. It is straight-forward to see the $\phi_{i,j}$ commute. Let $\phi_{d\times n}:(V^{\otimes dn})_{(n^d,(0))}\to(V^{\otimes dn})_{((0),d^n)}$ denote the composition of $\phi_{1,1}\cdots\phi_{d,b}$ restricted to $(V^{\otimes dn})_{(n^d,(0))}$. To see $h_{d,n}:S^d(S^n\mathbb{C}^N)\to S^n(S^d\mathbb{C}^N)$ is injective, it is sufficient to see it is injective on each irreducible submodule, in fact on the weight zero subspace of each irreducible submodule when $N=d$. By Gay's theorem 9.1.2.2 this is $(V^{\otimes dn})^{\mathcal{W}_E}_{(n^d,(0))}=(E^{\otimes dn})_0^{\mathcal{W}_E}$, where $\mathcal{W}=\mathfrak{S}_d$ is the Weyl group. ???*** proof**??? We need to show $\phi_{d\times(n-1)}$ injective implies $\phi_{d\times n}$ is injective.

Reorder and decompose

$$\phi_{d\times n}=[\phi_{1,1}\cdots\phi_{1,n-1}\phi_{2,1}\cdots\phi_{d,n-1}][\phi_{1,n}\cdots\phi_{d,n}]$$

and call the first term the left factor and the second the right factor. The injectivity of each term in the left factor follows from a straight-forward induction argument. It remains to show injectivity of each $\phi_{i,n}$, in fact injectivity of $\phi_{i,n}$ restricted to each $(((n-1)^{i-1},n^{d-i}),(0^{n-1},i-1))$ weight space. Each of these restrictions just deals with a rasing operator in the $\mathbb{C}^2$ with basis e_i,f_n, so we need to see the lowering map $((\mathbb{C}^2)^{\otimes n+i-1})_{(n,i-1)}\to((\mathbb{C}^2)^{\otimes n+i-1})_{(n-1,i)}$ is injective. Decompose

$$(\mathbb{C}^2)^{\otimes n+i-1}=\oplus_{\lambda_2=0}^{\lfloor\frac{n+i-1}{2}\rfloor}S_{n+i-1-\lambda_2,\lambda_2}\mathbb{C}^2.$$

The weight $(n-1,i)$ vector in each space may be written as $(e_i\wedge f_n)^{\otimes\lambda_2}\otimes(e_i^{n-\lambda_2}f_n^{i-1-\lambda_2})$. The lowering operator is zero on the first factor so this vector maps to $(e_i\wedge f_n)^{\otimes\lambda_2}\otimes(e_i^{n-\lambda_2-1}f_n^{i-\lambda_2})$ which is nonzero.

$\square$

Remark 9.1.4.14. The statements and proofs in [**BL89, McK08**] were regarding the map $h_{d,n:0}$ defined in §9.1.5 below.

Theorem 9.1.4.15. [**MN05**] *The map $h_{5,5}$ is not surjective.*

Remark 9.1.4.16. In [**MN05**] they showed the map $h_{5,5:0}$ defined in §9.1.5 below is not injective. A. Abdessalem realized their computation showed the map $h_{5,5}$ is not injective and pointed this out to them. Evidently there was some miscommunication because in [**MN05**] they mistakenly say the result comes from [**Bri02**] rather than their own paper.

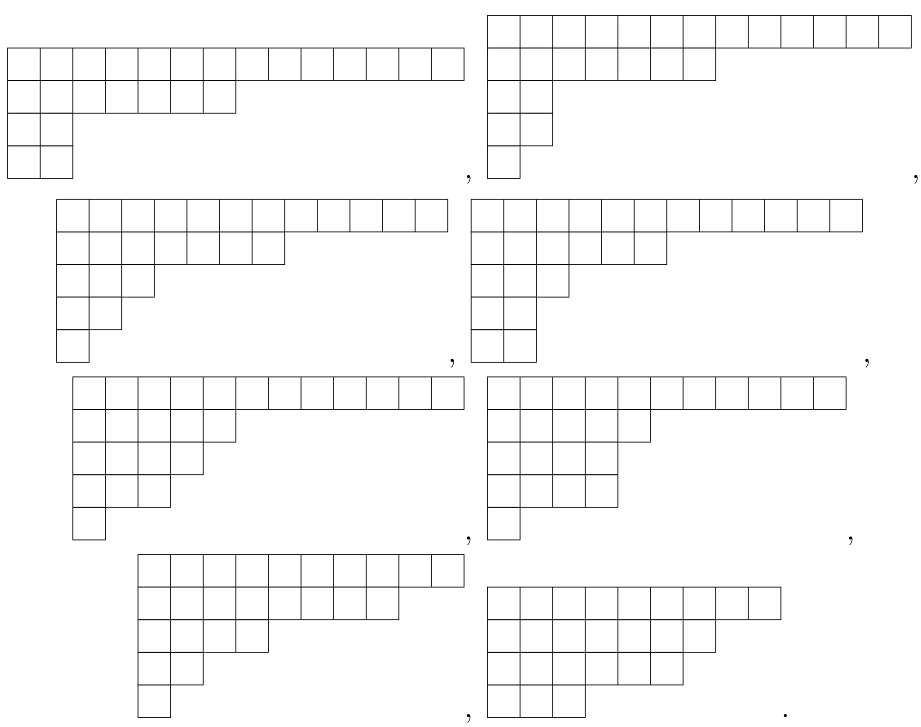
The $GL(V)$-module structure of the kernel of $h_{5,5}$ was determined by M-W Cheung, C. Ikenmeyer and S. Mkrtchyan as part of a 2012 AMS MRC program:

Proposition 9.1.4.17. [**CIM15**] *The kernel of $h_{5,5} : S^5(S^5\mathbb{C}^5) \to S^5(S^5\mathbb{C}^5)$ consists of irreducible modules corresponding to the following partitions:*

$$\{(14,7,2,2),(13,7,2,2,1),(12,7,3,2,1),(12,6,3,2,2),$$
$$(12,5,4,3,1),(11,5,4,4,1),(10,8,4,2,1),(9,7,6,3)\}.$$

All these occur with multiplicity one in the kernel, but not all occur with multiplicity one in $S^5(S^5\mathbb{C}^5)$. In particular, the kernel is not an isotypic component.

The Young diagrams of the kernel of $h_{5,5}$ are:



While the Hermite-Hadamard-Howe map is not always of maximal rank, it is "eventually" of maximal rank:

Theorem 9.1.4.18. [**Bri93, Bri97**] *The Hermite-Hadamard-Howe map*

$$h_{d,n} : S^d(S^nV^*) \to S^n(S^dV^*)$$

is surjective for d sufficiently large, in fact for $d \sim\geq n^2\binom{n+d}{d}$

I present the proof of Theorem 9.1.4.18 in §9.3.1.

Problem 9.1.4.19 (The Hadamard-Howe Problem)**.** Determine the function $d(n)$ such that $h_{d,n}$ is surjective for all $d \geq d(n)$.

A more ambitious problem would be:

Problem 9.1.4.20. Determine the kernel of $h_{d,n}$.

A less ambitious problem is as follows: when n is even, the module $S_{n^d}\mathbb{C}^n$ occurs in $S^d(S^n\mathbb{C}^n)$ with multiplicity one.

Conjecture 9.1.4.21 (Kumar [**Kum**])**.** *Let n be even, then for all $d \leq n$, $S_{n^d}\mathbb{C}^n \not\subset \ker h_{d,n}$, i.e., $S_{n^d}\mathbb{C}^n \subset \mathbb{C}[Ch_n(\mathbb{C}^n)]$.*

I discuss Conjecture 9.1.4.21 in §9.2. It turns out to be equivalent to a famous conjecture in combinatorics.

9.1.5. $\mathfrak{S}_{dn}$-formulation of the Hadamard-Howe problem. The dimension of V, as long as it is at least d, is irrelevant for the $GL(V)$-module structure of the kernel of $h_{d,n}$. *In this section assume* $\dim V = dn$.

If one restricts $h_{d,n}$ to the $\mathfrak{sl}(V)$-weight zero subspace, one obtains a $\mathcal{W}_V$-module map

$$h_{d,n:0} : S^d(S^nV)_0 \to S^n(S^dV)_0. \tag{9.1.1}$$

In other words, recalling the discussion in §9.1.2, as a $\mathcal{W}_V = \mathfrak{S}_{dn}$-module map, (9.1.1) is

$$h_{d,n:0} : \mathrm{Ind}_{\mathfrak{S}_n \wr \mathfrak{S}_d}^{\mathfrak{S}_{dn}} \mathrm{triv} \to \mathrm{Ind}_{\mathfrak{S}_d \wr \mathfrak{S}_n}^{\mathfrak{S}_{dn}} \mathrm{triv}\,. \tag{9.1.2}$$

Call $h_{d,n:0}$ the *Black-List map.* Moreover, since every irreducible module appearing in $S^d(S^nV)$ has a non-zero weight zero subspace, $h_{d,n}$ is the unique $GL(V)$-module extension of $h_{d,n:0}$.

The above discussion shows that one can deduce the kernel of $h_{d,n}$ from that of $h_{d,n:0}$ and vice versa. In particular, one is injective if and only if the other is.

The map $h_{d,n:0}$ was defined purely in terms of combinatorics in [**BL89**] as a path to try to prove the following conjecture of Foulkes:

Conjecture 9.1.5.1. [**Fou50**] *Let $d > n$, let π be a partition of dn and let $[\pi]$ denote the corresponding $\mathfrak{S}_{dn}$-module. Then,*

$$\operatorname{mult}([\pi], \operatorname{Ind}_{\mathfrak{S}_n \wr \mathfrak{S}_d}^{\mathfrak{S}_{dn}} \operatorname{triv}) \geq \operatorname{mult}([\pi], \operatorname{Ind}_{\mathfrak{S}_d \wr \mathfrak{S}_n}^{\mathfrak{S}_{dn}} \operatorname{triv}).$$

Equivalently,

$$\operatorname{mult}(S_\pi V, S^d(S^n V)) \geq \operatorname{mult}(S_\pi V, S^n(S^d V)). \tag{9.1.3}$$

Theorem 8.10.1.1 shows that equality holds asymptotically in (9.1.3). Conjecture 9.1.5.1 is still open in general.

9.1.6. Brill's equations. Set theoretic equations of $Ch_d(V)$ have been known since 1894. Here is a modern presentation elaborating the presentation in [**Lan12**, §8.6], which was suggested by E. Briand.

Our goal is a polynomial test to see if $f \in S^d V$ is a product of linear factors. We can first try to see just if P is divisible by a power of a linear form. The discussion in §8.4.2 will not be helpful as the conditions there are vacuous when $n - m = 1$. We could proceed as in §6.5.1 and check if $\ell x^{I_1} \wedge \cdots \wedge \ell x^{I_D} \wedge f = 0$ where the x^{I_j} are a basis of $S^{d-1}V$, but in this case there is a simpler test to see if a given linear form ℓ divides f:

Consider the map $\pi_{d,d} : S^d V \otimes S^d V \to S_{d,d} V$ obtained by projection. (By the Pieri rule 8.1.3.1, $S_{d,d}V \subset S^d V \otimes S^d V$ with multiplicity one.)

Lemma 9.1.6.1. *Let $\ell \in V$, $f \in S^d V$. Then $f = \ell h$ for some $h \in S^{d-1}V$ if and only if $\pi_{d,d}(f \otimes \ell^d) = 0$.*

Proof. Since $\pi_{d,d,}$ is linear, it suffices to prove the lemma when $f = \ell_1 \cdots \ell_d$. In that case $\pi_{d,d}(f \otimes \ell^d)$, up to a constant, is $(\ell_1 \wedge \ell) \cdots (\ell_d \wedge \ell)$. □

We would like a map that sends $\ell_1 \cdots \ell_d$ to $\sum_j \ell_j^d \otimes stuff_j$, as then we could apply $\pi_{d,d} \otimes \operatorname{Id}_{stuff}$ to f tensored with the result of our desired map to obtain our equations.

While it is not obvious how to obtain such a map for powers, there is an easy way to get elementary symmetric functions, namely the maps $f \mapsto f_{j,d-j}$ because $(\ell_1 \cdots \ell_d)_{j,d-j} = \sum_{|K|=j} \ell_K \otimes \ell_{K^c}$ where $\ell_K = \ell_{k_1} \cdots \ell_{k_j}$ and K^c denotes the complementary index set in $[d]$. We can try to convert this to power sums by the conversion formula obtained from the relation between generating functions (6.1.5):

$$p_d = \mathcal{P}_d(e_1, \dots, e_d) := \det \begin{pmatrix} e_1 & 1 & 0 & \cdots & 0 \\ 2e_2 & e_1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ de_d & e_{d-1} & e_{d-2} & \cdots & e_1 \end{pmatrix}. \tag{9.1.4}$$

The desired term comes from the diagonal e_1^d and the rest of the terms kill off the unwanted terms of e_1^d. This idea almost works- the only problem is that our naïve correction terms have the wrong degree on the right hand side. For example, when $d=3$, naïvely using $p_3 = e_1^3 - 3e_1e_2 + 3e_3$ would give, for the first term, degree $6 = 2+2+2$ on the right hand side of the tensor product, the second degree $3 = 2+1$ and the third degree zero. In general, the right hand side of the e_1^d term would have degree $(d-1)^d$, whereas the de_d term would have degree zero. In addition to fixing the degree mismatch, we need to formalize how we will treat the right hand sides.

To these ends, recall that for any two algebras $\mathcal{A}, \mathcal{B}$, one can give $\mathcal{A}\otimes\mathcal{B}$ the structure of an algebra by defining $(\alpha\otimes\beta)\cdot(\alpha'\otimes\beta') := \alpha\alpha'\otimes\beta\beta'$ and extending linearly. Give $Sym(V)\otimes Sym(V)$ this algebra structure. Define maps

$$
\begin{aligned}
E_j : S^\delta V &\to S^j V \otimes S^{\delta-1} V \\
f &\mapsto f_{j,\delta-j}\cdot(1\otimes f^{j-1}).
\end{aligned}
\tag{9.1.5}
$$

The $(1\otimes f^{j-1})$ fixes our degree problem. If $j>\delta$ define $E_j(f)=0$.

Our desired map is

$$
\begin{aligned}
Q_d : S^d V &\to S^d V \otimes S^{d(d-1)} V \\
f &\mapsto \mathcal{P}_d(E_1(f),\dots,E_d(f)).
\end{aligned}
\tag{9.1.6}
$$

Theorem 9.1.6.2 (Brill [**Bri93**], Gordan [**Gor94**], Gelfand-Kapranov-Zelevinski [**GKZ94**], Briand [**Bri10**])**.** *Consider the map*

$$
\mathcal{B} : S^d V \to S_{d,d} V \otimes S^{d^2-d} V \tag{9.1.7}
$$

$$
f \mapsto (\pi_{d,d}\otimes \mathrm{Id}_{S^{d^2-d}V})[f\otimes Q_d(f)]. \tag{9.1.8}
$$

Then $[f]\in Ch_d(V)$ if and only if $\mathcal{B}(f)=0$.

The proof will be by induction, that will require a generalization of Q_d. Define

$$
\begin{aligned}
Q_{d,\delta} : S^\delta V &\to S^d V \otimes S^{d(\delta-1)} V \\
f &\mapsto \mathcal{P}_d(E_1(f),\dots,E_d(f)).
\end{aligned}
\tag{9.1.9}
$$

Lemma 9.1.6.3. *If $f_1 \in S^\delta V$ and $f_2 \in S^{d'-\delta}V$, then*

$$
Q_{d,d'}(f_1f_2) = (1\otimes f_1^d)\cdot Q_{d,d'-\delta}(f_2) + (1\otimes f_2^d)\cdot Q_{d,\delta}(f_1).
$$

Assume Lemma 9.1.6.3 for the moment.

Proof of Theorem 9.1.6.2. Say $f = \ell_1\cdots\ell_d$. First note that for $\ell\in V$, $E_j(\ell^j) = \ell^j\otimes\ell^{j-1}$ and $Q_{d,1}(\ell) = \ell^d\otimes 1$. Next, compute $E_1(\ell_1\ell_2) = \ell_1\otimes\ell_2 +$

$\ell_2 \otimes \ell_1$ and $E_2(\ell_1\ell_2) = \ell_1\ell_2 \otimes \ell_1\ell_2$, so $Q_{2,2}(\ell_1\ell_2) = \ell_1^2 \otimes \ell_2^2 + \ell_2^2 \otimes \ell_1^2$. By induction and Lemma 9.1.6.3,

$$Q_{d,\delta}(\ell_1 \cdots \ell_\delta) = \sum_j \ell_j^d \otimes (\ell_1^d \cdots \ell_{j-1}^d \ell_{j+1}^d \cdots \ell_\delta^d).$$

We conclude $Q_d(f) = \sum_j \ell_j^d \otimes (\ell_1^d \cdots \ell_{j-1}^d \ell_{j+1}^d \cdots \ell_d^d)$ and $\pi_{d,d}(\ell_1 \cdots \ell_d, \ell_j^d) = 0$ for each j by Lemma 9.1.6.1.

For the other direction, first assume f is reduced, i.e., has no repeated factors. Let $z \in \mathrm{Zeros}(f)_{smooth}$, then $Q_d(f) = (E_1(f))^d + \sum \mu_j \otimes \psi_j$ where $\psi_j \in S^{d^2-d}V$, $\mu_j \in S^d V$ and f divides ψ_j for each j because $E_1(f)^d$ occurs as a monomial in the determinant (9.1.4) and all the other terms contain an $E_j(f)$ with $j > 1$, and so are divisible by f.

Thus $\mathcal{B}(f)(\cdot, z) = \pi_{d,d}(f \otimes (df_z)^d)$ because $E_1(f)^d = (f_{1,d-1})^d$ and $f_{1,d-1}(\cdot, z) = df_z$, and all the $\psi_j(z)$ are zero. By Lemma 9.1.6.1, df_z divides f for all $z \in \mathrm{Zeros}(f)$. But this implies the tangent space to f is constant in a neighborhood of z, i.e., that the component containing z is a linear space, and since every component of $\mathrm{Zeros}(f)$ contains a smooth point, $\mathrm{Zeros}(f)$ is a union of hyperplanes, which is what we set out to prove.

Finally, say $f = g^k h$ where g is irreducible of degree q and h is of degree $d - qk$ and is relatively prime to g. Apply Lemma 9.1.6.3:

$$Q_d(g(g^{k-1}h)) = (1 \otimes g^d) \cdot Q_{d,d-q}(g^{k-1}h) + (1 \otimes (g^{k-1}h)^d) \cdot Q_{d,q}(g).$$

A second application gives

$$Q_d(g^k h) = (1 \otimes g^d) \cdot [(1 \otimes g^d) \cdot Q_{d,d-2q}(g^{k-2}h) + (1 \otimes (g^{k-2}h)^d) \cdot Q_{d,q}(g) + (1 \otimes (g^{k-2}h)^d) \cdot Q_{d,q}(g)].$$

After $k-1$ applications one obtains:

$$Q_d(g^k h) = (1 \otimes g^{d(k-1)}) \cdot [k(1 \otimes h^d) \cdot Q_{d,q}(g) + (1 \otimes g^d) \cdot Q_{d,d-qk}(h)]$$

and $(1 \otimes g^{d(k-1)})$ will also factor out of $\mathcal{B}(f)$. Since $\mathcal{B}(f)$ is identically zero but $g^{d(k-1)}$ is not, we conclude

$$0 = \pi_{d,d} \otimes \mathrm{Id}_{S^{d^2-d}V}\, f \otimes [k(1 \otimes h^d) \cdot Q_{d,q}(g) + (1 \otimes g^d) \cdot Q_{d,d-qk}(h)]$$

Let $w \in \mathrm{Zeros}(g)$ be a general point, so in particular $h(w) \neq 0$. Evaluating at (z, w) with z arbitrary gives zero on the second term and the first implies $\pi_{d,d} \otimes \mathrm{Id}_{S^{d^2-d}V}(f \otimes Q_{d,q}(g)) = 0$ which implies dg_w divides g, so g is a linear form. □

Proof of Lemma 9.1.6.3. Define, for $u \in Sym(V) \otimes Sym(V)$,

$$\Delta_u : Sym(V) \to Sym(V) \otimes Sym(V)$$
$$f \mapsto \sum_j u^j \cdot f_{j,\deg(f)-j}.$$

Exercise 9.1.6.4: Show that $\Delta_u(fg) = (\Delta_u f)\cdot(\Delta_u g)$, and that the generating series for the $E_j(f)$ may be written as

$$\mathcal{E}_f(t) = \frac{1}{1\otimes f}\cdot \Delta_{t(1\otimes f)} f.$$

Note that $(1\otimes f)^{\cdot s} = 1\otimes f^s$ and $(1\otimes fg) = (1\otimes f)\cdot(1\otimes g)$. Thus

$$\mathcal{E}_{fg}(t) = [\frac{1}{1\otimes f}\cdot \Delta_{[t(1\otimes g)](1\otimes f)}(f)]\cdot[\frac{1}{1\otimes g}\cdot \Delta_{[t(1\otimes f)](1\otimes g)}(g)],$$

and taking the logarithmic derivative (recalling Equation (6.1.5)) we conclude. □

Remark 9.1.6.5. There was a gap in the argument in [**Gor94**], repeated in [**GKZ94**], when proving the "only if" part of the argument. They assumed that the zero set of f contains a smooth point, i.e., that the differential of f is not identically zero. This gap was fixed in [**Bri10**]. In [**GKZ94**] they use $G_0(d, \dim V)$ to denote $Ch_d(V)$.

9.1.7. Brill's equations as modules.

Brill's equations are of degree $d+1$ on S^dV^*. (The total degree of $S_{d,d}V\otimes S^{d^2-d}V$ is $d(d+1)$ which is the total degree of $S^{d+1}(S^dV)$.) Consider the $GL(V)$-module map

$$S_{dd}V\otimes S^{d^2-d}V \to S^{d+1}(S^dV)$$

given by Brill's equations. The components of the target are not known in general and the set of modules present grows extremely fast. One can use the Pieri formula 8.1.3.1 to get the components of the first. Using the Pieri formula, we conclude:

Proposition 9.1.7.1. *As a $GL(V)$-module, Brill's equations are multiplicity free.*

Exercise 9.1.7.2: Write out the decomposition and show that only partitions with three parts appear as modules in Brill's equations. ⊚

Remark 9.1.7.3. If $d < \mathbf{v} = \dim V$, then $Ch_d(V) \subset Sub_d(S^dV)$ so $I(Ch_d(V)) \supset \Lambda^{d+1}V^*\otimes\Lambda^{d+1}(S^{d-1}V^*)$. J. Weyman (in unpublished notes from 1994) observed that these equations are not in the ideal generated by Brill's equations. More precisely, the ideal generated by Brill's equations does not include modules $S_\pi V^*$ with $\ell(\pi) > 3$, so it does not cut out $Ch_d(V)$ scheme theoretically when $d < \mathbf{v}$. By Theorem 9.1.4.15 the same holds for $Ch_5(\mathbb{C}^5)$ and almost certainly holds for all $Ch_n(\mathbb{C}^n)$ with $n \geq 5$.

Problem 9.1.7.4. What is the kernel of $Brill : S_{n,n}V\otimes S^{n^2-n}V \to S^{n+1}(S^nV)$?

9.2. Conjecture 9.1.4.21 and a conjecture in combinatorics

Let $P \in S_{n^d}(\mathbb{C}^d) \subset S^d(S^n\mathbb{C}^d)$ be non-zero. Conjecture 9.1.4.21 may be stated as $P((x_1 \cdots x_n)^d) \neq 0$. Our first task is to obtain an expression for P.

9.2.1. Realization of the module. Let $V = \mathbb{C}^d$. For any even n, the one-dimensional module $S_{(n^d)}V$ occurs with multiplicity one in $S^d(S^nV)$ (cf. [**How87**, Prop. 4.3]). Fix a volume form on V so that $\det_d \in S^dV$ is well defined.

Proposition 9.2.1.1. [?] *Let n be even. The unique (up to scale) polynomial $P \in S_{(n^d)}V \subset S^d(S^nV)$ evaluates on*

$$x = (v_1^1 \cdots v_n^1)(v_1^2 \cdots v_n^2) \cdots (v_1^d \cdots v_n^d) \in S^d(S^nV^*), \ \ \textit{for any } v_j^i \in V^*,$$

to give

(9.2.1)

$$\langle P, x\rangle = \sum_{\sigma_1,\dots,\sigma_d \in \mathfrak{S}_n} \det{}_d(v^1_{\sigma_1(1)}, \dots, v^d_{\sigma_d(1)}) \cdots \det{}_d(v^1_{\sigma_1(n)}, \dots, v^d_{\sigma_d(n)}).$$

Proof. Let $\bar{P} \in (V)^{\otimes nd}$ be defined by the identity (9.2.1) (with P replaced by $\bar{P}$). It suffices to check that

(i) $\bar{P} \in S^d(S^nV)$,

(ii) $\bar{P}$ is $SL(V)$ invariant, and

(iii) $\bar{P}$ is not identically zero.

Observe that (iii) follows from the identity (9.2.1) by taking $v_j^i = x_i$ where $x_1, \dots, x_d$ is a basis of V^*, and (ii) follows because $SL(V)$ acts trivially on $\det_d$.

To see (i), we show (ia) $\bar{P} \in S^d((V)^{\otimes n})$ and (ib) $\bar{P} \in (S^nV)^{\otimes d}$ to conclude. To see (ia), it is sufficient to show that exchanging two adjacent factors in parentheses in the expression of x will not change (9.2.1). Exchange v_j^1 with v_j^2 in the expression for $j = 1, \dots, n$. Then, each individual determinant will change sign, but there are an even number of determinants, so the right hand side of (9.2.1) is unchanged. To see (ib), it is sufficient to show the expression is unchanged if we swap v_1^1 with v_2^1 in (9.2.1). If we multiply by $n!$, we may assume $\sigma_1 = \mathrm{Id}$, i.e.,

$$\langle \bar{P}, x\rangle =$$

$$n! \sum_{\sigma_2,\dots,\sigma_d \in \mathfrak{S}_n} \det{}_d(v_1^1, v^2_{\sigma_2(1)}, \dots, v^d_{\sigma_d(1)}) \det{}_d(v_2^1, v^2_{\sigma_2(2)}, \dots, v^d_{\sigma_d(2)}) \cdots \det{}_d(v_n^1, v^2_{\sigma_2(n)}, \dots, v^d_{\sigma_d(n)}).$$

With the two elements v_1^1 and v_2^1 swapped, we get

(9.2.2)
$$n! \sum_{\sigma_2,\dots,\sigma_d \in \mathfrak{S}_n} \det{}_d(v_2^1, v_{\sigma_2(1)}^2, \dots, v_{\sigma_d(1)}^d) \det{}_d(v_1^1, v_{\sigma_2(2)}^2, \dots, v_{\sigma_d(2)}^d) \cdots \det{}_d(v_n^1, v_{\sigma_2(n)}^2, \dots, v_{\sigma_d(n)}^d).$$

Now right compose each σ_s in (9.2.2) by the transposition $(1,2)$. The expressions become the same. □

Corollary 9.2.1.2. *The unique (up to scale) polynomial $P \in S_{(n^d)}V \subset S^d(S^nV)$ when n is even, is nonzero on $(y_1)^n + \cdots + (y_d)^n$ if the y_j are linearly independent. In particular, $S_{n^d}V \subset \mathbb{C}[\sigma_d(v_n(\mathbb{PC}^N))]$ for all $N \geq d$.*

Proof. The monomial $(y_1)^n \cdots (y_d)^n$ appears in $((y_1)^n + \cdots + (y_d)^n)$ and all other monomials appearing pair with P to be zero. □

Now specialize to the case $d = n$ (this is the critical case) and evaluate on $(x_1 \cdots x_n)^n$, where $x_1, \dots, x_n$ is a uni-modular basis of V^*.

(9.2.3)
$$\langle P, (x_1 \cdots x_n)^n \rangle = \sum_{\sigma_1,\dots,\sigma_n \in \mathfrak{S}_n} \det{}_d(x_{\sigma_1(1)}, \dots, x_{\sigma_n(1)}) \cdots \det{}_d(x_{\sigma_1(n)}, \dots, x_{\sigma_n(n)}).$$

For a fixed $(\sigma_1, \dots, \sigma_n)$ the contribution will either be $0, 1$ or -1. The contribution is zero unless for each j, the indices $\sigma_1(j), \dots, \sigma_n(j)$ are distinct. Arrange these numbers in an array:

$$\begin{pmatrix} \sigma_1(1) & \cdots & \sigma_n(1) \\ & \vdots & \\ \sigma_1(n) & \cdots & \sigma_n(n) \end{pmatrix}$$

The contribution is zero unless the array is a *Latin square*, i.e., an $n \times n$ matrix such that each row and column consists of the integers $\{1, \dots, n\}$. If it is a Latin square, the rows correspond to permutations, and the contribution of the term is the product of the signs of these permutations. Call this the *row sign* of the Latin square. There is a famous conjecture in combinatorics regarding the products of both the signs of the row permutations and the column permutations, called the *sign* of the Latin square:

Conjecture 9.2.1.3 (Alon-Tarsi [**AT92**]). *Let n be even. The number of sign -1 Latin squares of size n is not equal to the number of sign $+1$ Latin squares of size n.*

Conjecture 9.2.1.3 is known to be true when $n = p \pm 1$, where p is an odd prime; in particular, it is known to be true up to $n = 24$ [**Gly10, Dri97**].

On the other hand, in [**Alp14, CW16**] they show that the ratio of the number of sign -1 Latin squares of size n to the number of sign $+1$ Latin squares of size n tends to one as n goes to infinity.

In [**HR94**], Huang and Rota showed:

Theorem 9.2.1.4. [**HR94**, Identities 8,9] *The difference between the number of column even Latin squares of size n and the number of column odd Latin squares of size n equals the difference between the number of even Latin squares of size n and the number of odd Latin squares of size n, up to sign. In particular, the Alon-Tarsi conjecture holds for n if and only if the column-sign Latin square conjecture holds for n.*

Thus

Theorem 9.2.1.5. [**?**] *The Alon-Tarsi conjecture holds for n if and only if $S_{n^n}(\mathbb{C}^n) \in \mathbb{C}[Ch_n(\mathbb{C}^n)]$.*

In [**?**] several additional statements equivalent to the conjecture were given. In particular, for those familiar with integration over compact Lie groups, the conjecture holds for n if and only if

$$\int_{(g^i_j)\in SU(n)} \Pi_{1\leq i,j\leq n} g^i_j d\mu \neq 0$$

where $d\mu$ is Haar measure.

9.3. Asymptotic surjectivity of the Hadamard-Howe map

*This section is still in rough form*****

9.3.1. Coordinate ring of the normalization of the Chow variety.

*** introduction about normalization, and normal varieties to be added***

In this section I follow [**Bri93**]. There is another variety whose coordinate ring is as computable as the coordinate ring of the orbit, the normalization of the Chow variety. We work in affine space.

An affine variety Z is *normal* if $\mathbb{C}[Z]$ is integrally closed, that is if every element of $\mathbb{C}(Z)$, the field of fractions of $\mathbb{C}[Z]$, that is integral over $\mathbb{C}[Z]$ (i.e., that satisfies a monic polynomial with coefficients in $\mathbb{C}[Z]$) is in $\mathbb{C}[Z]$. To every affine variety Z one may associate a unique normal affine variety $Nor(Z)$, called the *normalization* of Z, such that there is a finite map $\pi : Nor(Z) \to Z$ (i.e. $\mathbb{C}[Nor(Z)]$ is integral over $\mathbb{C}[Z]$), in particular it is generically one to one and one to one over the smooth points of Z. For details see [**Sha94**, Chap II.5].

In particular, there is an inclusion $\mathbb{C}[Z] \to \mathbb{C}[Nor(Z)]$ given by pullback of functions, e.g., given $f \in \mathbb{C}[Z]$, define $\tilde{f} \in \mathbb{C}[Nor(Z)]$ by $\tilde{f}(z) = f(\pi(x))$. If the non-normal points of Z form a finite set, then the cokernel is finite dimensional. If Z is a G-variety, then $Nor(Z)$ will be too.

Recall that $Ch_n(V)$ is the projection of the Segre variety, but since we want to deal with affine varieties, we will deal with the cone over it. Consider the product map

$$\phi_n : V^{\times n} \to S^n V$$
$$(u_1, \ldots, u_n) \mapsto u_1 \cdots u_n$$

Note that i) the image of ϕ_n is $\hat{C}h_n(V)$, ii) ϕ_n is $\Gamma_n = T_V \ltimes \mathfrak{S}_n$ equivariant.

For any affine algebraic group Γ and any Γ-variety Z, define the *GIT quotient* $Z//\Gamma$ to be the affine algebraic variety whose coordinate ring is $\mathbb{C}[Z]^\Gamma$. (When Γ is finite, this is just the usual set-theoretic quotient. In the general case, Γ-orbits will be identified in the quotient when there is no Γ-invariant regular function that can distinguish them.) If Z is normal, then so is $Z//\Gamma$ (see, e.g. [**Dol03**, Prop 3.1]). In our case $V^{\times n}$ is an affine Γ_n-variety and ϕ_n factors through the GIT quotient because it is Γ_n-equivariant, so a map

$$\psi_n : V^{\times n}//\Gamma_n \to S^n V$$

whose image is $\hat{C}h_n(V)$. By unique factorization, ψ_n is generically one to one. Elements of $V^{\times n}$ of the form $(0, u_2, \ldots, u_n)$ cannot be distinguished from $(0, \ldots, 0)$ by Γ_n invariant functions, so they are identified with $(0, \ldots, 0)$ in the quotient, which is consistent with the fact that $\phi_n(0, u_2, \ldots, u_n) = 0$. Observe that ϕ_n and ψ_n are $GL(V) = SL(V) \times \mathbb{C}^*$ equivariant.

Consider the induced map on coordinate rings:

$$\psi_n^* : \mathbb{C}[S^n V] \to \mathbb{C}[V^{\times n}//\Gamma_n] = \mathbb{C}[V^{\times n}]^{\Gamma_n}.$$

For affine varieties, $\mathbb{C}[Y \times Z] = \mathbb{C}[Y] \otimes \mathbb{C}[Z]$ (see, e.g. [**Sha94**, §2.2]), so

$$\begin{aligned}\mathbb{C}[V^{\times n}] &= \mathbb{C}[V]^{\otimes n} \\ &= Sym(V^*) \otimes \cdots \otimes Sym(V^*) \\ &= \bigoplus_{i_1, \ldots, i_n \in \mathbb{Z}_{\geq 0}} S^{i_1} V^* \otimes \cdots \otimes S^{i_n} V^*.\end{aligned}$$

Taking torus invariants gives

$$\mathbb{C}[V^{\times n}]^{T_n^{SL}} = \bigoplus_i S^i V^* \otimes \cdots \otimes S^i V^*,$$

and finally

$$(\mathbb{C}[V^{\times n}]^{T_n^{SL}})^{\mathfrak{S}_n} = \bigoplus_i S^n(S^i V^*).$$

In summary,

$$\psi_n^* : Sym(S^n V^*) \to \oplus_i (S^n(S^i V^*)),$$

and this map respects GL-degree, so it gives rise to maps $\tilde{h}_{d,n} : S^d(S^n V^*) \to S^n(S^d V^*)$.

Proposition 9.3.1.1. $\tilde{h}_{d,n} = h_{d,n}$.

Proof. Since elements of the form $x_1^n \cdots x_d^n$ span $S^d(S^n V)$ it will be sufficient to prove the maps agree on such elements. By Exercise 9.1.4.3, $h_{d,n}(x_1^n \cdots x_d^n) = (x_1 \cdots x_d)^n$. On the other hand, in the algebra $\mathbb{C}[V]^{\otimes n}$, the multiplication is $(f_1 \otimes \cdots \otimes f_n) \circledcirc (g_1 \otimes \cdots \otimes g_n) = f_1 g_1 \otimes \cdots \otimes f_n g_n$ and this descends to the algebra $(\mathbb{C}[V]^{\otimes n})^{\Gamma_n}$ which is the target of the algebra map ψ_n^*, i.e.,

$$\begin{aligned}\tilde{h}_{d,n}(x_1^n \cdots x_d^n) &= \psi_n^*(x_1^n \cdots x_d^n)\\ &= \psi_n^*(x_1^n) \circledcirc \cdots \circledcirc \psi_n^*(x_d^n)\\ &= x_1^n \circledcirc \cdots \circledcirc x_d^n\\ &= (x_1 \cdots x_d)^n.\end{aligned}$$

□

Proposition 9.3.1.2. $\psi_n : V^{\times n}//\Gamma_n \to \hat{C}h_n(V)$ *is the normalization of* $\hat{C}h_n(V)$.

9.3.2. Brion's asymptotic surjectivity result. A regular (see, e.g. [**Sha94**, p.27] for the definition of regular) map between affine varieties $f : X \to Y$ such that $f(X)$ is dense in Y is defined to be *finite* if $\mathbb{C}[X]$ is integral over $\mathbb{C}[Y]$ (see, e.g. [**Sha94**, p. 61]). To prove the proposition, we will need a lemma:

Lemma 9.3.2.1. *Let* X, Y *be affine varieties equipped with polynomial* $\mathbb{C}^*$-*actions with unique fixed points* $0_X \in X$, $0_Y \in Y$, *and let* $f : X \to Y$ *be a* $\mathbb{C}^*$-*equivariant morphism such that as sets,* $f^{-1}(0_Y) = \{0_X\}$. *Then* f *is finite.*

Proof of Proposition 9.3.1.2. Since $V^{\times n}//\Gamma_n$ is normal and ψ_n is regular and generically one to one, it just remains to show ψ_n is finite.

Write $[0] = [0, \ldots, 0]$. To show finiteness, by Lemma 9.3.2.1, it is sufficient to show ${\psi_n}^{-1}(0) = [0]$ as a set, as $[0]$ is the unique $\mathbb{C}^*$ fixed point in $V^{\times n}//\Gamma_n$, and every $\mathbb{C}^*$ orbit closure contains $[0]$. Now $u_1 \cdots u_n = 0$ if and only if some $u_j = 0$, say $u_1 = 0$. The T-orbit closure of $(0, u_2, \ldots, u_n)$ contains the origin so $[0, u_2, \ldots, u_n] = [0]$. □

Proof of Lemma 9.3.2.1. $\mathbb{C}[X], \mathbb{C}[Y]$ are $\mathbb{Z}_{\geq 0}$-graded, and the hypothesis $f^{-1}(0_Y) = \{0_X\}$ states that $\mathbb{C}[X]/f^*(\mathbb{C}[Y]_{>0})\mathbb{C}[X]$ is a finite dimensional vector space. We want to show that $\mathbb{C}[X]$ is integral over $\mathbb{C}[Y]$. This is a graded version of Nakayama's Lemma (the algebraic implicit function theorem). □

In more detail (see, e.g. [**Kum13**, Lemmas 3.1,3.2], or [**Eis95**, p136, Ex. 4.6a]):

Lemma 9.3.2.2. *Let R, S be $\mathbb{Z}_{\geq 0}$-graded, finitely generated domains over $\mathbb{C}$ such that $R_0 = S_0 = \mathbb{C}$, and let $f^* : R \to S$ be an injective graded algebra homomorphism. If $f^{-1}(R_{>0}) = \{S_{>0}\}$ as sets, where $f : Spec(S) \to Spec(R)$ is the induced map on the associated schemes, then S is a finitely generated R-module. In particular, it is integral over R.*

Proof. The hypotheses on the sets says that $S_{>0}$ is the only maximal ideal of S containing the ideal $\mathfrak{m}$ generated by $f^*(R_{>0})$, so the radical of $\mathfrak{m}$ must equal $S_{>0}$, and in particular $S^d_{>0}$ must be contained in it for all $d > d_0$, for some d_0. So $S/\mathfrak{m}$ is a finite dimensional vector space, and by the next lemma, S is a finitely generated R-module. □

Lemma 9.3.2.3. *Let S be as above, and let M be a $\mathbb{Z}_{\geq 0}$-graded S-module. Assume $M/(S_{>0} \cdot M)$ is a finite dimensional vector space over $S/S_{>0} \simeq \mathbb{C}$. Then M is a finitely generated S-module.*

Proof. Choose a set of homogeneous generators $\{\overline{x}_1, \dots, \overline{x}_n\} \subset M/(S_{>0} \cdot M)$ and let $x_j \in M$ be a homogeneous lift of $\overline{x}_j$. Let $N \subset M$ be the graded S-submodule $Sx_1 + \cdots + Sx_n$. Then $M = S_{>0}M + N$, as let $a \in M$, consider $\overline{a} \in M/(S_{>0}M)$ and lift it to some $b \in N$, so $a - b \in S_{>0}M$, and $a = (a - b) + b$. Now quotient by N to obtain

$$S_{>0} \cdot (M/N) = M/N. \tag{9.3.1}$$

If $M/N \neq 0$, let d_0 be the smallest degree such that $(M/N)^{d_0} \neq 0$. But $S_{>0} \cdot (M/N)^{\geq d_0} \subset (M/N)^{\geq d_0 + 1}$ so there is no way to obtain $(M/N)^{d_0}$ on the right hand side. Contradiction. □

Theorem 9.3.2.4. [**Bri93**] *For all $n \geq 1$, ψ_n restricts to a map*

$$\psi_n^o : (V^{\times n}//\Gamma_n)\backslash[0] \to S^n V \backslash 0 \tag{9.3.2}$$

such that $\psi_n^{o*} : \mathbb{C}[S^n V \backslash 0] \to \mathbb{C}[(V^{\times n}//\Gamma_n)\backslash[0]]$ is surjective.

Corollary 9.3.2.5. [**Bri93**] *The Hermite-Hadamard-Howe map*

$$h_{d,n} : S^d(S^n V^*) \to S^n(S^d V^*)$$

is surjective for d sufficiently large.

Proof of Corollary. Theorem 9.3.2.4 implies $(\psi_n^*)_d$ is surjective for d sufficiently large, because the cokernel of ψ_n^* is supported at a point and thus must vanish in large degree. □

The proof of Theorem 9.3.2.4 will give a second proof that the kernel of ψ_n^* equals the ideal of $Ch_n(V)$.

Proof of Theorem. Since ψ_n is $\mathbb{C}^*$-equivariant, we can consider the quotient to projective space

$$\underline{\psi}_n : ((V^{\times n}//\Gamma_n)\backslash[0])/\mathbb{C}^* \to (S^nV\backslash 0)/\mathbb{C}^* = \mathbb{P}S^nV$$

and show that $\underline{\psi}_n^*$ is surjective. Note that $((V^{\times n}//\Gamma_n)\backslash[0])/\mathbb{C}^*$ is $GL(V)$-isomorphic to $(\mathbb{P}V)^{\times n}/\mathfrak{S}_n$, as

$$(V^{\times n}//\Gamma_n)\backslash[0] = (V\backslash 0)^{\times n}/\Gamma_n$$

and $\Gamma_n \times \mathbb{C}^* = (\mathbb{C}^*)^{\times n} \rtimes \mathfrak{S}_n$. So

$$\underline{\psi}_n : (\mathbb{P}V)^{\times n}/\mathfrak{S}_n \to \mathbb{P}S^nV.$$

It will be sufficient to show $\underline{\psi}_n^*$ is surjective on affine open subsets that cover the spaces. Let $w_1, \dots, w_{\mathbf{w}}$ be a basis of V and consider the affine open subset of $\mathbb{P}V$ given by elements where the coordinate on w_1 is nonzero, and the corresponding induced affine open subsets of $(\mathbb{P}V)^{\times n}$ and $\mathbb{P}S^nV$, call these $(\mathbb{P}V)_1^{\times n}$ and $(\mathbb{P}S^nV)_1$. We will show that the algebra of $\mathfrak{S}_n$-invariant functions on $(\mathbb{P}V)_1^{\times n}$ is in the image of $(\mathbb{P}S^nV)_1$. The restriction of the quotient by $\mathfrak{S}_n$ of $(\mathbb{P}V)^{\times n}$ composed with $\underline{\psi}_n$ to these open subsets in coordinates is

$$((w_1 + \sum_{s=2}^{\mathbf{w}} x_s^1 w_s), \dots, (w_1 + \sum_{s=2}^{\mathbf{w}} x_s^{\mathbf{w}} w_s) \mapsto \Pi_{i=1}^n (w_1 + \sum_{s=2}^{\mathbf{w}} x_s^i w_s).$$

Finally, by e.g., [**Wey97**, §II.3], the coordinates on the right hand side generate the algebra of $\mathfrak{S}_n$-invariant functions in the n sets of variables $(x_s^i)_{i=1,\dots,n}$. $\square$

With more work, in [**Bri97**, Thm 3.3], Brion obtains an explicit (but enormous) function $d_0(n, \mathbf{w})$ which is

$$d_0(n, \mathbf{w}) = (n-1)(\mathbf{w}-1)((n-1)\left\lfloor \frac{\binom{n+\mathbf{w}-1}{\mathbf{w}-1}}{\mathbf{w}} \right\rfloor - n) \tag{9.3.3}$$

for which the $h_{d,n}$ is surjective for all $d > d_0$ where $\dim V = \mathbf{v}$.

Problem 9.3.2.6. Improve Brion's bound to say, a polynomial bound in n when $n = \mathbf{w}$.

Problem 9.3.2.7. Note that $\mathbb{C}[Nor(Ch_n(V))] = \mathbb{C}[GL(V) \cdot (x_1 \cdots x_n)]_{\geq 0}$ and that the the boundary of the orbit closure is irreducible. Under what conditions will a $GL(V)$-orbit closure with reductive stabilizer that has an irreducible boundary will be such that the coordinate ring of the normalization of the orbit closure equals the positive part of the coordinate ring of the orbit?

9.3.3. Brion's qualitative theorem. We have a ring map

$$h_n : Sym(S^nV) \to \bigoplus_i S^n(S^iV) \tag{9.3.4}$$

The proof has three steps:

(1) Show $\mathbb{C}[Nor(Ch_n(V))]$ is generated in degree at most $(n-1)(\mathbf{v}-1)$ via vanishing of cohomology (Castelnuovo-Mumford regularity).

(2) Show that $h_n((v^n)^{d(n-1)} \cdot \mathbb{C}[Nor(Ch_n(V))] \subset \mathbb{C}[Ch_n(V)]$ via a localization argument to reduce to a question about multi-symmetric functions.

(3) Use that Zariski open subset of the polynomials of degree n in $\mathbf{v}$ variables can be written as a sum of r_0 n-th powers, where $r_0 \sim \frac{1}{n}\binom{\mathbf{v}+n-1}{n}$ (The Alexander-Hirschowitz theorem [**AH95**]).

Then we conclude that for $d \geq (n-1)(\mathbf{v}-1)(r_0(n-1)+n)$ that $h_{d,n}$ is surjective.

Proof of Step 1. We saw *** that $\mathbb{C}[Nor(Ch_n(V))] = (\mathbb{C}[V^{*\times n}]^{T_n})^{\mathfrak{S}_n}$ so it will be sufficient to show that $\mathbb{C}[V^{*\times n}]^{T_n}$ is generated in degree at most $(n-1)(\mathbf{v}-1)$. We translate this into a sheaf cohomology problem:

$$\begin{aligned}\mathbb{C}[V^{*\times n}]^{T_n} &= \bigoplus_{d=0}^{\infty} H^0(\mathbb{P}V^{*\times n}, \mathcal{O}_{\mathbb{P}V^*}(d)^{\times n}) \\ &= \bigoplus_{d=0}^{\infty} H^0(\mathbb{P}S^nV^*, proj_*\mathcal{O}_{\mathbb{P}V^*}(d)^{\times n})\end{aligned}$$

i.e., we want to know about the generators of the graded $Sym(S^nV)$-module associated to the sheaf $proj_*\mathcal{O}_{\mathbb{P}V^*}^{\times n}$. Castelnuovo-Mumford regularity [**Mum66**, Lect. 14] gives a bound in terms of vanishg of sheaf cohomology groups. Here we are dealing with groups we can compute: $H^j(\mathbb{P}V^{*\times n}, \mathcal{O}(d-j)^{\times n})$, and the result follows from this computation. □

Proof of Step 2. Let $v = v_{\mathbf{v}} \in V\backslash 0$, and let $v_1, \ldots, v_{\mathbf{v}}$ be a basis of V. Set $x_i = \frac{v_i}{v}$. Consider the localization of the coordinate ring of the normalization at v^n (the degree zero elements in the localization of $\mathbb{C}[Nor(Ch_n(V))][\frac{1}{v^n}]$:

$$\begin{aligned}\mathbb{C}[Nor(Ch_n(V))]_{v^n} := &\bigcup_{d\geq 0} S^n(S^dV)(v^n)^{-d}\\ &= S^n(\bigcup_{d\geq 0}(S^dV)(v^n)^{-d}\\ &= S^n\mathbb{C}[x_1,\dots,x_{\mathbf{v}-1}] =: S^n\mathbb{C}[\overline{x}]\\ &= [(\mathbb{C}[\overline{x}])^{\otimes n}]^{\mathfrak{S}_n}\\ &= (\mathbb{C}[\overline{x_1},\dots,\overline{x_n}])^{\mathfrak{S}_n}\end{aligned}$$

where $\overline{x_j} = (x_{1,j},\dots,x_{\mathbf{v}-1,j})$.

Similarly

$$\begin{aligned}Sym(S^nV)_{v^n} &= \bigcup_{d\geq 0} S^d(S^nV)(v^n)^{-d}\\ &= Sym(S^nV/v^n)\\ &= Sym(\bigoplus_{i=1}^n \mathbb{C}[\overline{x}]_i)\end{aligned}$$

We get a localized graded algebra map h_{n,v^n} between these spaces. Hence it is determined in degree one:

$$\bigoplus_{i=1}^n \mathbb{C}[\overline{x}]_i \to \mathbb{C}[\overline{x_1},\dots,\overline{x_n}]^{\mathfrak{S}_n}$$

that takes the degree at most n monomial $x_1^{a_1}\cdots x_{d-1}^{a_{d_1}}$ to the coefficient of $t_1^{a_1}\cdots t_{d-1}^{a_{d-1}}$ in the expansion of

$$\Pi_{i=1}^n(1+\overline{x_{i1}}t_1+\cdots+\overline{x_{id-1}}t_{d-1})$$

These are the *elementary multi-symmetric functions*. They generate the ring of multi-symmetric functions $\mathbb{C}[\overline{x_1},\dots,\overline{x_n}]^{\mathfrak{S}_n}$ [**AK81**]. Thus h_{n,v^n} is surjective.

Moreover, if $f \in \mathbb{C}[\overline{x_1},\dots,\overline{x_n}]^{\mathfrak{S}_n}$ has all its partial degrees at most d, then the total degree of f is at most dn in the $\overline{x_j}$'s, so it is a polynomial of degree at most dn in the elementary multi-symmetric functions. In other words, the map

$$S^{dn}(S^nV)(v^n)^{-dn} \to S^n(S^dV)(v^n)^{-d}$$

is surjective, so $h_n((v^n)^{d(n-1)}\mathbb{C}[Nor(Ch_n(V))] \subset \mathbb{C}[Ch_n(V)]$. □

We conclude by appeal to the Alexander-Hirschowitz theorem [**AH95**].

Chapter 10

Valiant's conjecture III: Results using algebraic geometry

Warning: this chapter is in rough form

10.1. Non-normality of $\mathcal{D}et_n$

give context be sure to include how SL-orbits are closed*** I follow [**Kum13**] in this section. Throughout this section I make the following assumptions and adopt the following notation:

Set up:

- V is a $GL(W)$-module,
- Let $\mathcal{P}^0 := GL(W) \cdot P$ and $\mathcal{P} := \overline{GL(W) \cdot P}$ denote its orbit and orbit closure, and let $\partial\mathcal{P} = \mathcal{P} \backslash \mathcal{P}^0$ denote its boundary, which we assume to be more than zero (otherwise $[\mathcal{P}]$ is homogeneous).

(10.1.1) **Assumptions** :

(1) $P \in V$ is such that the $SL(W)$-orbit of P is closed.

(2) The stabilizer $G_P \subset GL(W)$ is reductive, which is equivalent (by a theorem of Matsushima [**Mat60**]) to requiring that $\mathcal{P}^0$ is an affine variety.

This situation holds when $V = S^nW$, $\dim W = n^2$ and $P = \det_n$ or perm_n as well as when $\dim W = rn$ and $P = S^r_n := \sum_{j=1}^r x_1^j \cdots x_n^j$, the sum-product polynomial, in which case $\mathcal{P} = \hat{\sigma}_r(Ch_n(W))$.

Lemma 10.1.0.1. [**Kum13**] *Assumptions as in* (10.1.1). *Let $M \subset \mathbb{C}[\mathcal{P}]$ be a nonzero $GL(W)$-module, and let $Z(M) = \{y \in \mathcal{P} \mid f(y) = 0\ \forall f \in M\}$ denote its zero set. Then $0 \subseteq Z(M) \subseteq \partial\mathcal{P}$.*

If moreover $M \subset I(\partial\mathcal{P})$, then as sets, $Z(M) = \partial\mathcal{P}$.

Proof. Since $Z(M)$ is a $GL(W)$-stable subset, if it contains a point of $\mathcal{P}^0$ it must contain all of $\mathcal{P}^0$ and thus M vanishes identically on $\mathcal{P}$, which cannot happen as M is nonzero. Thus $Z(M) \subseteq \partial\mathcal{P}$. For the second assertion, since $M \subset I(\partial\mathcal{P})$, we also have $Z(M) \supseteq \partial\mathcal{P}$. $\square$

Proposition 10.1.0.2. [**Kum13**] *Assumptions as in* (10.1.1). *The space of $SL(W)$-invariants of positive degree in the coordinate ring of $\mathcal{P}$, $\mathbb{C}[\mathcal{P}]^{SL(W)}_{>0}$, is non-empty and contained in $I(\partial\mathcal{P})$. Moreover,*

(1) *any element of $\mathbb{C}[\mathcal{P}]^{SL(W)}_{>0}$ cuts out $\partial\mathcal{P}$ set-theoretically, and*

(2) *the components of $\partial\mathcal{P}$ all have codimension one in $\mathcal{P}$.*

Proof. To study $\mathbb{C}[\mathcal{P}]^{SL(W)}$, consider the GIT quotient $\mathcal{P}//SL(W)$ whose coordinate ring, by definition, is $\mathbb{C}[\mathcal{P}]^{SL(W)}$. It parametrizes the closed $SL(W)$-orbits in $\mathcal{P}$, so it is non-empty. Thus $\mathbb{C}[\mathcal{P}]^{SL(W)}$ is nontrivial.

Claim: every $SL(W)$-orbit in ∂P contains $\{0\}$ in its closure, i.e., $\partial\mathcal{P}$ maps to zero in the GIT quotient. This will imply any $SL(W)$-invariant of positive degree is in $I(\partial\mathcal{P})$ because any non-constant function on the GIT quotient vanishes on the inverse image of $[0]$. Thus (1) follows from Lemma 10.1.0.1. The zero set of a single polynomial, if it is not empty, has codimension one, which implies the components of $\partial\mathcal{P}$ are all of codimension one, proving (2).

It remains to show $\partial\mathcal{P}$ maps to zero in $\mathcal{P}//SL(W)$, where $\rho : GL(W) \to GL(V)$ is the representation. This GIT quotient inherits a $\mathbb{C}^*$ action via $\rho(\lambda Id)$, for $\lambda \in \mathbb{C}^*$. Its normalization is just the affine line $\mathbb{A}^1 = \mathbb{C}$. To see this, consider the $\mathbb{C}^*$-equivariant map $\sigma : \mathbb{C} \to \mathcal{P}$ given by $z \mapsto \rho(zId) \cdot P$, which descends to a map $\overline{\sigma} : \mathbb{C} \to \mathcal{P}//SL(W)$. Since the $SL(W)$-orbit of P is closed, for any $\lambda \in \mathbb{C}^*$, $\rho(\lambda Id)P$ does not map to zero in the GIT quotient, so we have $\overline{\sigma}^{-1}([0]) = \{0\}$ as a set. Lemma 9.3.2.1 applies so $\overline{\sigma}$ is finite and gives the normalization. Finally, were there a closed nonzero orbit in $\partial\mathcal{P}$, it would have to equal $SL(W) \cdot \sigma(\lambda)$ for some $\lambda \in \mathbb{C}^*$ since $\overline{\sigma}$ is surjective. But $SL(W) \cdot \sigma(\lambda) \subset \mathcal{P}^0$. $\square$

Remark 10.1.0.3. That each irreducible component of $\partial\mathcal{P}$ is of codimension one in $\mathcal{P}$ is due to Matsushima [**Mat60**]. It is a consequence of his result mentioned above.

The key to proving non-normality of $\hat{\mathcal{D}et_n}$ and $\hat{\mathcal{P}erm_n^n}$ is to find an $SL(W)$-invariant in the coordinate ring of the normalization (which has a $GL(W)$-grading), which does not occur in the corresponding graded component of the coordinate ring of S^nW, so it cannot occur in the coordinate ring of any $GL(W)$-subvariety.

Lemma 10.1.0.4. *Assumptions as in* (10.1.1)*. Let $P \in S^nW$ be such that $SL(W) \cdot P$ is closed and G_P is reductive. Let d be the smallest positive $GL(W)$-degree such that $\mathbb{C}[\mathcal{P}^0]_d^{SL(W)} \neq 0$. If n is even and $d < n\mathbf{w}$ (resp. n is odd and $d < 2n\mathbf{w}$) then $\mathcal{P}$ is not normal.*

Proof. Since $\mathcal{P}^0 \subset \mathcal{P}$ is a Zariski open subset, we have the equality of $GL(W)$-modules $\mathbb{C}(\mathcal{P}) = \mathbb{C}(\mathcal{P}^0)$. By restriction of functions $\mathbb{C}[\mathcal{P}] \subset \mathbb{C}[\mathcal{P}^0]$ and thus $\mathbb{C}[\mathcal{P}]^{SL(W)} \subset \mathbb{C}[\mathcal{P}^0]^{SL(W)}$. Now $\mathcal{P}^0//SL(W) = \mathcal{P}^0/SL(W) \simeq \mathbb{C}^*$, so $\mathbb{C}[\mathcal{P}^0]^{SL(W)} \simeq \oplus_{k\in\mathbb{Z}}\mathbb{C}\{z^k\}$. Under this identification, z has $GL(W)$-degree d. By Proposition 10.1.0.2, $\mathbb{C}[\mathcal{P}]^{SL(W)} \neq 0$. Let $h \in \mathbb{C}[\mathcal{P}]^{SL(W)}$ be the smallest element in positive degree. Then $h = z^k$ for some k. Were $\mathcal{P}$ normal, we would have $k = 1$.

But now we also have a surjection $\mathbb{C}[S^nW] \to \mathbb{C}[\mathcal{P}]$, and by Exercise **??** the smallest possible $GL(W)$-degree of an $SL(W)$-invariant in $\mathbb{C}[S^nW]$ when n is even (resp. odd) is $\mathbf{w}n$ (resp. $2\mathbf{w}n$) which would occur in $S^{\mathbf{w}}(S^nW)$ (resp. $S^{2\mathbf{w}}(S^nW)$). We obtain a contradiction. □

Theorem 10.1.0.5 (Kumar [**Kum13**])**.** *For all $n \geq 3$, $\mathcal{D}et_n$ and $\mathcal{P}erm_n^n$ are not normal. For all $n \geq 2m$ (the range of interest), $\mathcal{P}erm_n^m$ is not normal.*

I give the proof for $\mathcal{D}et_n$, the case of $\mathcal{P}erm_n^n$ is an easy exercise. Despite the variety being much more singular, the proof for $\mathcal{P}erm_n^m$ is more difficult, see [**Kum13**].

Proof. We will show that when n is congruent to 0 or 1 mod 4, $\mathbb{C}[\mathcal{D}et_n^0]_{n-GL}^{SL(W)} \neq 0$ and when n is congruent to 2 or 3 mod 4, $\mathbb{C}[\mathcal{D}et_n^0]_{2n-GL}^{SL(W)} \neq 0$. Since $n, 2n < (n^2)n$ Lemma 10.1.0.4 applies.

The $SL(W)$-trivial modules are $(\Lambda^{n^2}W)^{\otimes s} = S_{s^{n^2}}W$. Write $W = E \otimes F$. We want to determine the lowest degree trivial $SL(W)$-module that has a $G_{det_n} = (SL(E) \times SL(F)/\mu_n) \rtimes \mathbb{Z}_2$ invariant. We have the decomposition $(\Lambda^{n^2}W)^{\otimes s} = (\oplus_{|\pi|=n^2} S_\pi E \otimes S_{\pi'}F)^{\otimes s}$, where π' is the conjugate partition to π. Thus $(\Lambda^{n^2}W)^{\otimes s}$ contains the trivial $SL(E) \times SL(F)$ module $(\Lambda^n E)^{\otimes ns} \otimes (\Lambda^n F)^{\otimes ns}$ with multiplicity one. (In the language of §8.9.2,

$k_{s^{n^2},(sn)^n,(sn)^n} = 1$.) Now we consider the effect of the $\mathbb{Z}_2 \subset G_{\det_n}$ with generator $\tau \in GL(W)$. It sends $e_i \otimes f_j$ to $e_j \otimes f_i$, so acting on W it has $+1$ eigenspace $e_i \otimes f_j + e_j \otimes f_i$ for $i \leq j$ and -1 eigenspace $e_i \otimes f_j - e_j \otimes f_i$ for $1 \leq i < j \leq n$. Thus it acts on the one-dimensional vector space $(\Lambda^{n^2} W)^{\otimes s}$ by $((-1)^{\binom{n}{2}})^s$, i.e., by -1 if $n \equiv 2, 3 \bmod 4$ and s is odd and by 1 otherwise. We conclude that there is an invariant as asserted above. (In the language of §8.9.2, $sk^{s^{n^2}}_{(sn)^n,(sn)^n} = 1$ for all s when $\binom{n}{2}$ is even, and $sk^{s^{n^2}}_{(sn)^n,(sn)^n} = 1$ for even s when $\binom{n}{2}$ is odd and is zero for odd s.) $\square$

Exercise 10.1.0.6: Write out the proof of the non-normality of $\mathcal{P}erm^n_n$.

Exercise 10.1.0.7: Show the same method gives another proof that $Ch_n(W)$ is not normal, but that it fails (with good reason) to show $\sigma_n(v_d(\mathbb{P}^{n-1}))$ is not normal.

Exercise 10.1.0.8: Show a variant of the above holds for any reductive group with a nontrivial center (one gets a $\mathbb{Z}^k$-grading of modules if the center is k-dimensional), in particular it holds for $G = GL(A) \times GL(B) \times GL(C)$. Use this to show that $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ is not normal when $\dim A = \dim B = \dim C = r > 2$.

10.2. The minimal free resolution of the ideal generated by minors of size κ

This section follows the exposition in [**?**], which is based on the presentation in [**Wey03**]. The results are due to Lascoux [**Las78**].

10.2.1. Statement of the result.

Let $E, F = \mathbb{C}^n$, give $E \otimes F$ coordinates (x^i_j), with $1 \leq i, j \leq n$. Set $r = \kappa - 1$. Let $\hat{\sigma}_r = \hat{\sigma}_r(Seg(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1})) \subset \mathbb{C}^n \otimes \mathbb{C}^n = E^* \otimes F^*$ denote the variety of $n \times n$ matrices of rank at most r. By "degree $S_\pi E$", we mean $|\pi| = p_1 + \cdots + p_n$. Write $\ell(\pi)$ for the largest j such that $p_j > 0$. Write $\pi + \pi' = (p_1 + p'_1, \ldots, p_n + p'_n)$.

The *weight* (under $GL(E) \times GL(F)$) of a monomial $x^{i_1}_{j_1} \cdots x^{i_q}_{j_q} \in S^q(E \otimes F)$ is given by a pair of n-tuples $((w^E_1, \ldots, w^E_n), (w^F_1, \ldots, w^F_n))$ where w^E_s is the number of i_α's equal to s and w^F_t is the number of j_α's equal to t. A vector is a weight vector of weight $((w^E_1, \ldots, w^E_n), (w^F_1, \ldots, w^F_n))$ if it can be written as a sum of monomials of weight $((w^E_1, \ldots, w^E_n), (w^F_1, \ldots, w^F_n))$. Any $GL(E) \times GL(F)$-module has a basis of weight vectors, and any irreducible module has a unique *highest weight* which (if the representation is polynomial) is a pair of partitions, $(\pi, \mu) = ((p_1, \ldots, p_n), (m_1, \ldots, m_n))$, where we allow a string of zeros to be added to a partition to make it of length n. The corresponding $GL(E) \times GL(F)$-module is denoted $S_\pi E \otimes S_\mu F$.

Theorem 10.2.1.1. [**Las78**] *Let $0 \to F_N \to \cdots \to F_1 \to Sym(E \otimes F) = F_0 \to \mathbb{C}[\hat{\sigma}_r] \to 0$ denote the minimal free resolution of $\hat{\sigma}_r$. Then*

(1) *$N = (n-r)^2$, i.e., $\hat{\sigma}_r$ is arithmetically Cohen-Macaulay.*

(2) *$\hat{\sigma}_r$ is Gorenstein, i.e., $F_N = Sym(E \otimes F)$, generated by $S_{(n-r)^n} E \otimes S_{(n-r)^n} F$. In particular $F_{N-j} \simeq F_j$ as $SL(E) \times SL(F)$- modules, although they are not isomorphic as $GL(E) \times GL(F)$-modules.*

(3) *For $1 \leq j \leq N-1$, the space F_j has generating modules of degree $sr + j$ where $1 \leq s \leq \lfloor \sqrt{j} \rfloor$. The modules of degree $r + j$ form the generators of the linear strand of the minimal free resolution.*

(4) *The generating module of F_j is multiplicity free.*

(5) *Let α, β be (possibly zero) partitions such that $\ell(\alpha), \ell(\beta) \leq s$. Independent of the lengths (even if they are zero), write $\alpha = (\alpha_1, \dots, \alpha_s)$, $\beta = (\beta_1, \dots, \beta_s)$. The degree $sr + j$ generators of F_j, for $1 \leq j \leq N$ are*

$$M_{j,rs+j} = \bigoplus_{s \geq 1} \bigoplus_{\substack{|\alpha|+|\beta|=j-s^2 \\ \ell(\alpha),\ell(\beta)\leq s}} S_{(s)^{r+s}+(\alpha,0^r,\beta')} E \otimes S_{(s)^{r+s}+(\beta,0^r,\alpha')} F. \tag{10.2.1}$$

The Young diagrams of the modules are depicted in Figure 1 below.

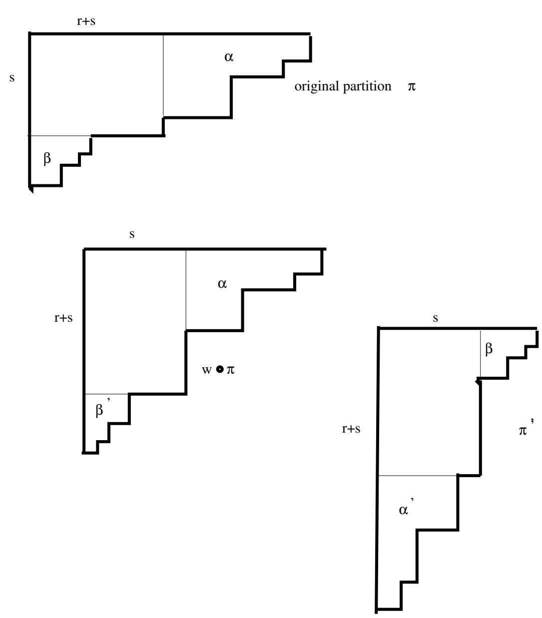


Figure 10.2.1. Partition π and pairs of partitions $(s)^{r+s} + (\alpha, 0^r, \beta') = w \cdot \pi$ and $(s)^{r+s} + (\beta, 0^r, \alpha') = \pi'$ it gives rise to in the resolution (see §10.2.4 for explanations).

(6) *In particular the generator of the linear component of F_j is*

$$M_{j,j+r} = \bigoplus_{a+b=j-1} = S_{a+1,1^{r+b}}E \otimes S_{b+1,1^{r+a}}F. \tag{10.2.2}$$

This module admits a basis as follows: form a size $r+j$ submatrix using $r+b+1$ distinct rows, repeating a subset of a rows to have the correct number of rows and $r+a+1$ distinct columns, repeating a subset of b columns, and then performing a "tensor Laplace expansion" as described below.

10.2.2. The Koszul resolution. If $\mathcal{I} = Sym(V)$, the minimal free resolution is given by the exact complex

$$\cdots \to S^{q-1}V \otimes \Lambda^{p+2}V \to S^qV \otimes \Lambda^{p+1}V \to S^{q+1}V \otimes \Lambda^pV \to \cdots \tag{10.2.3}$$

The maps are given by the transpose of exterior derivative (Koszul) map $d_{p,q} : S^qV^* \otimes \Lambda^{p+1}V^* \to S^{q-1}V^* \otimes \Lambda^{p+2}V^*$. Write $d^T_{p,q} : S^{q-1}V \otimes \Lambda^{p+2}V \to S^qV \otimes \Lambda^{p+1}V$. We have the $GL(V)$-decomposition $S^qV \otimes \Lambda^{p+1}V = S_{q,1^{p+1}}V \oplus S_{q+1,1^p}V$, so the kernel of $d^T_{p,q}$ is the first module, which also is the image of $d^T_{p+1,q-1}$.

Explicitly, $d^T_{p,q}$ is the composition of polarization ($\Lambda^{p+2}V \to \Lambda^{p+1}V \otimes V$) and multiplication:

$$S^{q-1}V \otimes \Lambda^{p+2}V \to S^{q-1}V \otimes \Lambda^{p+1}V \otimes V \to S^qV \otimes \Lambda^{p+1}V.$$

For the minimal free resolution of any ideal, the linear strand will embed inside (10.2.3).

Throughout this article, we will view $S_{q+1,1^p}V$ as a submodule of $S^qV \otimes \Lambda^{p-1}V$, $GL(V)$-complementary to $d^T_{p,q}(S^{q-1,1^p}V)$.

For $T \in S^\kappa V \otimes V^{\otimes j}$, and $P \in S^\ell V$, introduce notation for multiplication on the first factor, $T \cdot P \in S^{\kappa+\ell}V \otimes V^{\otimes j}$. Write $F_j = M_j \cdot Sym(V)$. As always, $M_0 = \mathbb{C}$.

10.2.3. Geometric interpretations of the terms in the linear strand (10.2.2). First note that $F_1 = M_1 \cdot Sym(E \otimes F)$, where $M_1 = M_{1,r+1} = \Lambda^{r+1}E \otimes \Lambda^{r+1}F$, the size $r+1$ minors which generate the ideal. The syzygies among these equations are generated by

$$M_{2,r+2} := S_{1^{r+2}}E \otimes S_{21^r}F \oplus S_{21^r}E \otimes S_{1^{r+2}}F \subset \mathcal{I}^{\sigma_r}_{r+2} \otimes V$$

(i.e., $F_2 = M_2 \cdot Sym(E \otimes F)$), where elements in the first module may be obtained by choosing $r+1$ rows and $r+2$ columns, forming a size $r+2$ square matrix by repeating one of the rows, then doing a 'tensor Laplace expansion" that we now describe:

In the case $r = 1$ we have highest weight vector

(10.2.4)
$$\begin{aligned} S^{1|12}_{123} &:= (x^1_2 x^2_3 - x^2_2 x^1_3) \otimes x^1_1 - (x^1_1 x^2_3 - x^2_1 x^1_3) \otimes x^1_2 + (x^1_1 x^2_2 - x^1_2 x^2_1) \otimes x^1_3 \\ &= M^{12}_{23} \otimes x^1_1 - M^{12}_{13} \otimes x^1_2 + M^{12}_{12} \otimes x^1_3 \end{aligned}$$

where in general M^I_J will denote the minor obtained from the submatrix with indices I, J. The expression (10.2.4) corresponds to the Young tableaux pair:

$$\begin{array}{|c|c|} \hline 1 & 1 \\ \hline 2 & \multicolumn{1}{c}{} \\ \cline{1-1} \end{array}\ , \quad \begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline 3 \\ \hline \end{array}\,.$$

To see (10.2.4) is indeed a highest weight vector, first observe that it has the correct weights in both E and F, and that in the F-indices $\{1, 2, 3\}$ it is skew and that in the first two E indices it is also skew. Finally to see it is a highest weight vector note that any raising operator sends it to zero. Also note that under the multiplication map $S^2V \otimes V \to S^3V$ the element maps to zero, because the map corresponds to converting a tensor Laplace expansion to an actual one, but the determinant of a matrix with a repeated row is zero.

In general, a basis of $S_\pi E \otimes S_\mu F$ is indexed by pairs of semi-standard Young tableau in π and μ. In the linear strand, all partitions appearing are hooks, a basis of $S_{a,1^b}E$ is given by two sequences of integers taken from $[n]$, one weakly increasing of length a and one strictly increasing of length b, where the first integer in the first sequence is at least the first integer in the second sequence.

A highest weight vector in $S_{21^r}E \otimes S_{1^{r+2}}F$ is

$$S^{1|1,\dots,r+1}_{1,\dots,r+2} = M^{1,\dots,r+1}_{2,\dots,r+2} \otimes x^1_1 - M^{1,\dots,r+1}_{1,3,\dots,r+1} \otimes x^1_2 + \cdots + (-1)^r M^{1,\dots,r+1}_{1,\dots,r+1} \otimes x^1_{r+2},$$

and the same argument as above shows it has the desired properties. Other basis vectors are obtained by applying lowering operators to the highest weight vector, so their expressions will be more complicated.

Remark 10.2.3.1. If we chose a size $r+2$ submatrix, and perform a tensor Laplace expansion of its determinant about two different rows, the difference of the two expressions corresponds to a linear syzygy, but these are in the span of M_2. These expressions are important for comparison with the permanent, as they are the only linear syzygies for the ideal generated by the size $r + 1$ sub-permanents, where one takes the permanental Laplace expansion.

Continuing, F_3 is generated by the module

$$M_{3,r+3} = S_{1^{r+3}}E \otimes S_{3,1^r}F \oplus S_{2,1^{r+1}}E \otimes S_{2,1^{r+1}}F \oplus S_{3,1^r}E \otimes S_{1^{r+3}}F \subset M_2 \otimes V.$$

These modules admit bases of double tensor Laplace type expansions of a square submatrix of size $r+3$. In the first case, the highest weight vector is obtained from the submatrix whose rows are the first $r+3$ rows of the original matrix, and whose columns are the first r-columns with the first column repeated three times. For the second module, the highest weight vector is obtained from the submatrix whose rows and columns are the first $r+2$ such, with the first row/column repeated twice. A highest weight vector for $S_{3,1^r}E \otimes S_{1^{r+3}}F$ is

$$\begin{aligned} S^{11|1,\dots,r+1}_{1,\dots,r+3} &= \sum_{1 \le \beta_1 < \beta_2 \le r+3} (-1)^{\beta_1+\beta_2} M^{1,\dots,r+1}_{1,\dots,\hat{\beta}_1,\dots,\hat{\beta}_2,\dots,r+3} \otimes (x^1_{\beta_1} \wedge x^1_{\beta_2}) \\ &= \sum_{\beta=1}^{r+3} (-1)^{\beta+1} S^{1|1,\dots,i_{r+1}}_{1,\dots,\hat{\beta},\dots,r+3} \otimes x^1_\beta. \end{aligned}$$

Here $S^{1|1,\dots,i_{r+1}}_{1,\dots,\hat{\beta},\dots,r+3}$ is defined in the same way as the highest weight vector.

A highest weight vector for $S_{2,1^{r+1}}E \otimes S_{2,1^{r+1}}F$ is

$$\begin{aligned} S^{1|1,\dots,r+3}_{1|1,\dots,r+2} &= \sum_{\alpha,\beta=1}^{r+3} (-1)^{\alpha+\beta} M^{1,\dots,\hat{\alpha},\dots,r+2}_{1,\dots,\hat{\beta},\dots,i+2} \otimes (x^\alpha_1 \wedge x^1_\beta) \\ &= \sum_{\beta=1}^{r+3} (-1)^{\beta+1} S^{1,\dots,r+2}_{1|1,\dots,\hat{\beta},\dots,r+2} \otimes x^1_\beta - \sum_{\alpha=1}^{r+3} (-1)^{\alpha+1} S^{1|1,\dots,\hat{\alpha},\dots,r+3}_{1,\dots,r+2} \otimes x^\alpha_1. \end{aligned}$$

Here $S^{1,\dots,r+2}_{1|1,\dots,\hat{\beta},\dots,r+2}$, $S^{1|1,\dots,\hat{\alpha},\dots,r+3}_{1,\dots,r+2}$ are defined in the same way as the corresponding highest weight vectors.

Proposition 10.2.3.2. *The highest weight vector of* $S_{p+1,1^{r+q}}E \otimes S_{q+1,1^{r+p}}F \subset M_{p+q+1,r+p+q+1}$ *is*

$$S^{1^p|1,\dots,r+q+1}_{1^q|1,\dots,r+p+1} = \sum_{\substack{I \subset [r+q+1], |I|=q, \\ J \subset [r+p+1], |J|=p}} (-1)^{|I|+|J|} M^{1,\dots,\hat{i}_1,\dots,\hat{i}_q,\dots,(r+q+1)}_{1,\dots,\hat{j}_1,\dots,\hat{j}_p,\dots,(r+p+1)} \otimes (x^1_{j_1} \wedge \dots \wedge x^1_{j_p} \wedge x^{i_1}_1 \wedge \dots \wedge x^{i_q}_1).$$

A hatted index is one that is omitted from the summation.

Proof. It is clear the expression has the correct weight and is a highest weight vector, and that it lies in $S^{r+1}V \otimes \Lambda^{p+q}V$. We now show it maps to zero under the differential.

Under the map $d^T : S^{r+1}V\otimes\Lambda^{p+q}V \to S^rV\otimes\Lambda^{p+q+1}V$, the element $S^{1^p|1,\dots,r+q+1}_{1^q|1,\dots,r+p+1}$ maps to:

$$\sum_{\substack{I\subset[r+q+1],|I|=q,\\ J\subset[r+p+1],|J|=p}} (-1)^{|I|+|J|}[\sum_{\alpha\in I}(-1)^{p+\alpha}M^{1,\dots,\hat{i}_1,\dots,\hat{i}_q,\dots,(r+q+1)}_{1,\dots,\hat{j}_1,\dots,\hat{j}_p,\dots,(r+p+1)}x^{i_\alpha}_1\otimes(x^1_{j_1}\wedge\cdots\wedge x^1_{j_p}\wedge x^{i_1}_1\wedge\cdots\wedge\hat{x}^{i_\alpha}_1\wedge\cdots\wedge x^{i_q}_1$$

$$+\sum_{\beta\in J}(-1)^{\beta}M^{1,\dots,\hat{i}_1,\dots,\hat{i}_q,\dots,(r+q+1)}_{1,\dots,\hat{j}_1,\dots,\hat{j}_p,\dots,(r+p+1)}x^1_{j_\beta}\otimes(x^1_{j_1}\wedge\cdots\wedge\hat{x}^1_{j_\beta}\wedge\cdots\wedge x^1_{j_p}\wedge x^{i_1}_1\wedge\cdots\wedge x^{i_q}_1$$

Fix I and all indices in J but one, call the resulting index set J', and consider the resulting term

$$\sum_{\beta\in[r+p+1]\backslash J'}(-1)^{f(\beta,J')}M^{1,\dots,\hat{i}_1,\dots,\hat{i}_q,\dots,(r+q+1)}_{1,\dots,\hat{j}'_1,\dots,\hat{j}'_{p-1},\dots,(r+p+1)}x^1_\beta\otimes(x^1_{j'_1}\wedge\cdots\wedge x^1_{j'_{p-1}}\wedge x^{i_1}_1\wedge\cdots\wedge x^{i_q}_1)$$

where $f(\beta,J')$ equals the number of $j'\in J$ less than β. This term is the Laplace expansion of the determinant of a matrix of size $r+1$ which has its first row appearing twice, and is thus zero. $\square$

Notice that if $q,p>0$, then $S^{1^p|1,\dots,r+q+1}_{1^q|1,\dots,r+p+1}$ is the sum of terms including $S^{1^p|1,\dots,r+q}_{1^{q-1}|1,\dots,r+p+1}\otimes x^{r+q+1}_1$ and $S^{1^{p-1}|1,\dots,r+q+1}_{1^q|1,\dots,r+p}\otimes x^1_{r+p+1}$. This implies the following corollary:

Corollary 10.2.3.3 (Roberts [**?**]). *Each module $S_{a,1^{r+b}}E\otimes S_{b,1^{r+a}}F$, where $a+b=j$ that appears with multiplicity one in $F_{j,j+r}$, appears with multiplicity two in $F_{j-1,j+r}$ if $a,b>0$, and multiplicity one if a or b is zero. The map $F_{j,j+r+1}\to F_{j-1,j+r+1}$ restricted to $S_{a,1^{r+b}}E\otimes S_{b,1^{r+a}}F$, maps non-zero to both $(S_{a-1,1^{r+b}}E\otimes S_{b,1^{r+a-1}}F)\cdot E\otimes F$ and $(S_{a,1^{r+b-1}}E\otimes S_{b-1,1^{r+a}}F)\cdot E\otimes F$.*

Proof. The multiplicities and realizations come from applying the Pieri rule. (Note that if a is zero the first module does not exist and if b is zero the second module does not exist.) That the maps to each of these is non-zero follows from the remark above. $\square$

Remark 10.2.3.4. In [**?**] it is proven more generally that all the natural realizations of the irreducible modules in M_j have non-zero maps onto every natural realization of the module in F_{j-1}. Moreover, the constants in all the maps are determined explicitly. The description of the maps is different than the one presented here.

10.2.4. Proof of Theorem 10.2.1.1. This subsection is less elementary and can be safely skipped. The variety $\hat{\sigma}_r$ admits a desingularization by the geometric method of [**Wey03**], namely consider the Grassmannian $G(r,E^*)$ and the vector bundle $p:\mathcal{S}\otimes F\to G(r,E^*)$ whose fiber over $x\in G(r,E^*)$ is

$x{\otimes}F$. (Although we are breaking symmetry here, it will be restored in the end.) The total space admits the interpretation as the incidence variety

$$\{(x,\phi)\in G(r,E^*)\times \mathrm{Hom}(F,E^*) \mid \phi(F)\subseteq x\},$$

and the projection to $\mathrm{Hom}(F,E^*)=E^*{\otimes}F^*$ has image $\hat{\sigma}_r$. One also has the exact sequence

$$0\to \mathcal{S}{\otimes}F^* \to \underline{E^*{\otimes}F^*} \to Q{\otimes}F^* \to 0$$

where $\underline{E^*{\otimes}F^*}$ denotes the trivial bundle with fiber $E^*{\otimes}F^*$ and $Q=\underline{E^*}/\mathcal{S}$ is the quotient bundle. As explained in [**Wey03**], letting $q:\mathcal{S}{\otimes}F^*\to E^*{\otimes}F^*$ denote the projection, q is a desingularization of $\hat{\sigma}_r$, the higher direct images $\mathcal{R}_iq^*(\mathcal{O}_{\mathcal{S}\otimes F^*})$ are zero for $i>0$, and so by [**Wey03**, Thm. 5.12,5.13] one concludes $F_i=M_i\cdot Sym(E{\otimes}F)$ where

$$\begin{aligned} M_i &= \oplus_{j\geq 0}H^j(G(r,E^*),\Lambda^{i+j}(\mathcal{Q}^*{\otimes}F))\\ &= \oplus_{j\geq 0}\oplus_{|\pi|=i+j} H^j(G(r,E^*),S_\pi Q){\otimes}S_{\pi'}F \end{aligned}$$

One now uses the Bott-Borel-Weil theorem to compute these cohomology groups. An algorithm for this is given in [**Wey03**, Rem. 4.1.5]: If $\pi=(p_1,\dots,p_q)$ (where we must have $p_1\leq n$ to have $S_{\pi'}F$ non-zero, and $q\leq n-r$ as $\mathrm{rank}Q=n-r$), then $S_\pi Q^*$ is the vector bundle corresponding to the sequence

$$(0^r,p_1,\dots,p_{n-r}). \tag{10.2.5}$$

The dotted Weyl action by $\sigma_i=(i,i+1)\in\mathfrak{S}_n$ is

$$\sigma_i\cdot(\alpha_1,\dots,\alpha_n)=(\alpha_1,\dots,\alpha_{i-1},\alpha_{i+1}-1,\alpha_i+1,\alpha_{i+2},\dots,\alpha_n)$$

and one applies simple reflections to try to transform α to a partition until one either gets a partition after u simple reflections, in which case H^u is equal to the module associated to the partition one ends up with and all other cohomology groups are zero, or one ends up on a wall of the Weyl chamber, i.e., at one step one has $(\beta_1,\dots,\beta_n)$ with some $\beta_{i+1}=\beta_i+1$, in which case there is no cohomology.

In our case, we need to move p_1 over to the first position in order to obtain a partition, which means we need $p_1\geq r+1$, and then if $p_2<2$ we are done, otherwise we need to move it etc... The upshot is we can get cohomology only if there is an s such that $p_s\geq r+s$ and $p_{s+1}<s+1$, in which case we get

$$S_{(p_1-r,\dots,p_s-r,s^r,p_{s+1},\dots,p_{n-r})}E{\otimes}S_{\pi'}F$$

contributing to H^{rs}. Say we are in this situation, then write $(p_1-r-s,\dots,p_s-r-s)=\alpha$, $(p_{s+1},\dots,p_{n-r})=\beta'$, so

$$(p_1-r,\dots,p_s-r,s^r,p_{s+1},\dots,p_{n-r})=(s^{r+s})+(\alpha,0^r,\beta')$$

and moreover we may write

$$\pi' = (s^{r+s}) + (\beta, 0^r, \alpha')$$

proving Theorem 10.2.1.1. The case $s = 1$ gives the linear strand of the resolution.

10.3. On the minimal free resolution of the ideal generated by sub-permanents

Let $E, F = \mathbb{C}^n$, $V = E \otimes F$, and let $\mathcal{I}_\kappa^{\mathrm{perm}_n,\kappa} \subset S^\kappa(E \otimes F)$ denote the span of the sub-permanents of size κ and let $\mathcal{I}^{perm_\kappa} \subset Sym(E \otimes F)$ denote the ideal it generates. Note that $\dim \mathcal{I}_\kappa^{perm_\kappa} = \binom{n}{\kappa}^2$. Fix complete flags $0 \subset E_1 \subset \cdots \subset E_n = E$ and $0 \subset F_1 \subset \cdots \subset F_n = F$. Write $\mathfrak{S}_{E_j}$ for the copy of $\mathfrak{S}_j$ acting on E_j and similarly for F.

Write $T_E \subset SL(E)$ for the maximal torus (diagonal matrices). By [**MM62**], the subgroup G_{perm_n} of $GL(E \otimes F)$ preserving the permanent is $[(T_E \times \mathfrak{S}_E) \times (T_F \times \mathfrak{S}_F)] \ltimes \mathbb{Z}_2$, divided by the image of the n-th roots of unity.

As an $\mathfrak{S}_{E_n} \times \mathfrak{S}_{F_n}$-module the space $\mathcal{I}_\kappa^{\mathrm{perm}_n,\kappa}$ decomposes as

(10.3.1)
$$Ind_{\tilde{\mathfrak{S}}_{E_\kappa} \times \tilde{\mathfrak{S}}_{F_\kappa}}^{\mathfrak{S}_{E_n} \times \mathfrak{S}_{F_n}} \widetilde{[\kappa]}_{E_\kappa} \otimes \widetilde{[\kappa]}_{F_\kappa} = ([n]_E \oplus [n-1,1]_E \oplus \cdots \oplus [n-\kappa,\kappa]_E) \otimes ([n]_F \oplus [n-1,1]_F \oplus \cdots \oplus [n-\kappa,\kappa]_F).$$

10.3.1. The linear strand.

Example 10.3.1.1. The space of linear syzygies $M_{2,\kappa+1} := \ker(\mathcal{I}_\kappa^{\mathrm{perm}_n,\kappa} \otimes V \to S^{\kappa+1}V)$ is the $\mathfrak{S}_{E_n} \times \mathfrak{S}_{F_n}$-module

$$M_{2,\kappa+1} = Ind_{\tilde{\mathfrak{S}}_{E_{\kappa+1}} \times \tilde{\mathfrak{S}}_{F_{\kappa+1}}}^{\mathfrak{S}_{E_n} \times \mathfrak{S}_{F_n}} (\widetilde{[\kappa+1]}_{E_{\kappa+1}} \otimes \widetilde{[\kappa,1]}_{F_{\kappa+1}} \oplus \widetilde{[\kappa,1]}_{E_{\kappa+1}} \otimes \widetilde{[\kappa+1]}_{F_{\kappa+1}}).$$

This module has dimension $2\kappa\binom{n}{\kappa+1}^2$. A spanning set for it may be obtained geometrically as follows: for each size $\kappa+1$ sub-matrix, perform the permanental "tensor Laplace expansion" along a row or column, then perform a second tensor Laplace expansion about a row or column and take the difference. An independent set of such for a given size $\kappa+1$ sub-matrix may be obtained from the expansions along the first row minus the expansion along the j-th for $j = 2, \ldots, \kappa+1$, and then from the expansion along the first column minus the expansion along the j-th, for $j = 2, \ldots, \kappa+1$.

Remark 10.3.1.2. Compare this with the space of linear syzygies for the determinant, which has dimension $\frac{2\kappa(n+1)}{n-\kappa}\binom{n}{\kappa+1}^2$. The ratio of their sizes is $\frac{n+1}{n-\kappa}$, so, e.g., when $\kappa \sim \frac{n}{2}$, the determinant has about twice as many linear syzygies, and if κ is close to n, one gets nearly n times as many.

Theorem 10.3.1.3. $\dim M_{j+1,\kappa+j} = \binom{n}{\kappa+j}^2\binom{2(\kappa+j-1)}{j}$. *As an $\mathfrak{S}_n\times\mathfrak{S}_n$-module,*

(10.3.2)
$$M_{j+1,\kappa+j} = Ind^{\mathfrak{S}_{E_n}\times\mathfrak{S}_{F_n}}_{\tilde{\mathfrak{S}}_{E_{\kappa+j}}\times\tilde{\mathfrak{S}}_{F_{\kappa+j}}}\Big(\bigoplus_{a+b=j}[\widetilde{\kappa+b,1^a}]_{E_{\kappa+j}}\otimes[\widetilde{\kappa+a,1^b}]_{F_{\kappa+j}}\Big).$$

The $\binom{n}{\kappa+j}^2$ is just the choice of a size $\kappa+j$ submatrix, the $\binom{2(\kappa+j-1)}{j}$ comes from choosing a set of j elements from the set of rows union columns. Naïvely there are $\binom{2(\kappa+j)}{j}$ choices but there is redundancy as with the choices in the description of M_2.

Proof. The proof proceeds in two steps. We first get "for free" the minimal free resolution of the ideal generated by $S^\kappa E\otimes S^\kappa F$. Write the generating modules of this resolution as $\tilde{M}_j$. We then locate the generators of the linear strand of the minimal free resolution of our ideal, whose generators we denote $M_{j+1,\kappa+j}$, inside $\tilde{M}_{j+1,\kappa+j}$ and prove the assertion.

To obtain $\tilde{M}_{j+1}$, we use the involution ω on the space of symmetric functions (see, e.g. [**Mac95**, §I.2]) that takes the Schur function s_π to $s_{\pi'}$. This involution extends to an endofunctor of $GL(V)$-modules and hence of $GL(E)\times GL(F)$-modules, taking $S_\lambda E\otimes S_\mu F$ to $S_{\lambda'}E\otimes S_{\mu'}F$ (see [**AW07**, §2.4]). This is only true as long as the dimensions of the vector spaces are sufficiently large, so to properly define it one passes to countably infinite dimensional vector spaces.

Applying this functor to the resolution (10.2.1), one obtains the resolution of the ideal generated by $S^\kappa E\otimes S^\kappa F\subset S^\kappa(E\otimes F)$. The $GL(E)\times GL(F)$-modules generating the linear component of the j-th term in this resolution are:

(10.3.3)
$$\begin{aligned}\tilde{M}_{j,j+\kappa-1} &= \bigoplus_{a+b=j-1} S_{(a,1^{\kappa+b})'}E\otimes S_{(b,1^{\kappa+a})'}F\\ &= \bigoplus_{a+b=j-1} S_{(\kappa+b+1,1^{a-1})}E\otimes S_{(\kappa+a+1,1^{b-1})}F.\end{aligned}$$

Moreover, by Corollary 10.2.3.3 and functoriality, the map from $S_{(\kappa+b+1,1^{a-1})}E\otimes S_{(\kappa+a+1,1^{b-1})}F$ into $\tilde{M}_{j-1,j+\kappa-1}$ is non-zero to the copies of $S_{(\kappa+b+1,1^{a-1})}E\otimes S_{(\kappa+a+1,1^{b-1})}F$ in

$$(S_{\kappa+b,1^{a-1}}E\otimes S_{\kappa+a+1,1^{b-2}F})\cdot(E\otimes F) \text{ and } (S_{\kappa+b+1,1^{a-2}}E\otimes S_{\kappa+a,1^{b-1}F})\cdot(E\otimes F),$$

when $a,b>0$.

Inside $S^\kappa E\otimes S^\kappa F$ is the ideal generated by the sub-permanents (10.3.1) which consists of the weight spaces $(p_1,\dots,p_n)\times(q_1,\dots,q_n)$, where all p_i,q_j are either zero or one. (Each sub-permanent has such a weight, and, given such a weight, there is a unique sub-permanent to which it corresponds.)

Call such a weight space *regular*. Note that the set of regular vectors in any $E^{\otimes m}{\otimes}F^{\otimes m}$ (where $m \leq n$ to have any) spans a $\mathfrak{S}_E \times \mathfrak{S}_F$-submodule.

The linear strand of the j-the term in the minimal free resolution of the ideal generated by (10.3.1) is thus a $\mathfrak{S}_E \times \mathfrak{S}_F$-submodule of $\tilde{M}_{j,j+\kappa-1}$. We claim this sub-module is the span of the regular vectors. In other words:

Lemma 10.3.1.4. $M_{j+1,\kappa+j} = (\tilde{M}_{j+1,\kappa+j})_{reg}$.

Assuming Lemma 10.3.1.4, Theorem 10.3.1.3 follows because if π is a partition of $\kappa+j$ then the weight $(1,\dots,1)$ subspace of $S_\pi E_{\kappa+j}$, considered as an $\mathfrak{S}_{E_{\kappa+j}}$-module, is $[\pi]$ (see, e.g., [**Gay76**]), and the space of regular vectors in $S_\pi E{\otimes}S_\mu F$ is $Ind_{\tilde{\mathfrak{S}}_{E_{\kappa+j}}\times\tilde{\mathfrak{S}}_{F_{\kappa+j}}}^{\mathfrak{S}_E\times\mathfrak{S}_F} \widetilde{[\pi]}_E{\otimes}\widetilde{[\mu]}_F$. □

Before proving Lemma 10.3.1.4 we establish conventions for the inclusions $S_{q+1,1^p}E \subset S_{q+1,1^{p-1}}E{\otimes}E$ and $S_{q+1,1^p}E \subset S_{q,1^p}E{\otimes}E$.

Let $\Theta(p,q) : S_{q+1,1^p}E \to S_{q+1,1^{p-1}}E{\otimes}E$ be the $GL(E)$-module map defined such that the following diagram commutes:

$$\begin{array}{ccc} S^qE{\otimes}\Lambda^{p+1}E & \to & S_{q+1,1^p}E \\ \downarrow & & \downarrow \Theta(p,q) \\ S^qE{\otimes}E{\otimes}\Lambda^pE & \to & S_{q+1,1^{p-1}}E{\otimes}E \end{array},$$

where the left vertical map is the identity tensored with the polarization $\Lambda^{p+1}E \to \Lambda^pE{\otimes}E$.

We define two $GL(E)$-module maps $S^qE{\otimes}\Lambda^{p+1}E \to S^{q-1}E{\otimes}E{\otimes}\Lambda^{p+1}E$: σ_1, which is the identity on the second component and polarization on the first, i.e. $S^qE \to S^{q-1}E{\otimes}E$, and σ_2, which is defined to be the composition of

$$S^qE{\otimes}\Lambda^{p+1}E \to (S^{q-1}E{\otimes}E){\otimes}(\Lambda^pE{\otimes}E) \to (S^{q-1}E{\otimes}E){\otimes}(\Lambda^pE{\otimes}E) \to S^{q-1}E{\otimes}E{\otimes}\Lambda^{p+1}E$$

where the first map is two polarizations, the second map swaps the two copies of E and the last is the identity times skew-symmetrization. Let $\Sigma(p,q)$: $S_{q+1,1^p}E \to S_{q+1,1^{p-1}}E{\otimes}E$ denote the unique (up to scale) $GL(E)$-module inclusion (unique because $S_{q+1,1^p}E$ has multiplicity one in $S_{q+1,1^{p-1}}E{\otimes}E$). A short calculation shows that the following diagram is commutative:

$$\begin{array}{ccc} S^qE{\otimes}\Lambda^{p+1}E & \to & S_{q+1,1^p}E \\ \sigma_2 - p\sigma_1 \downarrow & & \downarrow \Sigma(p,q) \\ S^{q-1}E{\otimes}E{\otimes}\Lambda^{p+1}E & \to & S_{q,1^p}E{\otimes}E. \end{array}$$

Proof of Lemma 10.3.1.4. We work by induction, the case $j=1$ was discussed above. Assume the result has been proven up to $M_{j,\kappa+j-1}$ and consider $M_{j+1,\kappa+j}$. It must be contained in $M_{j,\kappa+j-1}{\otimes}(E{\otimes}F)$, so all its weights are either regular, or such that one of the p_i's is 2, and/or one of the q_i's is 2, and all other p_u, q_u are zero or 1. Call such a weight *sub-regular*. It remains to show that no linear syzygy with a sub-regular weight can appear. To do this we show that no sub-regular weight vector in $(M_{j,\kappa+j})_{subreg}$ maps to zero in $(M_{j-1,\kappa+j-1})_{reg} \cdot (E \otimes F)$.

First consider the case where both the E and F weights are sub-regular, then (because the space is a $\mathfrak{S}_E \times \mathfrak{S}_F$-module), the weight $(2,1,\dots,1,0,\dots,0) \times (2,1,\dots,1,0,\dots,0)$ must appear in the syzygy. But the only way for this to appear is to have a term of the form $T \cdot x_1^1$, which cannot map to zero because, since x_1^1 is a non-zero-divisor in $Sym(V)$, our syzygy is a syzygy of degree zero multiplied by x_1^1. But by minimality no such syzygy exists.

Finally consider the case where there is a vector of weight $(2, 1^{j+\kappa-2}) \times (1^{j+\kappa})$ appearing. Consider the set of vectors of this weight as a module for $\mathfrak{S}_{j+\kappa-2} \times \mathfrak{S}_{j+\kappa}$. This module is

$$\bigoplus_{a+b=j} [\kappa+a, 1^b]/[2] \otimes [\kappa+b, 1^a]. \tag{10.3.4}$$

Here

$$[\kappa+a, 1^b]/[2] = [\kappa+a-2, 1^b] \oplus [\kappa+a-1, 1^{b-1}]$$

is called a *skew Specht module*.

By Howe-Young duality and Corollary 10.2.3.3 if $a, b > 0$, $S_{\kappa+a,1^b}E \otimes S_{\kappa+b,1^a}F \subset M_{a+b+1,\kappa+a+b}$ maps non-zero to the two distinguished copies of the same module in $M_{a+b,\kappa+a+b}$. This in turn implies that the two distinguished copies of $S_{\kappa+a,1^b}E \otimes S_{\kappa+b,1^a}F \subset M_{a+b,\kappa+a+b}$, each map non-zero to $M_{a+b-1,\kappa+a+b}$.

The module (10.3.4) will take image inside

$$\bigoplus_{c+d=j-1} Ind^{\mathfrak{S}_{j+\kappa-1} \times \mathfrak{S}_{j+\kappa+1}}_{(\mathfrak{S}_{j+\kappa-2} \times \mathfrak{S}_1) \times (\mathfrak{S}_{j+\kappa} \times \mathfrak{S}_1)} ([\kappa+c, 1^d]/[2] \otimes [1]) \otimes ([\kappa+d, 1^c] \otimes [1]).$$

Fix a term $[\kappa+b, 1^a]$ on the right hand side and examine the map on the left hand side. It is a map

$$[\kappa+a, 1^b]/[2] \to Ind^{\mathfrak{S}_{j+\kappa-1}}_{\mathfrak{S}_{j+\kappa-2} \times \mathfrak{S}_1}([\kappa+a, 1^{b-1}]/[2] \otimes [1]) \oplus Ind^{\mathfrak{S}_{j+\kappa-1}}_{\mathfrak{S}_{j+\kappa-2} \times \mathfrak{S}_1}([\kappa+a-1, 1^b]/[2] \otimes [1]).$$

If $b > 0$, the map to the first summand is the restriction of the map $\Theta(b, \kappa+a) : S_{\kappa+a+1,1^b}E \to S_{\kappa+a+1,1^{b-1}}E \otimes E$, and, due to the fact that it has to map to a sub-regular weight, there is no polarization because the basis vector e_1 has to stay on the left hand side. So the map is the identity, thus injective.

It remains to show that for $b = 0$, the map corresponding to the summand $b = 0, a = j$ which is the restriction of the injective map $\Sigma(0, \kappa+j-2) : S_{\kappa+j-1}E \to S_{\kappa+j-2}E \otimes E$ tensored with the map $\Theta(j-1, \kappa)$ injects into the cokernel of the summand corresponding to $c = 0, d = j-1$ modulo the image of the map coming from the summand $a = 1, b = j-1$. Both modules consist of just two irreducible $\mathfrak{S}_{E_{j+\kappa-1}} \times \mathfrak{S}_{F_{j+\kappa-1}}$-modules and, using formulas for Σ and Θ, the map is injective. This concludes the proof. $\square$

Example 10.3.1.5. For small n and κ, computer computations show no additional first syzygies on the $\kappa \times \kappa$ sub-permanents of a generic $n \times n$ matrix (besides the linear syzygies) in degree less than the Koszul degree 2κ. For example, for $\kappa = 3$ and $n = 5$, there are 100 cubic generators for the ideal and 5200 minimal first syzygies of degree six. There can be at most $\binom{100}{2} = 4950$ Koszul syzygies, so there must be additional non-Koszul first syzygies.

10.4. Young-flattenings and the cactus variety

10.5. The Hilbert scheme of points

10.6. Lower rank bounds

To go from border rank lower bounds to rank lower bounds, we examine not just the rank of flattenings, but the nature of the kernel.

For $f \in S^dV$, define the *apolar ideal of* f, $f^{ann} \subset Sym(V^*)$, the set of $P \in Sym(V^*)$ such that $P(f) = 0$. In other words, $f^{ann} = \oplus_{j=1}^{d} \ker f_{j,d-j} \oplus_{k=d+1}^{\infty} S^kV^*$.

The following lemma is critical:

Lemma 10.6.0.1 (Apolarity Lemma)**.** *$f \in \operatorname{span}\{\ell_1^d, \dots, \ell_r^d\} \subset S^dV$ if and only if $f^{ann} \supseteq I([\ell_1^d] \sqcup \cdots \sqcup [\ell_r^d])$.*

Note that $f^{ann} \supseteq I([\ell_1^d] \sqcup \cdots \sqcup [\ell_r^d])$ says that for all j, $\ker f_{j,d-j} \supseteq S^j({\ell_1}^{\perp} \cap \cdots \cap {\ell_r}^{\perp})$.

Exercise 10.6.0.2: Prove the apolarity lemma.

For ideals $I, J \subset Sym(V^*)$, introduce the ideal $I : J := \{P \in Sym(V^*) \mid PJ \subseteq I\}$.

Exercise 10.6.0.3: Prove that if $X = \operatorname{Zeros}(I)$ and $Y = \operatorname{Zeros}(J)$ are reduced, then $I : J$ is the ideal of polynomials vanishing on the set $X \backslash (X \cap Y)$.

Theorem 10.6.0.4. [**CCC$^+$15a**] *For $f \in S^dV$, and $L \in V^* \backslash \{0\}$,*

$$\mathbf{R}_S(f) \geq \sum_s \operatorname{Hilb}_s \left(\frac{Sym(V^*)}{(f^{ann} : (L)) + (L)} \right).$$

Remark 10.6.0.5. In [**CCC$^+$15a**] they prove a more general statement allowing arbitrary ideals generated in a single degree instead of just the linear form L.

Proof. For $J \subset Sym(V^*)$ a homogeneous ideal and $L \in V^*$ that is not a zero divisor in $Sym(V^*)/J$, we have an exact sequence

$$0 \to Sym(V^*)/J \to Sym(V^*)/J \to Sym(V^*)/[J + (L)] \to 0$$

where the first map is multiplication by L. In degree i we have

$$0 \to S^{i-1}V^*/J_{i-1} \to S^iV^*/J_i \to S^iV^*/[J + (L)]_i \to 0.$$

Thus for all s,

$$\dim(S^sV^*/J_s) = \sum_{i=0}^{s} \dim(S^iV^*/[J + (L)]_i).$$

If Zeros(J) is zero-dimensional and reduced, and $s >> 0$, then the Hilbert function of $Sym(V^*)/J$ will equal the number of points in Zeros(J).

In our setting $\mathbf{R}_S(f) = r$, $f = \ell_1^d + \cdots + \ell_r^d$, $X = \{[\ell_1], \ldots, [\ell_r]\} \subset \mathbb{P}V$. For any L, choose $J = I(X) : L$, so that by Exercise 10.6.0.3 J is the ideal of points on X not on the hyperplane determined by L.

Claim: L is not a zero divisor in $Sym(V^*)/J$. To see this, say $t \notin J$ and $t \cdot L \in J$. Then $t \cdot L^2 \in I(X)$, but $I(X)$ is reduced so $t \cdot L \in I(X)$ which means $t \in J$.

Putting it all together, assuming X gives a minimal decomposition of f,

$$\begin{aligned}
\mathbf{R}_S(f) &= \#X \\
&\geq \#\mathrm{Zeros}(J) = \sum_{i=0}^{s} \mathrm{Hilb}(Sym(V^*)/[I(X) : L + (L)], i) \\
&\geq \sum_{i=0}^{s} \mathrm{Hilb}_i\left(\frac{Sym(V^*)}{(f^{ann} : (L)) + (L)}\right).
\end{aligned}$$

where the last assertion is by the apolarity lemma. □

Returning to the study of elementary symmetric polynomials, take $L = \frac{\partial}{\partial x_n}$. Let $V' = L^\perp \subset V$. Then

$$\begin{aligned}
(e_{d,n}^{ann} : L) + (L) &= (\frac{\partial}{\partial x_n} e_{d,n})^{ann} + (\frac{\partial}{\partial x_n}) \\
&= e_{d,n-1}^{ann,V'} + (\frac{\partial}{\partial x_n})
\end{aligned} \tag{10.6.1}$$

where $e_{d,n-1}^{ann,V'}$ (resp. $e_{d,n-1}^{ann,V}$) is $e_{d,n-1}$ considered as an element of S^dV' (resp. S^dV). On the other hand, $\frac{\partial}{\partial x_n} \in e_{d,n-1}^{ann,V}$, so (10.6.1) equals $e_{d-1,n-1}^{ann,V}$.

Since $(S^jV^*/(e_{d-1,n-1}^{ann})_j)$ may be identified with the space square free monomials in degree j, we conclude:

Theorem 10.6.0.6. [**Lee16**] *For d odd,*

$$\mathbf{R}_S(e_{d,n}) = \sum_{j=0}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{j}.$$

In the case of even degree, one has a similar expression to Theorem 7.1.2.3 for $e_{d,n}$ with $2^{\frac{d}{2}}$ summands. The lower bound was independent of parity, so we get

Theorem 10.6.0.7. [**Lee16**] *For d even,*

$$\sum_{j=0}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{j} - \binom{n-1}{\frac{d}{2}} \leq \mathbf{R}_S(e_{d,n}) \leq \sum_{j=0}^{\lfloor \frac{d}{2} \rfloor} \binom{n}{j}.$$

Hints and Answers to Selected Exercises

Chapter 1.

1.1.15.1 In general, the trilinear map associated to a bilinear form is $(u, v, \gamma) \mapsto \gamma(T(u, v))$. Let z^{*u}_v denote the linear form that eats a matrix and returns its (u, v)-th entry. Since $(XY)^i_k = \sum_j x^i_j y^j_k$, the associated trilinear map is $(X, Y, z^{*u}_v) \mapsto \sum_j x^u_j y^j_v$. On the other hand, $\text{trace}(XYZ) = \sum_{i,j,k} x^i_j y^j_k z^k_i$. Now observe that both these agree, e.g., on basis vectors.

Chapter 2.

2.1.1.1 $v \in V$ goes to the map $\beta \mapsto \beta(v)$.

2.1.1.4 For the second assertion, a generic matrix will have nonzero determinant. In general, the complement to the zero set of any polynomial over the complex numbers has full measure. For the last assertion, first say $\text{rank}(f) = r' \leq r$ and let $v_1, \dots, v_{\mathbf{v}}$ be a basis of V such that the kernel is spanned by the last $\mathbf{v} - r'$ vectors. Then the matrix representing f will be nonzero only in the upper $r' \times r'$ block and thus all minors of size greater than r' will be zero. Next say $\text{rank}(f) = s > r$. Taking basis in the same manner, we see the upper right size s submatrix will have a nonzero determinant. Taking a Laplace expansion, we see at least one size $r + 1$ minor of it is nonzero. In any other choice of basis minors expressed in the new basis are linear combinations of minors expressed in the old, so we conclude. If you need help with the third assertion, use Proposition 3.1.6.1.

2.1.1.7 $\text{trace}(f)$.

2.1.2.1 A multi-linear map is determined by its action on bases of $A_1^*, \dots, A_n^*$.

2.1.2.4 See [**Lan12**, §2.4.4]

2.1.5.4 See §3.1.6.

2.1.6.1 For example, take $a_1{\otimes}b_1{\otimes}c_2+a_1{\otimes}b_2{\otimes}c_1+a_2{\otimes}b_1{\otimes}c_1+\sum_{j=3}^r a_j{\otimes}b_j{\otimes}c_j$.

2.1.6.2 If $T=\sum_{i=1}^r a_i{\otimes}b_i{\otimes}c_i$, then, letting $\pi_A : A\to A/(A')^{\perp}$ be the projection, and similarly for B,C, then $T_{A'\otimes B'\otimes C'}=\sum_{i=1}^r \pi_A(a_i){\otimes}\pi_B(b_i){\otimes}\pi(c_i)$.

2.1.7.2 First assume $\underline{\mathbf{R}}(T)=\mathbf{R}(T)$ and write $T=a_1{\otimes}b_1{\otimes}c_1+\cdots+a_r{\otimes}b_r{\otimes}c_r$. Then $T(A^*)=\mathrm{span}\{b_1{\otimes}c_1,\dots,b_r{\otimes}c_r\}$ so $\underline{\mathbf{R}}(T)\geq \mathrm{rank}T_A$. Now use that ranks of linear maps are determined by polynomials (the minors of the entries) to conclude.

2.2.1.2 Say $T=\sum_{j=1}^{\mathbf{b}} a_j{\otimes}b_j{\otimes}c_j$ and this is an optimal expression. Since T_A is injective, the a_j must be a basis. Let α^j be the dual basis, so $T(\alpha^j)=b_j{\otimes}c_j$ has rank one. These span. In the other direction, say the image is $\mathrm{span}\{b_1{\otimes}c_1,\dots,b_{\mathbf{b}}{\otimes}c_{\mathbf{b}}\}$. then for each j there must be some $\alpha^j\in A^*$ with $T(\alpha^j)=b_j{\otimes}c_j$. Since T_A is injective, these form a basis of A, so we must have $T=\sum_{j=1}^{\mathbf{b}} a_j{\otimes}b_j{\otimes}c_j$ with a_j the dual basis vectors.

2.2.2.2 Use Exercise 2.1.7.4, taking three matrices in A^*, e.g. Id, a matrix with all 1's just below the diagonal and zero elsewhere and a matrix with 1's just above the diagonal and zeros elsewhere.

2.6.1.2 It is sufficient to consider the case $q=p-1$. Say $X\in\ker(T_A^{\wedge p-1})$. Then $a\wedge X\in(T_A^{\wedge p})$ so $a\wedge X=0$ for all $a\in A$. But this is not possible.

2.6.2.1 Recall that for any vector space V, an element $v\in V$ is uniquely specified from knowning how a basis of V^* acts on it. As a bilinear map, the output is $XY\in W^*{\otimes}U$, where in bases, $(XY)^i_j=\sum_k X^i_kY^k_i$. Let $\zeta^s_t\in W{\otimes}U^*$ be the element that eats an element of $W^*{\otimes}U$ in bases and outputs its (s,t)-entry. The ζ^s_t form a basis of $W{\otimes}U^*$. Let $Z^s_t\in(W{\otimes}U^*)^*=W^*{\otimes}U$ denote the dual basis element. Then $\zeta^s_t(XY)=\mathrm{trace}(XYZ^s_t)$.

2.6.2.2 $M_{\langle U,V,W\rangle}\in(U^*{\otimes}V){\otimes}(V^*{\otimes}W){\otimes}(W^*{\otimes}U)\simeq(U^*{\otimes}U){\otimes}(V^*{\otimes}V){\otimes}(W^*{\otimes}W)$. Now re-arrange $\mathrm{Id}_U\otimes\mathrm{Id}_V\otimes\mathrm{Id}_W$ in bases.

2.6.2.4 Use Exercise **??**.2.6.2.3.

2.6.3.3 Extend the a_j to a basis of A and consider the induced basis of $\Lambda^{q+1}A$. Write out $X_j\wedge a_j$ with respect to the induced basis and compare coefficients.

2.6.3.1 Use the variant of the Cartan lemma below the exercise to show it is the whole kernel.

2.6.3.8 Apply the proof of Theorem 2.6.3.6 to $M_{\langle p,p,2\rangle}$.

2.7.2.2 First assume $T = e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$ and take $\mu = e^L$ and $\zeta = e^J$. Then

$$\mu \lrcorner T = \begin{cases} e_{I \backslash L} & \text{if} \quad L \subset I \\ 0 \text{ if } L \not\subset I \end{cases}$$

$$\zeta \lrcorner T = \begin{cases} e^{J \backslash I} & \text{if} \quad I \subset J \\ 0 \text{ if } I \not\subset J \end{cases}$$

and $\langle e^{J\backslash I}, e_{I\backslash L}\rangle = 0$, in fact they have no indices in common. By linearity we get zero for any linear combination of such e^J, e_L's so we see that $G(k,V)$ is in the zero set of the equations. (Any element of $G(k,V)$ is equivalent to $[e_I]$ after a change of basis and our equations are independent of the choice of basis.)

Now for simplicity assume $T = e_{I_1} + e_{I_2}$ where I_1, I_2 have at least one index different. Take $\zeta = e^{I_1 \cup F}$ where $F \subset I_2$, $F \not\subset I_1$ and $I_2 \not\subset I_1 \cup F$. Then $\zeta \lrcorner T = e^F$. Take $\mu = e^{I_2 \backslash F}$ so $\mu \lrcorner T = e_F$. We conclude.

The general case is similar, just with more bookkeeping.

Chapter 3.

3.1.4.3 The ideal is generated by $p_3^2 - p_2p_4, p_2^2 - p_0p_4$. Note that we simply are throwing away the polynomials with p_1. The point p_3, corresponding to the polynomial x^3y is on a tangent line to $v_4(\mathbb{P}^1)$, while the point p_{22}, corresponding to the polynomial x^2y^2 is not.

3.1.4.5 The ideal is generated by $p_2^2 - p_1p_3, p_1p_2 - p_0p_3, p_1^2 - p_0p_2$.

3.2.1.4 Recall from Exercise 2.6.2.4 that $\otimes_j M_{\langle \mathbf{l}_j, \mathbf{m}_j, \mathbf{n}_j\rangle} = M_{\langle \Pi_j \mathbf{l}_j, \Pi_k \mathbf{m}_k, \Pi_l \mathbf{n}_l\rangle}$. Set $N = \mathbf{nml}$ and consider $M_{\langle N\rangle} = M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle} \otimes M_{\langle \mathbf{n},\mathbf{l},\mathbf{m}\rangle} \otimes M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle}$.

3.1.4.6 Say $f(X) = Z_1 \cup Z_2$ and note that $X = f^{-1}(Z_1) \cup f^{-1}(Z_2)$.

3.2.2.1 Consider

$$\begin{pmatrix} \heartsuit & \heartsuit \\ \heartsuit & \spadesuit \\ \spadesuit & \spadesuit \end{pmatrix}$$

3.3.1.3 Since the border rank of points in $GL(A) \times GL(B) \times GL(C) \cdot T$ equals the border rank of T, the border rank of points in the closure cannot increase.

3.4.9.2 Instead of the curve $a_0 + ta_1$ use $a_0 + ta_1 + t^2 a_{q+1}$ and similarly for b, c.

3.4.6.3 Use Proposition 3.2.1.8.

3.5.3.3 When writing $T = \lim_{t\to 0} T(t)$ we may take $t \in \mathbb{Z}_{h+1}$.

3.5.3.4 If we are multiplying polynomials of degrees d_1 and d_2, then their product has degree d_1d_2, so the answer is the same as if we were working over $\mathbb{Z}_{d_1d_2}$.

Chapter 4.

4.3.2.2 If one uses the images of the standard basis vectors, one gets:

$$M_{\langle 2\rangle} = \begin{pmatrix}0 & -1\\ 1 & -1\end{pmatrix}^{\otimes 3} + \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}^{\otimes 3} + \begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix}^{\otimes 3} + \begin{pmatrix}0 & 0\\ -1 & 1\end{pmatrix}^{\otimes 3} + \langle \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} \otimes \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} \otimes \begin{pmatrix}1 & -1\\ 1 & -1\end{pmatrix} \rangle_{\mathbb{Z}_3}.$$

?? Note that while for $V^{\otimes 3}$, the kernels of $S^2V{\otimes}V \to S^3V$ and $\Lambda^2V{\otimes}V \to \Lambda^3V$ were isomorphic $GL(V)$-modules, the kernels of $S^3V{\otimes}V \to S^4V$ and $\Lambda^3V{\otimes}V \to \Lambda^4V$ are not. One can avoid dealing with spaces like $S_{21}V{\otimes}V$ by using maps like, e.g. the kernel of $S^2V{\otimes}S^2V \to S^4V$ and keeping track of dimensions of spaces uncovered. The answer is given by Theorem 8.7.1.2.

?? If $a_0 = b_0 = c_0 = Id$ then $({u_1}^{\perp}{\otimes}v_3){\otimes}({v_2}^{\perp}{\otimes}w_1){\otimes}({w_3}^{\perp}{\otimes}u_2)$ is mapped to $(u_2^T{\otimes}({w_3}^{\perp})^T){\otimes}(w_1^{T\perp}{\otimes}({v_2}^{\perp})^T){\otimes}(v_3^T{\otimes}({u_1}^{\perp})^T)$. The general case is just notationally more cumbersome.

?? See [**CHI**$^+$]. The calculation is a little more involved than indicated in the section.

Chapter 5.

5.1.4.4 First note that if x is generic, it is diagonalizable with distinct eigenvalues so if x is generic, then $\dim C(x) = \mathbf{b}$. Then observe that $\dim(C(x))$ is semi-continuous as the set $\{y \mid \dim C(y) \leq p\}$ is an algebraic variety. Alternatetively, and more painfully, compute the centralizer of elements in Jordan canonical form.

5.2.1.1 If $\underline{\mathbf{R}}(T) = m$, then T is a limit of points T_ϵ with $\mathbb{P}T_\epsilon(A^*) \cap Seg(\mathbb{P}B \times \mathbb{P}C) \neq \emptyset$.

5.3.1.4 For the lower bound use Koszul flattenings, for the upper, write T as the sum of the first AFT tensor and the remainder and bound the border rank of each.

5.3.1.8 For the lower bound, use the substitution method. For the upper, consider the rank decomposition of the structure tensor of $\mathbb{C}[\mathbb{Z}_{2\mathbf{m}-1}]$, which, using the DFT, has rank and border rank $\mathbf{m}$. Show that this tensor degenerates to the tensor corresponding to the centralizer of a regular nilpotent element.

5.4.3.3 Without loss of generality assume $2 \leq i \leq j$. For $j = 2, 3$ the inequality is straightforward to check, so assume $j \geq 4$. Prove the inequality 5.4.2 by induction on $\mathbf{n}$. For $\mathbf{n} = ij$ the inequality follows from the combinatorial interpretation of binomial coefficients and the fact that the middle one is the largest.

We have $\binom{\mathbf{n}+1-1+ij-1}{ij-1} = \binom{\mathbf{n}-1+ij-1}{ij-1}\frac{\mathbf{n}-1+ij}{\mathbf{n}}$, $\binom{\mathbf{n}+1-j+i-1}{i-1} = \binom{\mathbf{n}-j+i-1}{i-1}\frac{\mathbf{n}-j+i}{n-j+1}$ and $\binom{\mathbf{n}+1-i+j-1}{j-1} = \binom{\mathbf{n}-i+j-1}{j-1}\frac{\mathbf{n}-i+j}{\mathbf{n}-i+1}$. By induction it is enough to prove that:

$$\frac{\mathbf{n}-1+ij}{\mathbf{n}} \geq \frac{\mathbf{n}-j+i}{\mathbf{n}-j+1}\frac{\mathbf{n}-i+j}{\mathbf{n}-i+1}. \tag{10.6.2}$$

This is equivalent to:

$$ij-1 \geq \frac{\mathbf{n}(i-1)}{\mathbf{n}-j+1} + \frac{\mathbf{n}(j-1)}{\mathbf{n}-i+1} + \frac{\mathbf{n}(i-1)(j-1)}{(\mathbf{n}-j+1)(\mathbf{n}-i+1)}.$$

As the left hand side is independent from $\mathbf{n}$ and each fraction on the right hand side decreases with growing $\mathbf{n}$, we may set $\mathbf{n} = ij$ in inequality 10.6.2. Thus it is enough to prove:

$$2 - \frac{1}{ij} \geq (1 + \frac{i-1}{ij-j+1})(1 + \frac{j-1}{ij-i+1}).$$

Then the inequality is straightforward to check for $i = 2$, so assume $i \geq 3$.

5.4.3.2 $\binom{\mathbf{n}+j-2}{j-1} = \dim S^{j-1}\mathbb{C}^{\mathbf{n}-1}$ so the sum may be thought of as computing the dimension of $S^{m-1}\mathbb{C}^{\mathbf{n}}$ where each summand represents basis vectors (monomials) where e.g., x_1 appears to the power $m-j$.

5.6.2.1 Show that if $n \in \mathrm{Rad}(\mathcal{A})$ is not nilpotent, then there is some prime ideal of $\mathcal{A}$ not containing n.

5.6.1.4 $\mathcal{A}$ has basis $x_J := x_1^{j_1} \cdots x_n^{j_n}$ with $0 \leq j_s < a_s$. Let e_j be the dual basis. Then $T_{\mathcal{A}} = \sum_{i_s+j_s<a_s} e_I \otimes e_J \otimes x_{I+J}$. Write $x_K^* = x_1^{a_1-k_1-1} \cdots x_n^{a_n-k_n-1}$. Then $T_{\mathcal{A}} = \sum_{i_s+j_s+k_s<a_s} e_I \otimes e_J \otimes e_K$.

Chapter 6.

6.1.4.2 Use that $\frac{1}{1-\lambda t} = \sum_j \lambda^j t^j$.

6.2.3.1 $N^*_M \sigma^0_r = \ker M \otimes (\mathrm{Image}\, M)^\perp = \ker M \otimes \ker M^T \subset U \otimes V^*$. The second equality holds because for a linear map $f : V \to W$, $\mathrm{Image}(f)^\perp = \ker(f^T)$.

?? Recall that for a linear map $f : V \to W$, that $\ker f = (\mathrm{Image}\, f^T)^\perp$.

6.2.2.3 Consider $\lim_{\epsilon \to 0} \frac{1}{\epsilon}((x+\epsilon y)^n - x^n)$.

6.4.2.3 Parametrize C by a parameter s and $\tau(C)$ by s and a parameter for the line.

6.2.2.6 Respectively, taking $k = \lfloor \frac{n}{2} \rfloor$ one gets the ranks are $\binom{n}{\lfloor \frac{n}{2} \rfloor}$, $\binom{n}{\lfloor \frac{n}{2} \rfloor}^2$, and $\binom{n}{\lfloor \frac{n}{2} \rfloor}^2$.

6.3.2.3 The space of matrices with last two columns equal to zero is contained in $Z(\mathrm{perm}_m)_{sing}$.

6.3.3.5 Let $\hat{Q} \in S^2V$ be the corresponding quadratic form (defined up to scale). Take a basis $e_1, \ldots, e_{\mathbf{v}}$ of V such that $e_1, \ldots, e_k$ correspond to a linear space on Q, so $Q(e_s, e_t) = 0$ for $0 \leq s,t \leq k$. But Q being smooth says $\hat{Q}$ is non-degenerate, so for each e_s, there must be some $e_{f(s)}$ with $Q(e_s, e_{f(s)}) \neq 0$.

6.6.1.3 In this case the determinant is a smooth quadric.

6.6.2.1 $\{\mathrm{perm}_2 = 0\}$ is a smooth quadric.

6.4.6.1 First note that perm_m evaluated on a matrix whose entries are all one is $m!$. Then perform a permanental Laplace expansion about the first row.

6.5.2.2 Note that $\frac{\partial R}{\partial x_i} = \sum_j \frac{\partial^2 R}{\partial x_i \partial x_j}$ and now consider the last nonzero column.

Chapter 7.

7.1.2.6 Consider (where blank entries are zero)

$$\det \begin{pmatrix} 0 & & x_1 & & x_2 & & x_3 \\ x_1 & \ell & & & & & \\ & x_1 & \ell & & & & \\ x_2 & & & \ell & & & \\ & & & x_2 & \ell & & \\ x_3 & & & & & \ell & \\ & & & & & x_3 & \ell \end{pmatrix} = \ell^{7-3}(x_1^3 + x_3^2 + x_3^3).$$

7.4.1.4 Take $x_m = x_{m+1} = \cdots = x_N = 0$.

Chapter 8.

8.1.2.2 Say we have a highest weight vector $z \in V^{\otimes d}$ weight $(j_1, \ldots, j_{\mathbf{v}})$ with $j_i < j_{i+1}$. Consider the matrix g that is the identity plus a vector with one non-zero entry in the $(i, i+1)$ slot. Then gz is a non-zero vector of weight $(j_1, \ldots, j_{i-1}, j_{i+1}, j_{i+1} - 1, \ldots, j_{\mathbf{v}})$.

8.2.1.1 By linearity, for any P_1, P_2, the rank of the linear map $U^* \to W$ associated to $P_1 + P_2$ is at most the sum of the ranks of the maps associated to P_1 and P_2.

8.1.5.2 $g \cdot e_1 \wedge \cdots \wedge e_{\mathbf{v}} = \det(g) e_1 \wedge \cdots \wedge e_{\mathbf{v}}$

8.1.4.1 The weight of the one-dimensional representation $\det^{-1}$ is $(-1, \ldots, -1)$.

8.1.4.2 Consider the linear form $v \mapsto \det_{\mathbf{v}}(v_1, \ldots, v_{\mathbf{v}-1}, v)$.

8.4.1.2 A highest weight vector of any copy of $S_\pi V^*$ is constructed skew-symmetrizing over $\ell(\pi)$ vectors. For the other direction, the zero set of any $P \in S^\delta(S^d \mathbb{C}^k)$ is a proper subvariety of $S^d \mathbb{C}^k$.

8.6.8.2 We need $\mathrm{Hom}_{\mathfrak{S}_d}([\pi]^*, [\mu]) \neq 0$. But $[\pi]^* \simeq [\pi]$. By Schur's lemma $\mathrm{Hom}_{\mathfrak{S}_d}([\pi], [\mu]) \neq 0$ if and only if $[\pi] = [\mu]$.

8.5.2.5 Under the action of a basis vector in $\mathfrak{gl}(E\otimes F)$, since it is by Leibnitz rule, at most one variable in each monomial can be changed. So whatever highest weight vectors appear in the tangent space, their weight can differ by at most one in each of E, F from $((1,\ldots,n),(1,\ldots,n))$. But there is only one partition pair with this property that occurs in $S^n(E\otimes F)$, namely $(1^2,2,\ldots,n-1),(1^2,2,\ldots,n-1))$.

8.7.1.3

$$\begin{aligned} S^d(E\otimes F) &= [(E\otimes F)^{\otimes d}]^{\mathfrak{S}_d} \\ &= (E^{\otimes d}\otimes F^{\otimes d})^{\mathfrak{S}_d} \\ &= [(\oplus_{|\pi|=d} S_\pi E\otimes[\pi]))\otimes(\oplus_{|\mu|=d} S_\mu F\otimes[\mu]))]^{\mathfrak{S}_d} \\ &= \oplus_{|\mu|,|\pi|=d} S_\pi E\otimes S_\mu F\otimes([\pi]\otimes[\mu])^{\mathfrak{S}_d} \end{aligned}$$

Now use Exercise 8.6.8.2.

8.6.1.1 Prove an algebra version of Schur's lemma.

8.6.4.8.6.4.2 If V is an irreducible G-module, then $V^*\otimes V$ is an irreducible $G\times G$-module.)

8.7.2.2 $c_{\pi'} = \sum_{\sigma\in\mathfrak{S}^{def}_{\pi'}} \delta_\sigma \sum_{\sigma\in\mathfrak{S}^{def}_{\pi}} \mathrm{sgn}(\sigma)\delta_\sigma$. Now show $c_{(1^d)}c_\pi = c_{\pi'}$.

?? Apply an appropriate lowering operator (i.e., a lower triangular matrix) to the highest weight vector to bring the weight down to zero. The result is a (possibly zero) vector of weight zero. To show that some lowering operator applied to it is non-zero, note that if one reverses indices on a highest weight vector (sending e_j to $e_{\mathbf{v}-j}$) one gets a vector of weight less than zero in the module. But the module is generated by applying lowering operators to a highest weight vector.

?? Use the double commutant theorem.

8.7.2.4 The eigenvalues are $e^{\frac{\pm 2\pi i}{3}}$.

Chapter 9.

9.1.7.2 $Ch_d(\mathbb{C}^2) = \mathbb{P}S^d\mathbb{C}^2$.

9.1.4.6 Highest weight vectors here correspond to partitions with at most d parts.

Bibliography

[ABV15] J. Alper, T. Bogart, and M. Velasco, *A lower bound for the determinantal complexity of a hypersurface*, ArXiv e-prints (2015).

[AFLG15] Andris Ambainis, Yuval Filmus, and François Le Gall, *Fast matrix multiplication: limitations of the Coppersmith-Winograd method (extended abstract)*, STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing, ACM, New York, 2015, pp. 585–593. MR 3388238

[AFT11] Boris Alexeev, Michael A. Forbes, and Jacob Tsimerman, *Tensor rank: some lower and upper bounds*, 26th Annual IEEE Conference on Computational Complexity, IEEE Computer Soc., Los Alamitos, CA, 2011, pp. 283–291. MR 3025382

[AH95] J. Alexander and A. Hirschowitz, *Polynomial interpolation in several variables*, J. Algebraic Geom. **4** (1995), no. 2, 201–222. MR 96f:14065

[Ahl78] Lars V. Ahlfors, *Complex analysis*, third ed., McGraw-Hill Book Co., New York, 1978, An introduction to the theory of analytic functions of one complex variable, International Series in Pure and Applied Mathematics. MR 510197

[AJ15] N. R. Aravind and Pushkar S. Joglekar, *On the expressive power of read-once determinants*, CoRR **abs/1508.06511** (2015).

[AK81] L. A. Aĭzenberg and A. M. Kytmanov, *Multidimensional analogues of Newton's formulas for systems of nonlinear algebraic equations and some of their applications*, Sibirsk. Mat. Zh. **22** (1981), no. 2, 19–30, 235. MR 610765

[Alp14] L. Alpoge, *Square-root cancellation for the signs of Latin squares*, ArXiv e-prints (2014).

[AR03] Elizabeth S. Allman and John A. Rhodes, *Phylogenetic invariants for the general Markov model of sequence mutation*, Math. Biosci. **186** (2003), no. 2, 113–144. MR 2 024 609

[Aro58] S. Aronhold, *Theorie der homogenen Functionen dritten Grades von drei Veränderlichen*, J. Reine Angew. Math. **55** (1858), 97–191. MR 1579064

[AS13] V. B. Alekseev and A. V. Smirnov, *On the exact and approximate bilinear complexities of multiplication of* 4×2 *and* 2×2 *matrices*, Proceedings of the Steklov Institute of Mathematics **282** (2013), no. 1, 123–139.

[AT92] N. Alon and M. Tarsi, *Colorings and orientations of graphs*, Combinatorica **12** (1992), no. 2, 125–134. MR 1179249 (93h:05067)

[AV08] M. Agrawal and V. Vinay, *Arithmetic circuits: A chasm at depth four*, In Proc. 49th IEEE Symposium on Foundations of Computer Science (2008), 6775.

[AW07] Kaan Akin and Jerzy Weyman, *Primary ideals associated to the linear strands of Lascoux's resolution and syzygies of the corresponding irreducible representations of the Lie superalgebra* $\mathbf{gl}(m|n)$, J. Algebra **310** (2007), no. 2, 461–490. MR 2308168 (2009c:17007)

[Bar77] W. Barth, *Moduli of vector bundles on the projective plane*, Invent. Math. **42** (1977), 63–91. MR MR0460330 (57 #324)

[Bas14] Saugata Basu, *A complexity theory of constructible functions and sheaves*, arXiv:1309.5905 (2014).

[BB] Austin R. Benson and Grey Ballard, *A framework for practical parallel fast matrix multiplication*, arXiv:1409.2908.

[BCRL79] Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti, $O(n^{2.7799})$ *complexity for* $n \times n$ *approximate matrix multiplication*, Inform. Process. Lett. **8** (1979), no. 5, 234–235. MR MR534068 (80h:68024)

[BCS97] Peter Bürgisser, Michael Clausen, and M. Amin Shokrollahi, *Algebraic complexity theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 315, Springer-Verlag, Berlin, 1997, With the collaboration of Thomas Lickteig. MR 99c:68002

[Bea00] Arnaud Beauville, *Determinantal hypersurfaces*, Michigan Math. J. **48** (2000), 39–64, Dedicated to William Fulton on the occasion of his 60th birthday. MR 1786479 (2002b:14060)

[BGL13] Jaroslaw Buczyński, Adam Ginensky, and J. M. Landsberg, *Determinantal equations for secant varieties and the Eisenbud-Koh-Stillman conjecture*, J. Lond. Math. Soc. (2) **88** (2013), no. 1, 1–24. MR 3092255

[BGL14] H. Bermudez, S. Garibaldi, and V. Larsen, *Linear preservers and representations with a 1-dimensional ring of invariants*, Trans. Amer. Math. Soc. **366** (2014), no. 9, 4755–4780. MR 3217699

[BILR] Grey Ballard, Christian Ikenmeyer, J.M. Landsberg, and Nick Ryder, *The geometry of rank decompositions of matrix multiplication ii: 3x3 matrices*, preprint.

[Bin80] D. Bini, *Relations between exact and approximate bilinear algorithms. Applications*, Calcolo **17** (1980), no. 1, 87–97. MR 605920 (83f:68043b)

[BIP16] Peter Bürgisser, Christian Ikenmeyer, and Greta Panova, *No occurrence obstructions in geometric complexity theory*, CoRR **abs/1604.06431** (2016).

[BL89] S. C. Black and R. J. List, *A note on plethysm*, European J. Combin. **10** (1989), no. 1, 111–112. MR 977186 (89m:20011)

[BL14] Jaroslaw Buczyński and J. M. Landsberg, *On the third secant variety*, J. Algebraic Combin. **40** (2014), no. 2, 475–502. MR 3239293

[BL16] M. Bläser and V. Lysikov, *On degeneration of tensors and algebras*, ArXiv e-prints (2016).

[Blä00] Markus Bläser, *Lower bounds for the bilinear complexity of associative algebras*, Comput. Complexity **9** (2000), no. 2, 73–112. MR 1809686

[Blä01] ———, *Complete problems for Valiant's class of qp-computable families of polynomials*, Computing and combinatorics (Guilin, 2001), Lecture Notes in Comput. Sci., vol. 2108, Springer, Berlin, 2001, pp. 1–10. MR 1935355 (2003j:68051)

[Blä03] ———, *On the complexity of the multiplication of matrices of small formats*, J. Complexity **19** (2003), no. 1, 43–60. MR MR1951322 (2003k:68040)

[Blä13] Markus Bläser, *Fast matrix multiplication*, Graduate Surveys, no. 5, Theory of Computing Library, 2013.

[Blä14] Markus Bläser, *Explicit tensors*, Perspectives in Computational Complexity, Springer, 2014, pp. 117–130.

[BLMW11] Peter Bürgisser, J. M. Landsberg, Laurent Manivel, and Jerzy Weyman, *An overview of mathematical issues arising in the geometric complexity theory approach to* VP $\neq$ VNP, SIAM J. Comput. **40** (2011), no. 4, 1179–1209. MR 2861717

[BLR80] Dario Bini, Grazia Lotti, and Francesco Romani, *Approximate solutions for the bilinear form computational problem*, SIAM J. Comput. **9** (1980), no. 4, 692–697. MR MR592760 (82a:68065)

[BOC92] Micheal Ben Or and Richard Cleve, *Computing algebraic formulas using a constant number of registers*, SIAM J. Comput. **21** (1992), no. 21, 54–58.

[Bre70] R.P. Brent, *Algorithms for matrix multiplication*, Technical Report TR-CS-70-157 DCS, Stanford (1970), 1–52.

[Bre74] Richard P. Brent, *The parallel evaluation of general arithmetic expressions*, J. Assoc. Comput. Mach. **21** (1974), 201–206. MR 0660280 (58 #31996)

[Bri93] A. Brill, *über symmetrische functionen von variabelnpaaren*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen **20** (1893), 757–762.

[Bri93] Michel Brion, *Stable properties of plethysm: on two conjectures of Foulkes*, Manuscripta Math. **80** (1993), no. 4, 347–371. MR MR1243152 (95c:20056)

[Bri97] ———, *Sur certains modules gradués associés aux produits symétriques*, Algèbre non commutative, groupes quantiques et invariants (Reims, 1995), Sémin. Congr., vol. 2, Soc. Math. France, Paris, 1997, pp. 157–183. MR 1601139 (99e:20054)

[Bri02] Emmanuel Briand, *Polynômes multisymétriques*, Ph.D. thesis, Université de Rennes 1 et Universidad de Cantabria, 2002.

[Bri10] ———, *Covariants vanishing on totally decomposable forms*, Liaison, Schottky problem and invariant theory, Progr. Math., vol. 280, Birkhäuser Verlag, Basel, 2010, pp. 237–256. MR 2664658

[BS71] Jacek Bochnak and Józef Siciak, *Polynomials and multilinear mappings in topological vector spaces*, Studia Math. **39** (1971), 59–76. MR 0313810

[Bsh98] Nader H. Bshouty, *On the direct sum conjecture in the straight line model*, J. Complexity **14** (1998), no. 1, 49–62. MR 1617757 (99c:13056)

[BT15] Grigoriy Blekherman and Zach Teitler, *On maximum, typical and generic ranks*, Math. Ann. **362** (2015), no. 3-4, 1021–1031. MR 3368091

[Bur14] Vladimir P. Burichenko, *On symmetries of the strassen algorithm*, CoRR **abs/1408.6273** (2014).

[Bur15] ———, *Symmetries of matrix multiplication algorithms. I*, CoRR **abs/1508.01110** (2015).

[Cai90] Jin-Yi Cai, *A note on the determinant and permanent problem*, Inform. and Comput. **84** (1990), no. 1, 119–127. MR MR1032157 (91d:68028)

[CCC+15a] E. Carlini, M. V. Catalisano, L. Chiantini, A. V. Geramita, and Y. Woo, *Symmetric tensors: rank, Strassen's conjecture and e-computability*, ArXiv e-prints (2015).

[CCC15b] Enrico Carlini, Maria Virginia Catalisano, and Luca Chiantini, *Progress on the symmetric Strassen conjecture*, J. Pure Appl. Algebra **219** (2015), no. 8, 3149–3157. MR 3320211

[CHI+] Luca Chiantini, Jon Hauenstein, Christian Ikenmeyer, J.M. Landsberg, and Giorgio Ottaviani, *Towards algorithms for matrix multiplication*, in preparation.

[CIM15] M.-W. Cheung, C. Ikenmeyer, and S. Mkrtchyan, *Symmetrizing Tableaux and the 5th case of the Foulkes Conjecture*, ArXiv e-prints (2015).

[CKSU05] H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans, *Group-theoretic algorithms for matrix multiplication*, Proceedings of the 46th annual Symposium on Foundations of Computer Science (2005), 379–388.

[CKW10] Xi Chen, Neeraj Kayal, and Avi Wigderson, *Partial derivatives in arithmetic complexity and beyond*, Found. Trends Theor. Comput. Sci. **6** (2010), no. 1-2, front matter, 1–138 (2011). MR 2901512

[Com02] P. Comon, *Tensor decompositions, state of the art and applications*, Mathematics in Signal Processing V (J. G. McWhirter and I. K. Proudler, eds.), Clarendon Press, Oxford, UK, 2002, arXiv:0905.0454v1, pp. 1–24.

[Csa76] L. Csanky, *Fast parallel matrix inversion algorithms*, SIAM J. Comput. **5** (1976), no. 4, 618–623. MR 0455310 (56 #13549)

[CU03] H Cohn and C. Umans, *A group theoretic approach to fast matrix multiplication*, Proceedings of the 44th annual Symposium on Foundations of Computer Science (2003), no. 2, 438–449.

[CU13] ______, *Fast matrix multiplication using coherent configurations*, Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (2013), no. 2, 1074–1087.

[CW82] D. Coppersmith and S. Winograd, *On the asymptotic complexity of matrix multiplication*, SIAM J. Comput. **11** (1982), no. 3, 472–492. MR 664715 (83j:68047b)

[CW90] Don Coppersmith and Shmuel Winograd, *Matrix multiplication via arithmetic progressions*, J. Symbolic Comput. **9** (1990), no. 3, 251–280. MR 91i:68058

[CW16] NICHOLAS J. CAVENAGH and IAN M. WANLESS, *There are asymptotically the same number of latin squares of each parity*, Bulletin of the Australian Mathematical Society **FirstView** (2016), 1–8.

[dG78] Hans F. de Groote, *On varieties of optimal algorithms for the computation of bilinear mappings. I. The isotropy group of a bilinear mapping*, Theoret. Comput. Sci. **7** (1978), no. 1, 1–24. MR 0506377 (58 #22132)

[Die49] Jean Dieudonné, *Sur une généralisation du groupe orthogonal à quatre variables*, Arch. Math. **1** (1949), 282–287. MR 0029360 (10,586l)

[Dol03] Igor Dolgachev, *Lectures on invariant theory*, London Mathematical Society Lecture Note Series, vol. 296, Cambridge University Press, Cambridge, 2003. MR MR2004511 (2004g:14051)

[Dri97] Arthur A. Drisko, *On the number of even and odd Latin squares of order $p+1$*, Adv. Math. **128** (1997), no. 1, 20–35. MR 1451417 (98e:05018)

[DS13] A. M. Davie and A. J. Stothers, *Improved bound for complexity of matrix multiplication*, Proc. Roy. Soc. Edinburgh Sect. A **143** (2013), no. 2, 351–369. MR 3039815

[Dyn52] E. B. Dynkin, *Maximal subgroups of the classical groups*, Trudy Moskov. Mat. Obšč. **1** (1952), 39–166. MR 0049903 (14,244d)

[Eis95] David Eisenbud, *Commutative algebra*, Graduate Texts in Mathematics, vol. 150, Springer-Verlag, New York, 1995, With a view toward algebraic geometry. MR MR1322960 (97a:13001)

[ELSW16] K. Efremenko, J. M. Landsberg, H. Schenck, and J. Weyman, *The method of shifted partial derivatives cannot separate the permanent from the determinant*, ArXiv e-prints, to appear in MCOM (2016).

[FH91] William Fulton and Joe Harris, *Representation theory*, Graduate Texts in Mathematics, vol. 129, Springer-Verlag, New York, 1991, A first course, Readings in Mathematics. MR 1153249 (93a:20069)

[Fis94] Ismor Fischer, *Sums of Like Powers of Multivariate Linear Forms*, Math. Mag. **67** (1994), no. 1, 59–61. MR 1573008

[Fou50] H. O. Foulkes, *Concomitants of the quintic and sextic up to degree four in the coefficients of the ground form*, J. London Math. Soc. **25** (1950), 205–209. MR MR0037276 (12,236e)

[Fro97] G. Frobenius, *Über die Darstellung der endlichen Gruppen durch lineare Substitutionen*, Sitzungsber Deutsch. Akad. Wiss. Berlin (1897), 994–1015.

[Frö85] Ralf Fröberg, *An inequality for Hilbert series of graded algebras*, Math. Scand. **56** (1985), no. 2, 117–144. MR 813632 (87f:13022)

[FW84] Ephraim Feig and Shmuel Winograd, *On the direct sum conjecture*, Linear Algebra Appl. **63** (1984), 193–219. MR 766508 (86h:15022)

[Gal] Franois Le Gall, *Powers of tensors and fast matrix multiplication*, arXiv:1401.7714.

[Gat87] Joachim von zur Gathen, *Feasible arithmetic computations: Valiant's hypothesis*, J. Symbolic Comput. **4** (1987), no. 2, 137–172. MR MR922386 (89f:68021)

[Gay76] 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** (1976), no. 3, 449–455. MR MR0414794 (54 #2886)

[Ger61] Murray Gerstenhaber, *On dominance and varieties of commuting matrices*, Ann. of Math. (2) **73** (1961), 324–348. MR 0132079 (24 #A1926)

[Ges16] Fulvio Gesmundo, *Geometric aspects of iterated matrix multiplication*, J. Algebra **461** (2016), 42–64. MR 3513064

[GH79] Phillip Griffiths and Joseph Harris, *Algebraic geometry and local differential geometry*, Ann. Sci. École Norm. Sup. (4) **12** (1979), no. 3, 355–452. MR 81k:53004

[GHIL] Fulvio Gesmundo, Jonathan Hauenstein, Christian Ikenmeyer, and J. M. Landsberg, *Geometry and matrix rigidity*, arXiv:1310.1362.

[GHPS14] Zachary A. Griffin, Jonathan D. Hauenstein, Chris Peterson, and Andrew J. Sommese, *Numerical computation of the Hilbert function and regularity of a zero dimensional scheme*, Connections between algebra, combinatorics, and geometry, Springer Proc. Math. Stat., vol. 76, Springer, New York, 2014, pp. 235–250. MR 3213522

[GKKS13a] Ankit Gupta, Pritish Kamath, Neeraj Kayal, and Ramprasad Saptharishi, *Approaching the chasm at depth four*, Proceedings of the Conference on Computational Complexity (CCC) (2013).

[GKKS13b] ______, *Arithmetic circuits: A chasm at depth three*, Electronic Colloquium on Computational Complexity (ECCC) **20** (2013), 26.

[GKZ94] I. M. Gel′fand, M. M. Kapranov, and A. V. Zelevinsky, *Discriminants, resultants, and multidimensional determinants*, Mathematics: Theory & Applications, Birkhäuser Boston Inc., Boston, MA, 1994. MR 95e:14045

[Gly10] David G. Glynn, *The conjectures of Alon-Tarsi and Rota in dimension prime minus one*, SIAM J. Discrete Math. **24** (2010), no. 2, 394–399. MR 2646093 (2011i:05034)

[Gly13] ______, *Permanent formulae from the Veronesean*, Des. Codes Cryptogr. **68** (2013), no. 1-3, 39–47. MR 3046335

[Gor94] P. Gordon, *Das zerfallen der curven in gerade linien*, Math. Ann. (1894), no. 45, 410–427.

[Got78] Gerd Gotzmann, *Eine Bedingung für die Flachheit und das Hilbertpolynom eines graduierten Ringes*, Math. Z. **158** (1978), no. 1, 61–70. MR 0480478 (58 #641)

[Gre78] Edward L. Green, *Complete intersections and Gorenstein ideals*, J. Algebra **52** (1978), no. 1, 264–273. MR 0480472

[Gre98] Mark L. Green, *Generic initial ideals*, Six lectures on commutative algebra (Bellaterra, 1996), Progr. Math., vol. 166, Birkhäuser, Basel, 1998, pp. 119–186. MR 1648665 (99m:13040)

[Gre11] Bruno Grenet, *An Upper Bound for the Permanent versus Determinant Problem*, Manuscript (submitted), 2011.

[Gri86] B. Griesser, *A lower bound for the border rank of a bilinear map*, Calcolo **23** (1986), no. 2, 105–114 (1987). MR 88g:15021

[Gun] S. Gundelfinger.

[GW09] Roe Goodman and Nolan R. Wallach, *Symmetry, representations, and invariants*, Graduate Texts in Mathematics, vol. 255, Springer, Dordrecht, 2009. MR 2522486

[Had97] J. Hadamard, *Mémoire sur l'élimination*, Acta Math. **20** (1897), no. 1, 201–238. MR 1554881

[Had99] ______, *Sur les conditions de décomposition des formes*, Bull. Soc. Math. France **27** (1899), 34–47. MR 1504330

[Har95] Joe Harris, *Algebraic geometry*, Graduate Texts in Mathematics, vol. 133, Springer-Verlag, New York, 1995, A first course, Corrected reprint of the 1992 original. MR 1416564 (97e:14001)

[HIL13] Jonathan D. Hauenstein, Christian Ikenmeyer, and J. M. Landsberg, *Equations for lower bounds on border rank*, Exp. Math. **22** (2013), no. 4, 372–383. MR 3171099

[How87] Roger Howe, $(\mathrm{GL}_n, \mathrm{GL}_m)$*-duality and symmetric plethysm*, Proc. Indian Acad. Sci. Math. Sci. **97** (1987), no. 1-3, 85–109 (1988). MR MR983608 (90b:22020)

[HR94] Rosa Huang and Gian-Carlo Rota, *On the relations of various conjectures on Latin squares and straightening coefficients*, Discrete Math. **128** (1994), no. 1-3, 225–236. MR 1271866 (95i:05036)

[HS13] Jonathan D. Hauenstein and Andrew J. Sommese, *Membership tests for images of algebraic sets by linear projections*, Appl. Math. Comput. **219** (2013), no. 12, 6809–6818. MR 3027848

[Iar97] A. Iarrobino, *Inverse system of a symbolic power. III. Thin algebras and fat points*, Compositio Math. **108** (1997), no. 3, 319–356. MR 1473851 (98k:13017)

[IE78] A. Iarrobino and J. Emsalem, *Some zero-dimensional generic singularities; finite algebras having small tangent space*, Compositio Math. **36** (1978), no. 2, 145–188. MR 515043

[IK99] Anthony Iarrobino and Vassil Kanev, *Power sums, Gorenstein algebras, and determinantal loci*, Lecture Notes in Mathematics, vol. 1721, Springer-Verlag, Berlin, 1999, Appendix C by Iarrobino and Steven L. Kleiman. MR MR1735271 (2001d:14056)

[Ike15] C. Ikenmeyer, *On McKay's propagation theorem for the Foulkes conjecture*, ArXiv e-prints (2015).

[IL03] Thomas A. Ivey and J. M. Landsberg, *Cartan for beginners: differential geometry via moving frames and exterior differential systems*, Graduate Studies in Mathematics, vol. 61, American Mathematical Society, Providence, RI, 2003. MR 2 003 610

[IM05] Atanas Iliev and Laurent Manivel, *Varieties of reductions for $\mathfrak{gl}_n$*, Projective varieties with unexpected properties, Walter de Gruyter GmbH & Co. KG, Berlin, 2005, pp. 287–316. MR MR2202260 (2006j:14056)

[IP15] C. Ikenmeyer and G. Panova, *Rectangular Kronecker coefficients and plethysms in geometric complexity theory*, ArXiv e-prints (2015).

[JM86] Rodney W. Johnson and Aileen M. McLoughlin, *Noncommutative bilinear algorithms for 3×3 matrix multiplication*, SIAM J. Comput. **15** (1986), no. 2, 595–603. MR 837607

[Józ89] Tadeusz Józefiak, *Characters of projective representations of symmetric groups*, Exposition. Math. **7** (1989), no. 3, 193–247. MR 1007885 (90f:20018)

[JT86] Joseph Ja'Ja' and Jean Takche, *On the validity of the direct sum conjecture*, SIAM J. Comput. **15** (1986), no. 4, 1004–1020. MR MR861366 (88b:68084)

[KL14] Harlan Kadish and J. M. Landsberg, *Padded polynomials, their cousins, and geometric complexity theory*, Comm. Algebra **42** (2014), no. 5, 2171–2180. MR 3169697

[KLPSMN09] Abhinav Kumar, Satyanarayana V. Lokam, Vijay M. Patankar, and Jayalal Sarma M. N., *Using elimination theory to construct rigid matrices*, Foundations of software technology and theoretical computer science—FSTTCS 2009, LIPIcs. Leibniz Int. Proc. Inform., vol. 4, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2009, pp. 299–310. MR 2870721

[Koi] Pascal Koiran, *Arithmetic circuits: the chasm at depth four gets wider*, preprint arXiv:1006.4700.

[Kum] Shrawan Kumar, *A study of the representations supported by the orbit closure of the determinant*, arXiv:1109.5996.

[Kum13] ______, *Geometry of orbits of permanents and determinants*, Comment. Math. Helv. **88** (2013), no. 3, 759–788. MR 3093509

[Lad76] Julian D. Laderman, *A noncommutative algorithm for multiplying* 3×3 *matrices using* 23 *muliplications*, Bull. Amer. Math. Soc. **82** (1976), no. 1, 126–128. MR MR0395320 (52 #16117)

[Lan06] J. M. Landsberg, *The border rank of the multiplication of* 2×2 *matrices is seven*, J. Amer. Math. Soc. **19** (2006), no. 2, 447–459. MR 2188132 (2006j:68034)

[Lan10] ______, *P versus NP and geometry*, J. Symbolic Comput. **45** (2010), no. 12, 1369–1377. MR 2733384 (2012c:68065)

[Lan12] ______, *Tensors: geometry and applications*, Graduate Studies in Mathematics, vol. 128, American Mathematical Society, Providence, RI, 2012. MR 2865915

[Lan14a] ______, *Geometric complexity theory: an introduction for geometers*, ANNALI DELL'UNIVERSITA' DI FERRARA (2014), 1–53 (English).

[Lan14b] ______, *New lower bounds for the rank of matrix multiplication*, SIAM J. Comput. **43** (2014), no. 1, 144–149. MR 3162411

[Lan15] ______, *Nontriviality of equations and explicit tensors in* $\mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$ *of border rank at least* $2m - 2$, J. Pure Appl. Algebra **219** (2015), no. 8, 3677–3684. MR 3320240

[Las78] Alain Lascoux, *Syzygies des variétés déterminantales*, Adv. in Math. **30** (1978), no. 3, 202–237. MR 520233 (80j:14043)

[Lee16] Hwangrae Lee, *Power sum decompositions of elementary symmetric polynomials*, Linear Algebra Appl. **492** (2016), 89–97. MR 3440150

[Lei16] Arielle Leitner, *Limits under conjugacy of the diagonal subgroup in* $SL_n(\mathbb{R})$, Proc. Amer. Math. Soc. **144** (2016), no. 8, 3243–3254. MR 3503693

[Lic85] Thomas Lickteig, *Typical tensorial rank*, Linear Algebra Appl. **69** (1985), 95–120. MR 87f:15017

[Lit06] Dudley E. Littlewood, *The theory of group characters and matrix representations of groups*, AMS Chelsea Publishing, Providence, RI, 2006, Reprint of the second (1950) edition. MR MR2213154 (2006m:20013)

[LM04] J. M. Landsberg and Laurent Manivel, *On the ideals of secant varieties of Segre varieties*, Found. Comput. Math. **4** (2004), no. 4, 397–422. MR MR2097214 (2005m:14101)

[LM08] ______, *Generalizations of Strassen's equations for secant varieties of Segre varieties*, Comm. Algebra **36** (2008), no. 2, 405–422. MR MR2387532

[LM15] J. M. Landsberg and M. Michalek, *Abelian Tensors*, ArXiv e-prints, to appear in JMPA (2015).

[LM16a] ______, *A* $2n^2 - log(n) - 1$ *lower bound for the border rank of matrix multiplication*, ArXiv e-prints (2016).

[LM16b] ______, *On the geometry of border rank algorithms for matrix multiplication and other tensors with symmetry*, ArXiv e-prints (2016).

[LMR13] Joseph M. Landsberg, Laurent Manivel, and Nicolas Ressayre, *Hypersurfaces with degenerate duals and the geometric complexity theory program*, Comment. Math. Helv. **88** (2013), no. 2, 469–484. MR 3048194

[LO13] J. M. Landsberg and Giorgio Ottaviani, *Equations for secant varieties of Veronese and other varieties*, Ann. Mat. Pura Appl. (4) **192** (2013), no. 4, 569–606. MR 3081636

[LO15] Joseph M. Landsberg and Giorgio Ottaviani, *New lower bounds for the border rank of matrix multiplication*, Theory Comput. **11** (2015), 285–298. MR 3376667

[LR0] J. M. Landsberg and Nicholas Ryder, *On the geometry of border rank algorithms for n 2 by 2 2 matrix multiplication*, Experimental Mathematics **0** (0), no. 0, 1–12.

[LR06] Hong Liu and Kenneth W. Regan, *Improved construction for universality of determinant and permanent*, Inform. Process. Lett. **100** (2006), no. 6, 233–237. MR 2270826 (2007f:68084)

[LR15] J. M. Landsberg and N. Ressayre, *Permanent v. determinant: an exponential lower bound assumingsymmetry and a potential path towards Valiant's conjecture*, ArXiv e-prints (2015).

[Mac95] I. G. Macdonald, *Symmetric functions and Hall polynomials*, second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1995, With contributions by A. Zelevinsky, Oxford Science Publications. MR 1354144 (96h:05207)

[Man97] Laurent Manivel, *Applications de Gauss et pléthysme*, Ann. Inst. Fourier (Grenoble) **47** (1997), no. 3, 715–773. MR MR1465785 (98h:20078)

[Man98] ———, *Gaussian maps and plethysm*, Algebraic geometry (Catania, 1993/Barcelona, 1994), Lecture Notes in Pure and Appl. Math., vol. 200, Dekker, New York, 1998, pp. 91–117. MR MR1651092 (99h:20070)

[Man15a] ———, *On the asymptotics of Kronecker coefficients*, J. Algebraic Combin. **42** (2015), no. 4, 999–1025. MR 3417256

[Man15b] ———, *On the asymptotics of Kronecker coefficients, 2*, Sém. Lothar. Combin. **75** (2015), Art. B75d, 13. MR 3461556

[Mat60] Yozô Matsushima, *Espaces homogènes de Stein des groupes de Lie complexes*, Nagoya Math. J **16** (1960), 205–218. MR MR0109854 (22 #739)

[McK08] Tom McKay, *On plethysm conjectures of Stanley and Foulkes*, J. Algebra **319** (2008), no. 5, 2050–2071. MR 2394689 (2008m:20023)

[MM61] Marvin Marcus and Henryk Minc, *On the relation between the determinant and the permanent*, Illinois J. Math. **5** (1961), 376–381. MR 0147488 (26 #5004)

[MM62] Marvin Marcus and F. C. May, *The permanent function*, Canad. J. Math. **14** (1962), 177–189. MR MR0137729 (25 #1178)

[MN05] Jurgen Müller and Max Neunhöffer, *Some computations regarding Foulkes' conjecture*, Experiment. Math. **14** (2005), no. 3, 277–283. MR MR2172706 (2006e:05186)

[MO34] S. Mazur and W. Orlicz, *Grundlegende eigenschaften der polynomischen operationen*, Studia Math. **5** (1934), 50–68.

[MP] D. Maulik and R. Pandharipande, *Gromov-Witten theory and Noether-Lefschetz theory*, preprint, 0705.1653.

[MR04] Thierry Mignon and Nicolas Ressayre, *A quadratic bound for the determinant and permanent problem*, Int. Math. Res. Not. (2004), no. 79, 4241–4253. MR MR2126826 (2006b:15015)

[MR13] Alex Massarenti and Emanuele Raviolo, *The rank of $n \times n$ matrix multiplication is at least $3n^2 - 2\sqrt{2}n^{\frac{3}{2}} - 3n$*, Linear Algebra Appl. **438** (2013), no. 11, 4500–4509. MR 3034546

[MS01] Ketan D. Mulmuley and Milind Sohoni, *Geometric complexity theory. I. An approach to the P vs. NP and related problems*, SIAM J. Comput. **31** (2001), no. 2, 496–526 (electronic). MR MR1861288 (2003a:68047)

[MS08] ———, *Geometric complexity theory. II. Towards explicit obstructions for embeddings among class varieties*, SIAM J. Comput. **38** (2008), no. 3, 1175–1206. MR MR2421083

[Mul] K. Mulmuley, *The gct chasm*, lecture.

[Mul99] Ketan Mulmuley, *Lower bounds in a parallel model without bit operations*, SIAM J. Comput. **28** (1999), no. 4, 1460–1509 (electronic). MR 1681069

[Mum66] David Mumford, *Lectures on curves on an algebraic surface*, With a section by G. M. Bergman. Annals of Mathematics Studies, No. 59, Princeton University Press, Princeton, N.J., 1966. MR 0209285

[Mum95] ———, *Algebraic geometry. I*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Complex projective varieties, Reprint of the 1976 edition. MR 1344216 (96d:14001)

[MV97] Meena Mahajan and V. Vinay, *Determinant: combinatorics, algorithms, and complexity*, Chicago J. Theoret. Comput. Sci. (1997), Article 5, 26 pp. (electronic). MR 1484546 (98m:15016)

[Nis91] Noam Nisan, *Lower bounds for non-commutative computation*, Proceedings of the Twenty-third Annual ACM Symposium on Theory of Computing (New York, NY, USA), STOC '91, ACM, 1991, pp. 410–418.

[NR16] J.F. Nash and M.T. Rassias, *Open problems in mathematics*, Springer International Publishing, 2016.

[NW97] Noam Nisan and Avi Wigderson, *Lower bounds on arithmetic circuits via partial derivatives*, Comput. Complexity **6** (1996/97), no. 3, 217–234. MR 1486927 (99f:68107)

[OV90] A. L. Onishchik and È. B. Vinberg, *Lie groups and algebraic groups*, Springer Series in Soviet Mathematics, Springer-Verlag, Berlin, 1990, Translated from the Russian and with a preface by D. A. Leites. MR 1064110 (91g:22001)

[Pan66] V. Ja. Pan, *On means of calculating values of polynomials*, Uspehi Mat. Nauk **21** (1966), no. 1 (127), 103–134. MR 0207178

[Pop75] A. M. Popov, *Irreducible simple linear Lie groups with finite standard subgroups in general position*, Funkcional. Anal. i Priložen. **9** (1975), no. 4, 81–82. MR MR0396847 (53 #707)

[Pra94] V. V. Prasolov, *Problems and theorems in linear algebra*, Translations of Mathematical Monographs, vol. 134, American Mathematical Society, Providence, RI, 1994, Translated from the Russian manuscript by D. A. Leĭtes. MR 1277174

[Pro07] Claudio Procesi, *Lie groups*, Universitext, Springer, New York, 2007, An approach through invariants and representations. MR MR2265844 (2007j:22016)

[RS11] Kristian Ranestad and Frank-Olaf Schreyer, *On the rank of a symmetric form*, J. Algebra **346** (2011), 340–342. MR 2842085 (2012j:13037)

[Sch81] A. Schönhage, *Partial and total matrix multiplication*, SIAM J. Comput. **10** (1981), no. 3, 434–455. MR MR623057 (82h:68070)

[Seg10] C. Segre, *Preliminari di una teoria delle varieta luoghi di spazi*, Rend. Circ. Mat. Palermo XXX (1910), 87–121.

[Sha94] Igor R. Shafarevich, *Basic algebraic geometry. 1*, second ed., Springer-Verlag, Berlin, 1994, Varieties in projective space, Translated from the 1988 Russian edition and with notes by Miles Reid. MR MR1328833 (95m:14001)

[Sha13] ______, *Basic algebraic geometry. 1*, third ed., Springer, Heidelberg, 2013, Varieties in projective space. MR 3100243

[Shp02] Amir Shpilka, *Affine projections of symmetric polynomials*, J. Comput. System Sci. **65** (2002), no. 4, 639–659, Special issue on complexity, 2001 (Chicago, IL). MR 1964647

[Sip92] Michael Sipser, *The history and status of the p versus np question*, STOC '92 Proceedings of the twenty-fourth annual ACM symposium on Theory of computing (1992), 603–618.

[Smi13] A. V. Smirnov, *The bilinear complexity and practical algorithms for matrix multiplication*, Comput. Math. Math. Phys. **53** (2013), no. 12, 1781–1795. MR 3146566

[Spi79] Michael Spivak, *A comprehensive introduction to differential geometry. Vol. I*, second ed., Publish or Perish Inc., Wilmington, Del., 1979. MR MR532830 (82g:53003a)

[SS42] R. Salem and D. C. Spencer, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U. S. A. **28** (1942), 561–563. MR 0007405

[Ste89] John R. Stembridge, *Shifted tableaux and the projective representations of symmetric groups*, Adv. Math. **74** (1989), no. 1, 87–134. MR 991411 (90k:20026)

[Sto] A. Stothers, *On the complexity of matrix multiplication*, PhD thesis, University of Edinburgh, 2010.

[Str69] Volker Strassen, *Gaussian elimination is not optimal*, Numer. Math. **13** (1969), 354–356. MR 40 #2223

[Str73] ______, *Vermeidung von Divisionen*, J. Reine Angew. Math. **264** (1973), 184–202. MR MR0521168 (58 #25128)

[Str75] ______, *Die Berechnungskomplexität der symbolischen Differentiation von Interpolationspolynomen*, Theor. Comput. Sci. **1** (1975), no. 1, 21–25. MR 0395147 (52 #15945)

[Str83] V. Strassen, *Rank and optimal computation of generic tensors*, Linear Algebra Appl. **52/53** (1983), 645–685. MR 85b:15039

[Str87] ______, *Relative bilinear complexity and matrix multiplication*, J. Reine Angew. Math. **375/376** (1987), 406–443. MR MR882307 (88h:11026)

[Str91] ______, *Degeneration and complexity of bilinear maps: some asymptotic spectra*, J. Reine Angew. Math. **413** (1991), 127–180. MR 92m:11038

[SVW01] A. J. Sommese, J. Verschelde, and C. W. Wampler, *Using monodromy to decompose solution sets of polynomial systems into irreducible components*, Applications of algebraic geometry to coding theory, physics and computation (Eilat, 2001), NATO Sci. Ser. II Math. Phys. Chem., vol. 36, Kluwer Acad. Publ., Dordrecht, 2001, pp. 297–315. MR 1866906

[SVW02] Andrew J. Sommese, Jan Verschelde, and Charles W. Wampler, *Symmetric functions applied to decomposing solution sets of polynomial systems*, SIAM J. Numer. Anal. **40** (2002), no. 6, 2026–2046 (2003). MR 1974173

[SW01] Amir Shpilka and Avi Wigderson, *Depth-3 arithmetic circuits over fields of characteristic zero*, Comput. Complexity **10** (2001), no. 1, 1–27. MR 1867306 (2003a:68048)

[SY09] Amir Shpilka and Amir Yehudayoff, *Arithmetic circuits: a survey of recent results and open questions*, Found. Trends Theor. Comput. Sci. **5** (2009), no. 3-4, 207–388 (2010). MR 2756166

[Syl52] J. J. Sylvester, *On the principles of the calculus of forms*, Cambridge and Dublin Mathematical Journal (1852), 5297.

[Ter11] A. Terracini, *Sulla v_k per cui la varieta degli $s_h(h+1)$-seganti ha dimensione minore dell'ordinario*, Rend. Circ. Mat. Palermo **31** (1911), 392–396.

[Toe77] Emil Toeplitz, *Ueber ein Flächennetz zweiter Ordnung*, Math. Ann. **11** (1877), no. 3, 434–463. MR 1509924

[Tra84] B. A. Trakhtenbrot, *A survey of Russian approaches to perebor (brute-force search) algorithms*, Ann. Hist. Comput. **6** (1984), no. 4, 384–400. MR 763733

[Val77] Leslie G. Valiant, *Graph-theoretic arguments in low-level complexity*, Mathematical foundations of computer science (Proc. Sixth Sympos., Tatranská Lomnica, 1977), Springer, Berlin, 1977, pp. 162–176. Lecture Notes in Comput. Sci., Vol. 53. MR 0660702 (58 #32067)

[Val79a] L. G. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci. **8** (1979), no. 2, 189–201. MR MR526203 (80f:68054)

[Val79b] Leslie G. Valiant, *Completeness classes in algebra*, Proc. 11th ACM STOC, 1979, pp. 249–261.

[VSBR83] L. G. Valiant, S. Skyum, S. Berkowitz, and C. Rackoff, *Fast parallel computation of polynomials using few processors*, SIAM J. Comput. **12** (1983), no. 4, 641–644. MR 721003 (86a:68044)

[vzG87] Joachim von zur Gathen, *Permanent and determinant*, Linear Algebra Appl. **96** (1987), 87–100. MR MR910987 (89a:15005)

[Wah91] Jonathan Wahl, *Gaussian maps and tensor products of irreducible representations*, Manuscripta Math. **73** (1991), no. 3, 229–259. MR 1132139 (92m:14066a)

[Wey97] Hermann Weyl, *The classical groups*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997, Their invariants and representations, Fifteenth printing, Princeton Paperbacks. MR 1488158 (98k:01049)

[Wey03] Jerzy Weyman, *Cohomology of vector bundles and syzygies*, Cambridge Tracts in Mathematics, vol. 149, Cambridge University Press, Cambridge, 2003. MR MR1988690 (2004d:13020)

[Wil] Virginia Williams, *Breaking the coppersimith-winograd barrier*, preprint.

[Win71] S. Winograd, *On multiplication of 2×2 matrices*, Linear Algebra and Appl. **4** (1971), 381–388. MR 45 #6173

[Ye11] Ke Ye, *The stabilizer of immanants*, Linear Algebra Appl. **435** (2011), no. 5, 1085–1098. MR 2807220 (2012e:15017)

[Zui15] J. Zuiddam, *A note on the gap between rank and border rank*, ArXiv e-prints (2015).

Index