NOTES FOR TRENTO SUMMER SCHOOL

J.M. LANDSBERG

These notes are meant as a supplement to my lecture series. I will cover part of them in detail, other parts are more background material that will be covered only briefly, and further parts consist of supplementary material that I probably will not have time to touch upon. In addition to these notes, the lectures will include basics on the representation theory of GL_n. I suggest reading sections 3 and 4 before you arrive and doing the exercises there.

1. Tensor decompositions arising in numerical analysis

1.1. **Motivation.** For a proper introduction to the uses of tensors in numerical analysis, I suggest [9]. There are good techniques for approximating linear maps between vector spaces equipped with inner products by linear maps of low rank. (Take the singular value decomposition and truncate.) One would like to do something similar for tensors. A central issue for applications is storage of information, and if we are in $A_1 \otimes \cdots \otimes A_d$, even if $\dim A_i = 2$, there is an exponential growth in storage costs with the number of factors. One wants to approximate tensors in a way that the cost grows slower than exponentially with the number of factors. A first idea, to use tensors of low tensor rank has two drawbacks: first, we do not know how to find low rank approximations, and more seriously, tensor rank is not semi-continuous, so it gives rise to instability. In physics one often uses *tensor network states*. I will use a modified version of W. Hackbusch's notation to describe these in the binary tree case. I restrict to the tree case because other cases also suffer from not being closed, see [16]. I restrict further to the binary tree case because they are most amenable to adopting approximation methods from linear algebra, and the main purpose of these notes is to present a question of Hackbusch and some auxiliary geometric questions.

1.2. **Tree tensor network states.** Take a binary tree Γ where we write the root at the top, call it v_f, and the vertices with a single edge (called the offspring) at the bottom, which we assume are d in number, denoted $v_1, \cdots, v_d$ and we associate the vector spaces A_i to these vertices. To all other vertices v_α we associate a natural number r_α, and let $\overline{r} = (r_1, \cdots, r_f)$ be the vector of these natural numbers. Then define $TNS_{\Gamma,\overline{r}} \subset A_1 \otimes \cdots \otimes A_d$ to be the set of tensors $T \in A_1 \otimes \cdots \otimes A_d$ such that there exists subspaces, $U_j \subset A_j$ of dimension r_j, and for all other vertices, there exist U_v of dimension r_v where U_v is a subspace of the tensor product of the U's associated to its two children, and such that $T \in U_f$.

For example, if $d = 3$ and Γ is

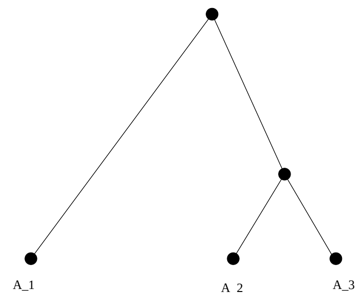
Then $TNS_{\Gamma,(r_1,r_2,r_3,r_{23},r_{1,23})}$ consists of tensors T such that there exist $U_j \subset A_j$ of dimension r_j, $U_{23} \subset U_2 \otimes U_3$ of dimension r_{23} and $U \subset U_1 \otimes U_{23}$ of dimension $r_{1,23}$ such that $T \in U$.

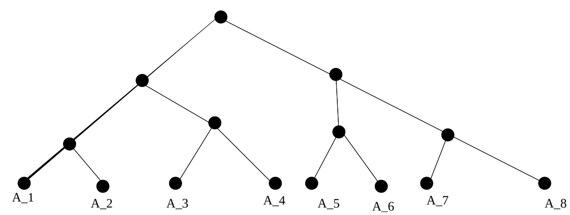
Among these there are two cases that appear frequently in the literature, the *tensor train* or *matrix product states*, denoted TT, where the graph is:

and the hierarchical format, denoted HF, where the graph is a binary tree:

When we are in one of the above formats I write $TT_{\overline{r}}$ or $HF_{\overline{r}}$ instead. Note that I should really include an ordering of the offspring vertices, but I suppress that in the notation.

Landsberg supported by NSF grant DMS-1006353.





Exercise 1.1: Write down parametrizations of $TT_{\overline{r}}$ and $HF_{\overline{r}}$.

Proposition 1.2. *For all trees Γ, the set $TNS_{\Gamma,\overline{r}} \subset A_1 \otimes \cdots \otimes A_d$ is an algebraic variety, which is irreducible for any fixed ordering of the roots.*

Exercise 1.3: Prove Proposition 1.2. Hint: consider bundles over Grassmannians.

Problem 1.4. Give a nice presentation of generators of the ideal, especially in the TT and HF cases. What are the dimensions of these varieties?

Exercise 1.5: [9, Prop 12.5] Let $T \in TT_{\overline{r}}$ with $r_j \leq r$ for all j, then $T \in HF_{\overline{r}}$ with $r_s \leq r^2$.

Conjecture 1.6 (Hackbusch)**.** *For all $d = 2^k$ with $k \geq 3$, there exists $T \in HF_{\overline{r}}$ with $r_s \leq r$ and $T \notin TT_{\overline{r}}$ with $r_j < d^{\frac{1}{2}\log_2 r}$ for any ordering of the vector spaces.*

2. Matrix multiplication

The workhorse of scientific computation is matrix multiplication. The standard algorithm for multiplying $n \times n$ matrices uses on the order of n^3 arithmetic operations, whereas addition of matrices only uses n^2. For a $10,000 \times 10,000$ matrix this means 10^{12} arithmetic operations for multiplication compared with 10^8 for addition. Wouldn't it be great if all matrix operations were as easy as addition? As "pie in the sky" as this wish sounds, it might not be far from reality. In 1969 Strassen discovered an algorithm for multiplying 2×2 matrices using seven multiplications and more generally $n \times n$ matrices using on the order of $n^{2.81}$ arithmetic operations. More recent advances have brought the number of operations needed even closer to the n^2 of addition, and it is generally conjectured in the computer science community that asymptotically one can get arbitrarily close to n^2.

2.1. The standard algorithm.

Let A, B be 2×2 matrices

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

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 $n \times n$ matrices, it uses n^3 multiplications and $n^3 - n^2$ additions.

2.2. **Strassen's algorithm for multiplying 2×2 matrices using only seven scalar multiplications** [22]**.** Set

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

Exercise 2.1: Show that if $C = AB$, then

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

2.3. **Fast multiplication of $n \times n$ 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 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, one can multiply $n \times n$ matrices using approximately $n^{\log_2(7)} \simeq n^{2.81}$ arithmetic operations. To see this, let $n = 2^k$ and write $7^k = (2^k)^a$ so $k \log_2 7 = ak \log_2 2$ so $a = \log_2 7$.

Definition 2.2. The *exponent* ω of matrix multiplication is

$$\omega := \inf\{h \in \mathbb{R} \mid Mat_{n \times n} \text{ may be multiplied using } O(n^h) \text{ arithmetic operations}\}$$

Strassen's algorithm shows $\omega \leq \log_2(7) < 2.81$. Determining ω is a central open problem in complexity theory. The current "world record" is $\omega < 2.373$, see [24, 7, 20].

2.4. **Geometry and the complexity of matrix multiplication.** Had someone asked him, Terracini, in 1913 would have been able to predict the existence of something like Strassen's algorithm from geometric considerations alone. He would have been able to tell you that even a generic 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 ϵ" using seven multiplications. We will see how to prove this.

We will work with two different ways of counting arithmetic operations, *(tensor) rank* and *border rank*, that are particularly well suited to be studied using geometry. In particular we will derive lower bounds for the complexity of matrix multiplication and see how to arrive at Strassen's algorithm from purely geometric considerations.

We will define a sequence of nested subsets of the space of bilinear maps $\mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N$ (where for matrix multiplication we will take $N = n^2$). These subsets will be *algebraic varieties*, that is they will be defined by polynomial equations. (They go under the name *secant varieties of Segre varieties.*) To prove a bilinear map T has "complexity" at least r, we will find a polynomial that vanishes on the r-th subset, such that the polynomial does not vanish on T.

The matrix multiplication operator has considerable symmetry. For example, $M(X,Y) = M(XA, A^{-1}Y)$ for any invertible matrix A. Notice that this already gives us an n^2 parameter family of ways of writing down M as a bilinear map. In fact we will see there is considerably more symmetry.

Exploiting symmetry will be critical to this and every problem we discuss in this class. The systematic way to exploit symmetry goes under the name *representation theory*, and I will introduce different aspects of it as we need them.

3. Matrix multiplication and multi-linear algebra

3.1. **Matrix multiplication as a bilinear map.** Let's study rectangular matrix multiplication:

$$M = M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle} : Mat_{\mathbf{n}\times\mathbf{m}} \times Mat_{\mathbf{m}\times\mathbf{l}} \to Mat_{\mathbf{n}\times\mathbf{l}}.$$

Note that matrix multiplication is a *bilinear map*, i.e., for all $X_j, X \in Mat_{n\times m}$, $Y_j, Y \in Mat_{M\times k}$ and $a_j, b_j \in \mathbb{C}$,

$$M(a_1X_1 + a_2X_2, Y) = a_1M(X_1,Y) + a_2M(X_2,Y), \quad \text{and}$$
$$M(X, b_1Y_1 + b_2Y_2) = b_1M(X,Y_1) + b_2M(X,Y_2).$$

Our first task will be to describe matrix multiplication without reference to coordinates - this will reveal extra symmetry that we missed in the last lecture.

To do this I will review basic facts from linear and multi-linear algebra. For more details on this topic, see [12, Chap. 2].

3.2. **Linear maps.** In what follows it will be essential to work without bases, so instead of writing $\mathbb{C}^{\mathbf{v}}$ I will work with a complex vector space V of dimension $\mathbf{v}$. Given V, one can form a second vector space, called the *dual space* to V, 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 $v_1, \cdots, v_{\mathbf{v}}$ of V, it determines a basis $\alpha^1, \cdots, \alpha^{\mathbf{v}}$ of V^* by $\alpha^i(v_j) = \delta_{ij}$.

Exercise 3.1: Assuming V is finite dimensional, write down a canonical isomorphism $V \to (V^*)^*$. ⊚

Let $V^*\otimes W$ denote the 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 $\alpha\otimes w(v) := \alpha(v)w$. Such a linear map is said to have *rank one*.

Define the *rank* of an element $f \in V^*\otimes W$ is the smallest r such f may be expressed as a sum of r rank one linear maps.

Given a linear map $f : V \to W$ we also have 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 3.2: 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 $b_f(\beta, v) := \beta(f(v))$.

Exercise 3.3: Show that $(V^*\otimes W)^* \simeq V\otimes W^*$ where $\alpha\otimes w(v\otimes\beta) := \alpha(w)\beta(v)$. Now let $V = W$ and let $Id_V \in V^*\otimes V \simeq (V^*\otimes V)^*$ denote the identity map. What is $Id_V(f)$ for $f \in V^*\otimes V$? ⊚

Theorem 3.4 (Fundamental theorem of linear algebra)**.** *Let V, W be finite dimensional vector spaces and let $f : V \to W$ be a linear map. Then*

(1) $\operatorname{rank}(f) = \dim f(V) = \dim f^T(W^*)$, *in particular* $\operatorname{rank}(f) \leq \min\{\dim V, \dim W\}$.
(2) *For generic f,* $\operatorname{rank}(f) = \min\{\dim V, \dim W\}$.
(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 $r+1$ minors of the matrix representing f are all zero.*

Exercise 3.5: Prove the last assertion. ⊚

3.3. **Basic definitions from representation theory.** Let $GL(V)$ denote the group of invertible linear maps $V \to V$. If we have chosen a basis, this is the set of invertible $\mathbf{v}\times\mathbf{v}$ matrices. We will say $GL(V)$ *acts on* V and that V is a *G-module*. Write $v \mapsto g\cdot v$ for the action. Then $GL(V)$ also acts on V^* where we define $g\cdot\alpha$ by $(g\cdot\alpha)(v) := \alpha(g^{-1}v)$. The reason we take the inverse is to have $g_1\cdot(g_2\cdot\alpha) = (g_1g_2)\cdot\alpha$.

More generally, for any group G, given a group homomorphism $\mu : G \to GL(V)$, we will say G *acts on* V. We will say the action is *irreducible* if there does not exist a proper subspace $U \subset V$ such that $\mu(g)\cdot u \in U$ for all $u \in U$, $g \in G$, and otherwise we say the action is *decomposable*. The word *reducible* is used for the action when there exists both a subspace U that and a complementary subspace U^c both of which are preserved by G. For example $GL(V)$ acts irreducibly on V and $GL(V)\times GL(W)$ acts irreducibly on $V^*\otimes W$.

Definition 3.6. If W_1 and W_2 are G-modules, i.e., if $\rho_j : G \to GL(W_j)$ are linear 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 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. A module W is *trivial* if for all $g \in G$, $\rho(g) = \mathrm{Id}_W$.

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 3.7: Show that the image and kernel of a G-module homomorphism are G-modules.

Lemma 3.8 (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 $f = \lambda\,\mathrm{Id}$ for some constant λ.*

Exercise 3.9: Prove Schur's Lemma.

Example 3.10. The permutation group $\mathfrak{S}_d$ acts on $\mathbb{C}^d$ equipped with its standard basis $e_1, \cdots, e_d$ by $\sigma(e_j) = e_{\sigma(j)}$ and extending linearly. This action is reducible, it decomposes into the direct sum of the trivial representation, given by the span of $e_1 + \cdots + e_d$, and a second irreducible representation which has basis $e_1 - e_2, \cdots, e_1 - e_d$. The first is usually denoted $[d]$, and the second is denoted $[d-1,1]$. (These notations are explained in §**??**.)

There is a second one-dimensional reprsentation of $\mathfrak{S}_d$, denoted $[1^d]$, namely, letting $v \in \mathbb{C}^1$, define $\sigma \cdot v = \mathrm{sgn}(\sigma)v$. This is called the *sign reprsentation*.

Exercise 3.11: 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. Show that χ_ρ is constant on conjugacy classes of G. (In general, 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*.) Show that for representations $\rho_j : G \to GL(V_j)$, that $\chi_{\rho_1 \oplus \rho_2} = \chi_{\rho_1} + \chi_{\rho_2}$.

3.4. **Multi-linear maps and tensors.** We are interested in properties of linear and multi-linear maps that are invariant under changes of bases.

Proposition 3.12. *$f \in V^* \otimes W$ is completely determined up to $GL(V) \times GL(W)$-equivalence by its rank. In other words, there is a choice of bases such that the matrix representing f has r 1's along the diagonal and is zero elsewhere.*

If one considers linear maps $V \to V$ under changes of bases in V the situation is quite different. There are continuous parameters of invariants (the eigenvalues), and even these do not tell the whole story. We'll return to such issues later in the course. For now, we observe the following:

Proposition 3.13. *The action of $GL(V)$ on $V^* \otimes V$ is reducible. It decomposes as $V^* \otimes V = \mathbb{C}\{Id_V\} \oplus \mathfrak{sl}(V)$, where Id_V is the identity map and $\mathfrak{sl}(V)$ are the linear maps $f : V \to V$ such that, $Id_{V^*}(f) = 0$, where we consider $Id_{V^*} \in V \otimes V^* \simeq (V^* \otimes V)^*$.*

Remark 3.14. Note that in any choice of basis, $\mathfrak{sl}(V)$ will be identified with the traceless matrices, and by Exercise 3.3, the map $f \mapsto Id_{V^*}(f)$ is just $f \mapsto \mathrm{trace}(f)$, i.e., f maps to the sum of its eigenvalues. For more details see [12, §2.3,§2.5].

Exercise 3.15: Prove the decomposition. ⊚ Show that moreover $\mathbb{C}\{Id\}$ is a trivial $GL(V)$-module.

We say $V \otimes W$ is the *tensor product* of V with W. More generally, for vector spaces $A_1, \cdots, A_n$ 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..

Give $V^{\otimes d}$ the structure of a $GL(V)$-module by $g \cdot (v_1 \otimes \cdots \otimes v_d) := (g \cdot v_1) \otimes \cdots \otimes (g \cdot v_d)$ and extending linearly.

Remark 3.16. I may (and will) 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 the *re-ordering isomorphism*.

Exercise 3.17: Give $V^{\otimes d}$ the structure of a $\mathfrak{S}_d$-module by defining $\sigma \cdot (v_1 \otimes \cdots \otimes v_d) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)}$. Show that the actions of $\mathfrak{S}_d$ and $GL(V)$ on $V^{\otimes d}$ commute with each other.

Definition 3.18. A tensor $T \in V^{\otimes d}$ is said to be *symmetric* if $T(\alpha_1, \cdots, \alpha_d) = T(\alpha_{\sigma(1)}, \cdots, \alpha_{\sigma(d)})$ for all $\sigma \in \mathfrak{S}_d$, and *skew-symmetric* if $T(\alpha_1, \cdots, \alpha_d) = \mathrm{sgn}(\sigma) T(\alpha_{\sigma(1)}, \cdots, \alpha_{\sigma(d)})$ for 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.

Definition 3.19. If $l > k$, define a *contraction map* $V^{\otimes k} \times V^{* \otimes l} \to V^{* \otimes l-k}$ by

$$(X, \phi) = (x_1 \otimes \cdots \otimes x_k, \phi_1 \otimes \cdots \otimes \phi_l) \mapsto \phi \lrcorner X := \phi_1(x_1) \cdots \phi_k(x_k) \phi_{k+1} \otimes \cdots \otimes \phi_l.$$

3.5. **Exercises.**

(1) If $\dim A_i = \mathbf{a}_i$, show that $\dim(A_1\otimes\cdots\otimes A_n) = \mathbf{a}_1\cdots\mathbf{a}_n$.
(2) Show that the action of $GL(V)$ on $V^{\otimes d}$ is well-defined.
(3) Show that $Id_V\otimes Id_W = Id_{V\otimes W} \in (V^*\otimes V)\otimes(W^*\otimes W) \simeq (V\otimes W)^*\otimes(V\otimes W)$.
(4) Show that the $\mathfrak{S}_2$-module $V^{\otimes 2}$ is reducible. Show that in matrices, this decomposition corresponds to writing an $\mathbf{v}\times\mathbf{v}$ matrix as the direct sum of a symmetric matrix and a skew-symmetric matrix. Show that the action of $GL(V)$ on $V\otimes V$ also preserves this decomposition.
(5) Given $T\in S^dV$, show that T defines a homogeneous polynomial of degree d on V^*, which, for the purposes of this exercise, denote it by P_T, via $P_T(\alpha) = T(\alpha,\cdots,\alpha)$. Similarly, given a homogeneous polynomial P of degree d on V^*, define a tensor T_P by $T_P(\alpha_1,\cdots,\alpha_d)$ is the coefficient of $t_1\cdots t_d$ in the polynomial $P(t_1\alpha_1+\cdots+t_d\alpha_d)$ **check coefficients ok**. The latter process is often called the *polarization* of P. Because of this I will use the the same notation for symmetric tensors and homogenous polynomials in these notes.
(6) Let $P\in S^2\mathbb{C}^2$. If $P\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = a\alpha^2+b\alpha\beta+c\beta^2$, what is $\overline{P}(\begin{pmatrix}\alpha_1\\ \beta_1\end{pmatrix},\begin{pmatrix}\alpha_1\\ \beta_1\end{pmatrix})$?
(7) Given a linear map $f: V\to W$, one obtains linear maps $f^{\otimes k}: V^{\otimes k}\to W^{\otimes k}$ defined by $f(v_1\otimes\cdots\otimes v_k) = f(v_1)\otimes\cdots\otimes f(v_k)$. Verify that this is well defined. Show that it descends to give linear maps $f^{\circ k}: S^kV\to S^kW$ and $f^{\wedge k}:\Lambda^kV\to\Lambda^kW$. Show that if $\dim V = \dim W = \mathbf{v}$, the map $f^{\wedge\mathbf{v}}:\Lambda^{\mathbf{v}}V\to\Lambda^{\mathbf{v}}W$ is multiplication by a scalar. If $V = W$, show that the scalar is the determinant of the matrix of f with respect to any choices of basis.
(8) Let χ_ρ be as in Exercise 3.11, show that $\chi_{\rho_1\otimes\rho_2} = \chi_{\rho_1}\chi_{\rho_2}$.

3.6. **Matrix multiplication.** By the discussion above, we may view matrix multiplication

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

as a trilinear map which I will still denote by $M_{\langle U,V,W\rangle}$:

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

Exercise 3.20: Show that in bases this map is $(X,Y,Z)\mapsto\mathrm{trace}(XYZ)$.

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

Exercise 3.21: Show that as a tensor $M_{\langle U,V,W\rangle} = Id_U\otimes Id_V\otimes Id_W$.

Thanks to Exercises 3.21 and 3.15, we conclude

Proposition 3.22. *Matrix multiplication $M_{\langle U,V,W\rangle}$ is invariant under $GL(U)\times GL(V)\times GL(W)$.*

Exercise 3.23: Let $U = V = W$ so we are dealing with multiplication of square matrices. Show that $M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}$ is also invariant under the group $\mathbb{Z}_3$ of cyclic permutations of the three matrices. Since $\mathrm{trace}\,X^T = \mathrm{trace}\,X$ show that there is also a $\mathbb{Z}_2$-invariance, but note that this $\mathbb{Z}_2$ is *not* contained in the $\mathfrak{S}_3$ permuting the factors.

We conclude there is a map $\rho: GL_{\mathbf{n}}^{\times 3}\ltimes\mathbb{Z}_3\ltimes\mathbb{Z}_2\to G_{M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}}$, where $G_{M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}}\subset GL(A)\times GL(B)\times GL(C)$ denotes the subgroup preserving the tensor $M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}$.

Exercise 3.24: Determine $\ker(\rho)$. ⊚

3.7. **Complexity of bilinear maps.** 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 $T = a_1\otimes\cdots\otimes a_n$, i.e., considering $T: A_1^*\times\cdots\times A_n^*\to\mathbb{C}$, we have $T(\alpha_1,\cdots,\alpha_n) = \alpha_1(a_1)\cdots\alpha_n(a_n)$.

We will use the following measure of complexity:

Definition 3.25. 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 $\hat{\sigma}_r^0 = \hat{\sigma}_{r,A_1\otimes\cdots\otimes A_n}^0 \subset A_1 \otimes \cdots \otimes A_n$ 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., [3, §14.1]. It is roughly the number of multiplications required to compute T on a triple of vectors, and is comparable to the total number of arithmetic operations needed to compute T.

For example, letting x^i_α, y^α_u, z^u_i 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 is

$$M_{\langle \mathbf{l},\mathbf{m},\mathbf{n}\rangle} = \sum_{i=1}^{\mathbf{n}} \sum_{\alpha=1}^{\mathbf{m}} \sum_{u=1}^{\mathbf{l}} x^i_\alpha \otimes y^\alpha_u \otimes z^u_i$$

so we conclude $\mathbf{R}(M_{\langle \mathbf{n},\mathbf{m},\mathbf{l}\rangle}) \leq \mathbf{nml}$ and Strassen's algorithm shows $\mathbf{R}(M_{\langle 2,2,2\rangle}) \leq 7$.

Exercise 3.26: Write Strassen's algorithm out as a tensor. ⊚

Here is the plan:

To prove lower complexity bounds for matrix multiplication, find a polynomial on $A\otimes B\otimes C$ that
(i) vanishes on all tensors of rank at most r, and
(ii) does not vanish on $M_{\langle U,V,W\rangle}$.
Then one can conclude $\mathbf{R}(M_{\langle U,V,W\rangle}) \geq r+1$.

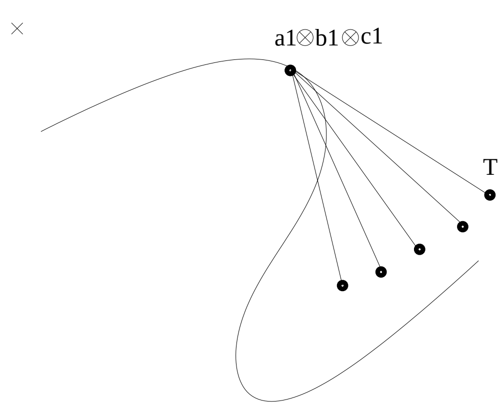
3.8. **Caution: The Fundamental theorem of linear algebra is false for tensors.**

Theorem 3.27. *Let $n \geq 3$.*

1. *If $T \in A_1 \otimes \cdots \otimes A_n$ is outside the zero set of a finite collection of polynomials (in particular outside a certain set of measure zero), then $R(T) \sim \frac{\mathbf{a}_1 \cdots \mathbf{a}_n}{\mathbf{a}_1 + \cdots + \mathbf{a}_n} >> \mathbf{a}_i$.*
2. *Rank can jump up (or down) under limits, i.e., $\hat{\sigma}_r^0$ is not closed when $r > 1$.*

For the proof, see [12, §2.4,§5.5].

To envision why rank can jump up while taking limits, consider the following picture, where the curve represents the points of $\hat{\sigma}_1^0$. The points of $\hat{\sigma}_2^0$ are those on a secant line to $\hat{\sigma}_1^0$, and the points where the rank jumps up 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.)



To see this explicitly, 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

$$\lim_{t \to 0} T(t) = a_1 \otimes b_1 \otimes c_2 + a_1 \otimes b_2 \otimes c_1 + \alpha_2 \otimes b_1 \otimes c_1$$

which does not have rank two (exercise).

I should make more precise what I mean by "closed", since *a priori* there could be two possible meanings: a set is *closed in the Zariski topology* if it is the zero set of a collection of polynomials. A set is *closed in the Euclidean topology* if it is closed under taking limits. In Theorem 3.27 both notions coincide:

Proposition 3.28. *Let $U \subset \mathbb{P}V$ or V. Then the Euclidean closure of U is contained in the Zariski closure of U. If U is a Zariski open subset of its Zariski closure and the closure is irreducible, then the two closures coincide.*

For the proof, see, e.g. [19, Thm. 2.33].

3.9. **Border rank.** To "fix" this problem, we let $\hat{\sigma}_r$ denote the closure of $\hat{\sigma}_r^0$, and define the *border rank* of $T \in A_1 \otimes \cdots \otimes A_n$, denoted $\underline{\mathbf{R}}(T)$, to be the smallest r such that $T \in \hat{\sigma}_r$.

By definition, border rank is semi-continuous. 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, we only know it is at most twice the border rank [5]. In contrast, the maximal border rank is known in many cases. For example, in $\mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$, it is known to be $\lceil \frac{m^3-1}{3m-2} \rceil$ for all $m \neq 3$, and is 5 when $m = 3$ [18]. The method of proof is a differential-geometric calculation that dates back to Terracini [23].

So our lower bounds obtained from polynomials will actually be lower bounds on border rank.

Both rank and border rank give rise to the same exponent of matrix muliplication, see §**??**.

3.10. **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 3.29. $\underline{\mathbf{R}}(T) \geq \mathrm{rank}(T_A)$.

Exercise 3.30: Prove Proposition 3.29. ⊚

Exercise 3.31: Find a choice of bases such that

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

where $x = (x^i_j)$ is $\mathbf{n} \times \mathbf{n}$.

Exercise 3.32: Show that $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n} \rangle}) \geq \mathbf{n}^2$.

Exercise 3.33: Let $\mathbf{b} = \mathbf{c}$. Show that if $T(A)$ is a diagonalizable subspace, then $\mathbf{R}(T) \leq \mathbf{b}$. This implies that if $T(A)$ is a limit of diagonalizable subspaces then $\underline{\mathbf{R}}(T) \leq \mathbf{b}$.

Exercise 3.34: Assume $\mathbf{b} = \mathbf{c}$. Show that if $T(A)$ is (simultaneously) diagonalizable, then $\mathbf{R}(T) \leq \mathbf{b}$. Thus if it is a limit of diagonalizable subspaces, then $\underline{\mathbf{R}}(T) \leq \mathbf{b}$.

Exercise 3.35: Show $\underline{\mathbf{R}}(M_{\langle \mathbf{m},\mathbf{n},1 \rangle}) = \mathbf{mn}$ and $\underline{\mathbf{R}}(M_{\langle \mathbf{m},1,1 \rangle}) = \mathbf{m}$.

4. Definitions and examples from algebraic geometry

We will need language from algebraic geometry to describe the "space" of decompositions of a tensor. In this section I establish that language as well as additional language that we will need later.

4.1. **Varieties.** Let $\hat{X} \subset V$ be the common zero set of a collection of homogeneous polynomials. Provisionally we will call such a set an *algebraic variety*. Since we only deal with homogeneous polynomials, the zero set will be invariant under re-scaling. For this, and other reasons, it will be convenient to work in projective space $\mathbb{P}V := (V\backslash 0)/\sim$ where $v \sim w$ iff $v = \lambda w$ for some $\lambda \in \mathbb{C}^*$. 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.

4.2. **First examples of varieties.**

(1) Projective space

(2) The *Veronese variety*

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

(3) The *Segre variety*

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

(4) The *Grassmannian*

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

(5) The *Chow variety*

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

By definition, projective space is a variety (the zero set of no equations) and the Chow variety is a variety, as the overline denotes Zariski closure.

Exercise 4.1: Show that the other three are varieties. Explicitly:

(1) Show that the Veronese is the zero set of the 2×2 minors of the linear map, for $P \in S^dV$, $P_{1,d-1} : V^* \to S^{d-1}V$, where the map is contraction. (V^* may also be thought of as the space of first order homogeneous linear differential operators on S^dV.)

(2) Show that the Segre is the common zero set of the two by two minors of the linear maps, given $T \in A_1 \otimes \cdots \otimes A_n$, define $T_j : A_j^* \to A_1 \otimes \cdots \otimes A_{j-1} \otimes A_{j+1} \otimes \cdots \otimes A_n$.

(3) Show that the Grassmannian is the zero set of the equations spanned 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 $Y \in \Lambda^{k-2j}V^*$ and $Z \in \Lambda^{k+2j}V^*$, recalling Definition 3.19, define $P_{Y\otimes Z}(T) := (T \lrcorner Z)(Y \lrcorner T)$ (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.

In each of the above cases, the equations are the minors of some matrix and are naturally expressed as such. Most varieties do not admit such natural determinental equations, and it will be a re-occuring theme to determine which varieties do.

While by definition, the Chow variety is a variety, one would like to have its equations. Equations for the Chow are known, see [12, §8.6.2]. However (for those familiar with the terminology), generators of the ideal of the Chow variety are not known explicitly.

For each complexity problem I will discuss, the problem could be solved by finding a set of defining equations for a variety associated to the problem and testing the equations on certain points.

4.3. **Secant Varieties.** Given a variety $X \subset \mathbb{P}V$, define the *X-rank* of $p \in V$, $R_X(p)$, to be the smallest r such that there exist $x_1, \cdots, x_r \in \hat{X}$ such that p is in the span of of $x_1, \cdots, x_r$, and the *X-border rank* $\underline{\mathbf{R}}_X(p)$ is defined to be the smallest r such that there exist curves $x_1(t), \cdots, x_r(t) \in \hat{X}$ such that p is in the span of the limiting plane $\lim_{t\to 0}\langle x_1(t), \cdots, x_r(t)\rangle$. I will use the same terms and notation for points in projective space.

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*.

When $X = \sigma_1 = Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)$ is the set of rank one tensors, $\sigma_r(X) = \sigma_r$.

One could determine the complexity of matrix multiplication by finding a set of defining equations for $\sigma_r(Seg(\mathbb{P}\mathbb{C}^{\mathbf{n}^2} \times \mathbb{P}\mathbb{C}^{\mathbf{n}^2} \times \mathbb{P}\mathbb{C}^{\mathbf{n}^2}))$ and testing those equations on $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$.

4.4. **G-varieties.** The Segre, Veronese and Grassmannian are examples of *homogeneous varieties $G/P \subset \mathbb{P}V$*, which by definition are varieties consisting of the orbit of some point $x \in \mathbb{P}V$ under the action of some group $G \subset GL(V)$. In this case $P = G_x$, the subgroup of G fixing x. Most orbits are not varieties, so one takes their *orbit closures* to obtain a variety. When $d \leq \mathbf{v}$ the Chow variety is an orbit closure, and when $r = \mathbf{a}_i$ for $1 \leq i \leq n$, $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ is an orbit closure. Higher secant varieties of the Segre are not orbit closures in general, but they are *invariant* under the action of $G := GL(A_1) \times \cdots \times GL(A_n) \subset GL(A_1 \otimes \cdots \otimes A_n)$, in the sense that if $x \in \sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ and $g \in G$, then $g \cdot x \in \sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$. A variety invariant under the action of a group is called a *G-variety*.

4.5. **Dimension.** See e.g. [12, §4.9.4] or any standard algebraic geometry text for the definition of the dimension of a variety - for now one could take it to be the number of parameters needed to describe a variety locally. A naïve parameter count gives that if $\dim X = n$, one expects $\dim \sigma_r(X) = \min\{rn + r - 1, \mathbf{v} - 1\}$, because to locate a point on $\sigma_r(X)$, one gets to pick r points on x and then a point on the $\mathbb{P}^{r-1}$ that they span. We will call this number the *expected dimension* of $\sigma_r(X)$.

4.6. **The abstract secant variety.** We are now in a position to construct a space corresponding to the set of decompositions of a tensor. We make the construction in the more general context of secant varieties.

Let $X \subset \mathbb{P}V$ be a variety. Consider the set

$$S_r(X)^0 := \{(x_1, \cdots, x_r, z) \in X^{\times r} \times \mathbb{P}V \mid z \in \operatorname{span}\{x_1, \cdots, x_r\}\} \subset Seg(X^{\times r} \times \mathbb{P}V) \subset \mathbb{P}V^{\otimes 3}$$

and let $S_r(X) := \overline{S_r(X)^0}$ denote its Zariski closure. (For those familiar with quotients, it would be better to deal with $X^{(\times r)} := X^{\times r}/\mathfrak{S}_r$.) We have a map $\pi : S_r(X) \to \mathbb{P}V$, given by projection onto the third factor and the image is $\sigma_r(X)$. We will call $S_r(X)$ the *abstract r-th secant variety of X*. Its main use for us will be to study decompositions of a point on $\sigma_r(X)$. This is because, if $\sigma_r(X)$ is of the expected dimension and $z \in \sigma_r(X)$ is a general point, $\pi^{-1}(z)$ 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 $\overline{x_j} \in \hat{x}_j$, $\overline{z} \in \hat{z}$. If the fiber of z 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 will show this is indeed the case for $M_{\langle 2,2,2\rangle} \in \sigma_7(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$.

First observe that if X is a G-variety, so $\sigma_r(X)$ 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 $\pi^{-1}(z)$, and every distinct (up to trivialities) point in its orbit will correspond to a distinct decomposition of z. Letting $q \in \pi^{-1}(X)$, 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.

4.7. **Decompositions of $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$.** Let $A = \mathbb{C}^{\mathbf{n}^2}$. The variety $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}A \times \mathbb{P}A)) \subset \mathbb{P}(A \otimes A \otimes A)$ is a $G = GL(A) \times GL(A) \times GL(A) \ltimes \mathfrak{S}_3$-variety. The tensor $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle} \in A \otimes A \otimes A$ is invariant under $SL_{\mathbf{n}} \times SL_{\mathbf{n}} \times SL_{\mathbf{n}} \ltimes \mathbb{Z}_3 \ltimes \mathbb{Z}_2$. Thus there is potentially $(3\mathbf{n}^2 - 3)$-parameters worth of decompositions for $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$ in its minimal decomposition.

Remark 4.2. This potentially huge family of optimal decompositions illustrates again just how special matrix multiplication is. Note that $\dim S_{r-1}(Seg(\mathbb{P}A \times \mathbb{P}A \times \mathbb{P}A)) = (r-1)(3\mathbf{n}^2 - 3) + r - 2 = 3\mathbf{n}^2(r-1) - 2r + 1$ while $\dim G_{M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}} = 3\mathbf{n}^2 - 3$, so $\dim S_{r-1}(Seg(\mathbb{P}A \times \mathbb{P}A \times \mathbb{P}A)) + \dim G_{M_{\langle 2\rangle}} = 3\mathbf{n}^2 r - 2r - 2$ whereas $\dim S_r(Seg(\mathbb{P}A \times \mathbb{P}A \times \mathbb{P}A)) = 3\mathbf{n}^2 r - 2r - 1$, so if the stabilizer of $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$ in $G_{M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}}$ is finite the dimensions fall just one short of one expecting an intersection with $\dim S_{r-1}(Seg(\mathbb{P}A \times \mathbb{P}A \times \mathbb{P}A))$, and this is independent of r. Note however since the ambient space is of $S_r(Seg(\mathbb{P}A \times \mathbb{P}A \times \mathbb{P}A))$ is not projective space, we would not be guaranteed an intersection even if the dimensions did match up.

5. Decompositions of tensors, especially $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$

Unfortunately we only know which σ_r it sits in optimally when $\mathbf{n} = 2$, so we now restrict to that case. We expect similar results for all $\mathbf{n}$. Before continuing, I would like to discuss the discrete part of the symmetry. For this we will have to take a short detour.

5.1. **The generalized Comon conjecture.** In 2008 there was an AIM workshop 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 conjecture that if a tensor happens to be symmetric and of rank r (as a tensor), then it will admit a decomposition as a sum of r rank one symmetric tensors. The algebraic geometers in the audience reacted to the conjecture with, to put it mildly, skepticism. The conjecture is still open, see [2] for a discussion. Here is a generalization:

Conjecture 5.1 (Generalized Comon Conjecture)**.** [4] *Let $T \in A^{\otimes d}$ and say $\mathbf{R}(T) = r$ and T is invariant under some $\Gamma \subseteq \mathfrak{S}_d$. Then T admits a Γ-invariant decomposition as a sum of r rank one tensors.*

In particular, one expects an optimal decomposition of matrix multiplication to be $\mathbb{Z}_3$-invariant.

5.2. $\mathbb{Z}_3$ **invariant tensors in** $V^{\otimes 3}$**.** To understand $\mathbb{Z}_3$-invariant decompositions, we need to decompose $V^{\otimes 3}$ under the action of $\mathfrak{S}_3 \times GL(V)$. First let's review the decomposition of $V^{\otimes 2}$ under $\mathfrak{S}_2 \times GL(V)$. If one likes to think in bases, the action on $V^{\otimes 2}$ considered as the space of $\mathbf{v} \times \mathbf{v}$ matrices X (thought of as the space of bilinear forms on V^*, where $X(v,w) = v^T X w$) by $g \in GL(V)$ is $g \cdot X = gXg^T$, and the action by the nontrivial element of $\mathfrak{S}_2$ is $X \mapsto X^T$. So the decomposition is into symmetric and skew-symmetric matrices, invariantly, is $V^{\otimes 2} = S^2V \oplus \Lambda^2 V$ where both S^2V and Λ^2V are irreducible $\mathfrak{S}_2 \times GL(V)$-modules. As an $\mathfrak{S}_2$-module, S^2V is $\binom{\mathbf{v}+1}{2}$ copies of the trivial representation and $\Lambda^2 V$ is $\binom{\mathbf{v}}{2}$ copies of the sign representation. As a $GL(V)$-module, both S^2V and Λ^2V are irreducible. (To see this note that the $GL(V)$-orbit of any non-zero vector will never lie in a proper subspace.)

There are three irreducible representations of $\mathfrak{S}_3$, the trivial $[3]$, the sign $[1,1,1]$, and the complement of the trivial for the action of $\mathfrak{S}_3$ on $\mathbb{C}^3$, $[2,1]$ (see Example 3.10).

In analogy with the $d = 2$ case, one sees $S^3V, \Lambda^3 V$ are both irreducible for $GL(V)$, and are the isotypic components respectively of the trivial representation and the sign representation.

By counting dimensions, we see we have not accounted for everything. Thus there must be a subspace corresponding to the isotypic component of $[2,1]$, of dimension $\mathbf{v}^3 - \binom{\mathbf{v}+2}{3} - \binom{\mathbf{v}}{3} = \frac{2}{3}(n^3 - n)$. To get something complementary to S^3V and $\Lambda^3 V$, consider the symmetrization map $V \otimes S^2 V \to S^3 V$. It is a $GL(V)$-map, so its kernel must be a $GL(V)$-module. Since the map is surjective, the dimension of the kernel is $\mathbf{v}\binom{\mathbf{v}+1}{2} - \binom{\mathbf{v}+2}{3} = \frac{2}{6}(n^3 - n)$, exactly half of what we are missing. To get the other half we could take, e.g., the kernel of the map $S^2V \otimes V \to S^3 V$. Both of these kernels are irreducible $GL(V)$-modules and they are isomorphic. We'll see a proof of this later, for now I record that, labeling the irreducible module $S_{21}V$, we have

$$V^{\otimes 3} = (S^3 V \otimes [3]) \oplus (S_{21} V \otimes [2,1]) \oplus (\Lambda^3 V \otimes [1,1,1]) \tag{2}$$

as an irreducible $GL(V) \times \mathfrak{S}_3$-module.

We now consider how $\mathbb{Z}_3 \subset \mathfrak{S}_3$ acts on each factor. It is generated by a cyclic permutation.

Exercise 5.2: Show that $(1,2,3)$ acts trivially on $S^3V \oplus \Lambda^3 V$ (i.e., these are the $+1$ eigenspaces for $(1,2,3)$, and $S_{21}V \otimes [2,1]$ splits into a direct sum of eigenspaces for ω and ω^2 where $\omega = e^{\frac{2\pi i}{3}}$.

For more details on this decomposition, see [12, §2.7,2.8].

We now have enough representation theory to study decompositions of $M_{\langle 2 \rangle}$.

5.3. **The decompositions of $M_{\langle 2 \rangle}$.** We are looking for a nine-parameter family of $\mathbb{Z}_3$-invariant decompositions of $M_{\langle 2 \rangle}$, parametrized by $SL_2 \times SL_2 \times SL_2$. To keep better track of this, we return to our invariant description of $M_{\langle 2 \rangle}$: let $U, V, W = \mathbb{C}^2$, and write $M_{\langle 2 \rangle} = Id_U \otimes Id_V \otimes Id_W \in (U^* \otimes V) \otimes (V^* \otimes W) \otimes (W^* \otimes U) = A \otimes B \otimes C$. To take advantage of $\mathbb{Z}_3$, we will need to identify A, B, C, which we may do by identifying U, V, W. Explicitly, choose $a_0 : U \to V$ identifying $U \simeq V$ and $b_0 : V \to W$, which determines $c_0 = {a_0}^{-1} {b_0}^{-1} : W \to U$. These elements, in bases for V, W induced from those of U, will correspond to identity matrices. The choices have already killed off two of our SL_2's and we are only left with one, say $SL(U)$.

Now observe that the first term in Strassen's algorithm, when it is expressed as a tensor (see [12, eqn. 2.4.5]), is $Id_2 \otimes Id_2 \otimes Id_2$, indicating our first term in the expression should be $a_0 \otimes b_0 \otimes c_0$. We now need to exactly use up $SL(U)$'s worth of freedom to get the rest of the expression. Now SL_2 acts three-transitively on $\mathbb{CP}^1$, that is, given any three distinct points in $\mathbb{CP}^1 \simeq S^2$, there exists exactly one element of SL_2 that will move them to e.g., $0, 1, \infty$, i.e., $[1,0], [1,1], [0,1]$. So we think of our algorithm as being obtained from three distinct points $[u_1], [u_2], [u_3] \in \mathbb{P}U$. Together with the choice of a_0, b_0, this should uniquely determine the algorithm. Notice that our choices also give us three points $[{u_1}^{\perp}], [{u_2}^{\perp}], [{u_3}^{\perp}] \in \mathbb{P}U^*$, as the annihilator of a line in $\mathbb{C}^2$ is a line in $\mathbb{C}^{2*}$. Applying a_0 etc.. gives us triples of points in each of $\mathbb{P}V, \mathbb{P}V^*, \mathbb{P}W, \mathbb{P}W^*$ as well.

Examining Strassen's algorithm, notice that all terms in it except the first are represented by a triple of rank one matrices, and we now have a collection of rank one matrices up to scale: ${u_i}^{\perp} \otimes v_j \in U^* \otimes V$, ${v_i}^{\perp} \otimes w_j \in V^* \otimes W$, ${w_i}^{\perp} \otimes u_j \in W^* \otimes U$. Set $\lambda = {u_1}^{\perp}(v_2) {v_3}^{\perp}(w_1) {w_2}^{\perp}(u_3)$ *need to check normalization***

Theorem 5.3. [4] *The following is a 9 parameter family, parameterized by $SL(U) \times SL(V) \times SL(W)$, of $\mathbb{Z}_3$-invariant rank seven expressions for $M_{\langle 2 \rangle}$:*

$$\begin{aligned} M_{\langle 2 \rangle} = & a_0 \otimes b_0 \otimes c_0 \\ & + \frac{1}{\lambda} \langle ({u_1}^{\perp} \otimes v_3) \otimes ({v_2}^{\perp} \otimes w_1) \otimes ({w_3}^{\perp} \otimes u_2) \rangle_{\mathbb{Z}_3} \\ & + \frac{1}{\lambda} \langle ({u_1}^{\perp} \otimes v_2) \otimes ({v_3}^{\perp} \otimes w_1) \otimes ({w_2}^{\perp} \otimes u_3) \rangle_{\mathbb{Z}_3} \end{aligned} \tag{3}$$

where, using our identifications $U \simeq V \simeq W$, $\langle \alpha\otimes\beta\otimes\gamma\rangle_{\mathbb{Z}_3} := \alpha\otimes\beta\otimes\gamma + \beta\otimes\gamma\otimes\alpha + \gamma\otimes\alpha\otimes\beta$.

To prove this is indeed matrix multiplication, of course one could check it directly, e.g., taking $u_1 = (1,0)$, $u_2 = (0,1)$, $u_3 = (1,1)$, ${u_1}^{\perp} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, ${u_2}^{\perp} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, ${u_3}^{\perp} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, and $a_0 = b_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

However instead I will give a geometric proof.

Proof. In §*** we will prove that the matrix multiplication tensor is characterized by its symmetry group, in the sense that no other tensor in $\mathbb{C}^{\mathbf{n}^2}\otimes\mathbb{C}^{\mathbf{n}^2}\otimes\mathbb{C}^{\mathbf{n}^2}$ is preserved by a group that contains $SL_{\mathbf{n}}\times SL_{\mathbf{n}}\times SL_{\mathbf{n}}$. Thus to prove we have equality it is sufficient to show the right hand side is, preserved by $SL(U)\times SL(V)\times SL(W)$, or equivalently (via differentiation) annihilated by $\mathfrak{sl}(U)+\mathfrak{sl}(V)+\mathfrak{sl}(W)$.

By $\mathbb{Z}_3$-invariance it is sufficient to show the right hand side is annihilated by $\mathfrak{sl}(U)$, which has basis ${u_1}^{\perp}\otimes u_1, {u_2}^{\perp}\otimes u_2, {u_3}^{\perp}\otimes u_3$.

By the symmetry of the algorithm it is sufficient to show it is annihilated by ${u_1}^{\perp}\otimes u_1$, which is easy to check.

□

Exercise 5.4: Verify that ${u_1}^{\perp}\otimes u_1, {u_2}^{\perp}\otimes u_2, {u_3}^{\perp}\otimes u_3$ is a basis of $\mathfrak{s}(U)$.

Exercise 5.5: Verify that the right hand side of (3) is annhilated by ${u_1}^{\perp}\otimes u_1$ to complete the proof.

6. Strassen's equations and generalizations

6.1. **Our first equations revisited.** Recall that our lower bound $\underline{\mathbf{R}}(M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \geq \mathbf{n}^2$ was obtained from the following equations: for $T \in A \otimes B \otimes C$, the maximal minors of the linear map $T_A : A^* \to B \otimes C$ (and the two similar maps which I suppress reference to in what follows). To extract more information, we should examine what the image of this map looks like for a tensor of border rank a little larger than the dimensions of A, B, C. Assume $\mathbf{b} = \mathbf{c}$ so the image is a space of linear maps $\mathbb{C}^{\mathbf{b}} \to \mathbb{C}^{\mathbf{b}}$. First say $\mathbf{R}(T) = \mathbf{b}$. Then $T_A(A^*)$ will be spanned by $\mathbf{b}$ rank one linear maps.

Exercise 6.1: Show that, $\mathbf{R}(T) = \mathbf{b}$ if and only if $T_A(A^*)$ is spanned by $\mathbf{b}$ rank one linear maps.

Assume for the moment furthermore that $\mathbf{a} = \mathbf{b}$. 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 $T_A(A^*)$ will be diagonalizable. Being diagonalizable is not stable under degenerations of linear maps. In other words, the set of diagonalizable matrices is *not* an algebraic variety. However being an *abelian* subspace of linear maps is stable. Namely a space of linear maps $U \in End(V)$ is abelian iff $[u_i, u_j] = 0$ for $u_1, \cdots, u_{\mathbf{u}}$ a basis of U. These quadratic equations on U would give us additional equations for $\sigma_{\mathbf{a}}$, except that our family of linear maps is between two different vector spaces, so we cannot compose them or take their commutators. If we pick one invertible map and use it to identify B with C, we could then test the commutators of the remaining maps, using this identification. However this causes a problem if no linear map in $T_A(A^*)$ is invertible.

6.2. **A substitute for the inverse of a linear map.** We can avoid this problem as follows: Recall a linear map $f : V \to W$ induces linear maps $f^{\wedge k} : \Lambda^k V \to \Lambda^k W$. Assume $\mathbf{v} = \mathbf{w}$. Fixing a volume form on V, i.e., a nonzero element of $\Omega \in \Lambda^{\mathbf{v}} V^*$, we may identify $\Lambda^{\mathbf{v}-1} V \simeq V^*$ by the map

$$\begin{aligned} \Lambda^{\mathbf{v}-1} V &\to V^* \\ \alpha &\mapsto \alpha \lrcorner \Omega. \end{aligned}$$

Fixing volume elements, we may consider $f^{\wedge \mathbf{v}-1} : V^* \to W^*$, or, taking transpose, as a map $W \to V$.

Exercise 6.2: What is the relationship between $f^{\wedge \mathbf{v}-1}$ and f^{-1} when f is invertible?

6.3. **Additional equations for $\sigma_{\mathbf{b}}$.** Combining the discussions of the previous two subsections, we obtain:

Proposition 6.3. *Let $T \in A \otimes B \otimes C$ and assume $\mathbf{b} = \mathbf{c}$. Then $\underline{\mathbf{R}}(T) \leq \mathbf{b}$ implies that for all $\alpha, \alpha_1, \alpha_2 \in A^*$, the size two minors of the linear map $[(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_1), (T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_2)]$ are all zero. Here $T(\alpha) := T_A(\alpha)$, and the bracket denotes the commutator of linear maps.*

Note that in bases, these equations are of degree two in the entries of the commutator, the entries of $T(\alpha_j)$ are linear in the entries of T, and the entries of $(T(\alpha)^{\wedge \mathbf{a}-1})^T$ are of degree $\mathbf{a}-1$ in the entries of T, so the polynomials are of degree $2((\mathbf{a}-1)+1) = 4\mathbf{a}$. We can lower the degree in two ways. First note that we have

$$\begin{aligned} &[(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_1), (T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_2)] \\ &\qquad = (T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_1)(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_2) - T(\alpha_2)(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_1) \end{aligned}$$

we can work instead with minors of

$$T(\alpha_1)(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_2) - T(\alpha_2)(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_1),$$

dropping the degree by $2(\mathbf{a}-1)$. (This could potentially enlarge the zero set if $T(\alpha)$ is not invertible.) More significantly, both for dropping the degree and for later generalizations, the border rank of T cannot increase if we restrict to subspaces.

Exercise 6.4: Prove that if $T \in A\otimes B\otimes C$ and $T' := T|_{A'\otimes B'\otimes C'}$ for some $A' \subseteq A^*$, $B' \subseteq B^*$, $C' \subseteq C^*$, then $\underline{\mathbf{R}}(T) \geq \underline{\mathbf{R}}(T')$. ⊚

Thus we have:

Proposition 6.5 (Strassen). *Let $T \in A\otimes B\otimes C$. Then $\underline{\mathbf{R}}(T) \leq \mathbf{a}$ implies that for all $\alpha, \alpha_1, \alpha_2 \in A^*$, if one constructs the linear maps $T(\alpha_j) : A^* \to C$ and $(T(\alpha)^{\wedge \mathbf{a}-1})^T : C^* \to A^*$, that the size two minors of the linear map*

$$T(\alpha_1)(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_2) - T(\alpha_2)(T(\alpha)^{\wedge \mathbf{a}-1})^T T(\alpha_1)$$

are all zero. These equations are of degree $\mathbf{a}+1$. As one ranges over all possible choices of subspaces, elements and bases, one spans a module of polynomials for $\sigma_{\mathbf{a}}$.

6.4. Strassen's equations.

If T is "close to" having rank $\mathbf{a}$, one expects that $T_A(A^*)$ will be "close to" being abelian. The following theorem makes this precise:

Theorem 6.6 (Strassen). [21] *Let $T \in A\otimes B\otimes C$. Then $\underline{\mathbf{R}}(T) \leq r$ implies that for all $\alpha, \alpha_1, \alpha_2 \in A^*$, the size $r(\mathbf{a}-1)+1$ minors of the linear map*

$$T(\alpha_1)(T(\alpha)^{\wedge 2})^T T(\alpha_2) - T(\alpha_2)(T(\alpha)^{\wedge 2})^T T(\alpha_1)$$

are all zero.

So we now have potential tests for border rank for tensors in $\mathbb{C}^N\otimes\mathbb{C}^N\otimes\mathbb{C}^N$ up to $r = \frac{3}{2}N$.

A natural question arises: we actually have 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?

Representation theory will enable us to answer these questions.

6.5. Exercises.

(1) Write out Strassen's equations for $\sigma_3(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2))$ in bases.
(2) Write out the proof of Theorem 6.6.
(3) Show that $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \geq \frac{3}{2}n^2$. ⊚
(4) Show that when $r = 1$, any two of the sets are not completely redundant, but the third set is contained in the span of the other two. ⊚
(5) Show that if $T \in \sigma_r$ then for all $\alpha_1, \cdots, \alpha_4 \in A^*$,
$$\text{rank}[(T(\alpha_1)^{\wedge \mathbf{a}-1})^T T(\alpha_2), (T(\alpha_3)^{\wedge \mathbf{a}-1})^T T(\alpha_4)] \leq 3(r-\mathbf{a}).$$

6.6. Reformulation of Strassen's equations.

Let $\dim A = 3$ and $\dim B = \dim C = \mathbf{b}$. We augment the linear map $T_B : B^* \to A\otimes C$ by tensoring it with Id_A, to get a linear map

$$T_B\otimes Id_A : B^* \to A \to A\otimes A\otimes C.$$

So far we have done nothing, but the target of this map decomposes as $(\Lambda^2 A\otimes C) \oplus (S^2 A\otimes C)$, so we may consider the two linear maps, which I will denote

$$T_A^{\wedge} : A\otimes B^* \to \Lambda^2 A\otimes C \text{ and } T_A^{\circ} : A\otimes B^* \to S^2 A\otimes C. \tag{4}$$

Exercise 6.7: Show that if $T = a\otimes b\otimes c$ is a rank one tensor, then $\text{rank}(T_A^{\wedge}) = 2$ and $\text{rank}(T_A^{\circ}) = 3$.

Thus, if T has rank r, $\mathrm{rank}(T_A^\wedge) \le 2r$ and $\mathrm{rank}(T_A^\circ) \le 3r$. Since the dimension of the sources are $3\mathbf{b}$, by this method one respectively gets potential equations for σ_r up to $r = \frac{3}{2}\mathbf{b} - 1$ and $r = \mathbf{b} - 1$. Thus only the first gives interesting bounds. In fact the first set has the same zero set as Strassen's equations, as I now show in coordinates.

Remark 6.8. As presented, this derivation of the equations seems a bit "rabbit out of the hat". However it comes from a long tradition of finding determinantal equations for algebraic varieties that is out of the scope of this course. 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 had been discoved by Barth in this context. Variants of Strassen's equations had been discovered previous to Strassen by Frahm-Toeplitz and Aronhold. See [12, §3.8.5] for precise statements.

Let a_1, a_2, a_3 be a basis of A, choose bases of B, C so for any $T \in A\otimes B\otimes C$ we may write $T = a_1\otimes X_1 + a_2\otimes X_2 + a_3\otimes X_3$. Then $T_A^{\wedge 1}$ will be expressed by a $3\mathbf{b} \times 3\mathbf{b}$ matrix. Ordering the basis of $A\otimes B^*$ by $a_1\otimes\beta^1, \cdots, a_1\otimes\beta^{\mathbf{b}}, a_2\otimes\beta^1, \cdots, a_1\otimes\beta^{\mathbf{b}}, a_3\otimes\beta^1, \cdots, a_1\otimes\beta^{\mathbf{b}}$, and that of $\Lambda^2 A\otimes C$ by $(a_2\wedge a_3)\otimes c_1, \cdots, (a_2\wedge a_3)\otimes c_{\mathbf{b}}, (a_1\wedge a_3)\otimes c_1, \cdots, (a_1\wedge a_3)\otimes c_{\mathbf{b}}, (a_1\wedge a_2)\otimes c_1, \cdots, (a_1\wedge a_2)\otimes c_{\mathbf{b}}$, we obtain the block matrix

$$T_A^\wedge = \begin{pmatrix} 0 & X_1 & -X_2 \\ -X_1 & 0 & X_3 \\ X_2 & -X_3 & 0 \end{pmatrix} \tag{5}$$

Now recall the basic identity about determinants, assuming W is invertible:

$$\det\begin{pmatrix} X & Y \\ Z & W \end{pmatrix} = \det(W)\det(X - YW^{-1}Z), \tag{6}$$

Now assume X_3 is invertible and change bases such that it is the identity matrix. Using the $(\mathbf{b}, 2\mathbf{b}) \times (\mathbf{b}, 2\mathbf{b})$ blocking (so $X = 0$ in (6))

$$\det Mat(T_A^\wedge) = \det(X_1X_2 - X_2X_1) = \det([X_1, X_2]).$$

Now observe that if $T(\alpha_1)$ is of full rank and normalized to be the identity in the previous presentation of Strassen's equations, we obtain the same equation.

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

$$T_A^{\wedge p} : B^*\otimes\Lambda^p A \to \Lambda^{p+1}A\otimes C \tag{7}$$

obtained by first taking $T_B\otimes Id_{\Lambda^p A} : B^*\otimes\Lambda^p A \to \Lambda^p A\otimes A\otimes C$, and then projecting to $\Lambda^{p+1}A\otimes C$. Call the map $T_A^{\wedge p}$ a *Koszul flattening.* Note that if $T = a\otimes b\otimes c$ has rank one, then $\mathrm{rank}(T_A^{\wedge p}) = \binom{2p}{p}$ as the image is $a \wedge \Lambda^p A\otimes c$. In summary:

Proposition 6.9. *If* $\mathrm{rank}(T_A^{\wedge p}) \ge r$, *then*

$$\underline{\mathbf{R}}(T) \ge \frac{r}{\binom{2p}{p}}.$$

Since both source and target have dimension $\binom{2p+1}{p}\mathbf{b}$, we potentially obtain equations up to border rank

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

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 $2p+1$.

6.8. Exercises.

(1) Consider the symmetric cousins of Koszul flattenings: for $P \in S^d V$, first take the map $P_{d-k,k} : S^{d-k}V^* \to S^k V$, tensor it with $Id_{\Lambda^p}V$, and project to get a map $P^{\wedge p}_{d-k,k} : S^{d-k}V^* \otimes \Lambda^p V \to S^{k-1}V \otimes \Lambda^{p+1}V$. What is best the potential lower bound on symmetric border rank that one could attain this way? For those familiar with the rudiments of differential geometry, note that the map $S^kV\otimes\Lambda^pV \to S^{k-1}V\otimes\Lambda^{p+1}V$ is just the exterior derivative. We will see later that $S^kV\otimes\Lambda^pV$ decomposes into the direct sum of two irreducible modules, called $S_{k,1^p}V$ and $S_{k+1,1^{p-1}}V$, so $P^{\wedge p}_{d-k,k}$ can at best surject onto $S_{k,1^p}V$.

(2) For those handy with computers: determine the rank of $(M_{\langle 3\rangle})^{\wedge 4}_A$. What lower bound does this give on the border rank?

Answer for those not handy: 13, the same as Strassen.

(3) For those handy with computers: determine the rank of $(M_{\langle 4\rangle})^{\wedge 7}_A$ restricting to a 15-dimensional subspace of $\mathbb{C}^{16}$. What lower bound does this give on the border rank?

This beats Strassen by one. For those familiar with Lickteig's bound [17]: $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \geq \frac{3n^2}{2} + \frac{n}{2} - 1$, you may now be thinking this is going to lead to Lickteig's bound.

(4) For those handy with computers: determine the rank of $(M_{\langle 5\rangle})^{\wedge 12}_A$. What lower bound does this give on the border rank?

Even more evidence for this being Lickteig's bound.

(5) For those handy with computers: determine the rank of $(M_{\langle 6\rangle})^{\wedge 17}_A$. What lower bound does this give on the border rank?

Now it looks like something interesting, as $58 > 56 = \frac{3(6^2)}{2} + \frac{6}{2} - 1$.

How would one possibly determine the rank for arbitrary n?

For any tensor $T = t^{ijk}a_i\otimes b_j\otimes c_k$, where a_i, b_j, c_k are bases of A, B, C and I use the summation convention, and any $f_1, \cdots, f_p \in A$ and $\beta \in B^*$,

$$T^{\wedge p}_A(\beta\otimes f_1\wedge\cdots\wedge f_p) = t^{ijk}\beta(b_j)a_i\wedge f_1\wedge\cdots\wedge f_p\otimes c_k \tag{8}$$

Consider the skew-symmetrization map $A\otimes\Lambda^pA \to \Lambda^{p+1}A$. It is a surjective $GL(A)$-module map, so its kernel is a $GL(A)$-module generalizing $S_{21}A$ that we saw in §5.2. Its isomorphism class as a $GL(A)$-module is denoted $S_{2,1^{p-1}}A$. We will prove later that it is irreducible and distinct from the irreducible module $\Lambda^{p+1}A =: S_{1^{p+1}}A$.

Thus the projection map, since it is a $GL(A)\times GL(C)$-module map, $A\otimes\Lambda^pA\otimes C \to \Lambda^{p+1}A\otimes C$ has kernel isomorphic to $S_{2,1^{p-1}}A\otimes C$, so if we had a way to compute the dimension of $S_{2,1^{p-1}}A\otimes C\cap (T_B(B^*)\otimes\Lambda^pA)$, we would be able to solve our problem.

Recall the invariant description of matrix multiplication,

$$M = Id_U\otimes Id_V\otimes Id_W \in (U^*\otimes V)\otimes(V^*\otimes W)\otimes(W^*\otimes U).$$

Our map is $(M^{\langle U,V,W\rangle})^{\wedge p}_A : V\otimes W^*\otimes\Lambda^p(U^*\otimes V) \to \Lambda^{p+1}(U^*\otimes V)\otimes(W^*\otimes U)$. The presence of $Id_W = Id_{W^*}$ means the map factors as $M^{\wedge p}_A = (M_{\langle \mathbf{u},\mathbf{v},1\rangle})^{\wedge p}_A\otimes Id_{W^*}$, where

$$(M_{\langle \mathbf{u},\mathbf{v},1\rangle})^{\wedge p}_A : V\otimes\Lambda^p(U^*\otimes V) \to \Lambda^{p+1}(U^*\otimes V)\otimes U. \tag{9}$$

Equation (8), when $T = M_{\langle \mathbf{u},\mathbf{v},1\rangle}$ becomes, for $v, e_j \in V$ and $\xi^j \in U$, letting $u_1, \cdots, u_{\mathbf{u}}$ be a basis of u with dual basis $\gamma^1, \cdots, \gamma^{\mathbf{u}}$ of U^*, so $Id_U = \sum_{s=1}^n \gamma^s \otimes u_s$,

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

To compute the kernel of $(M_{\langle \mathbf{u},\mathbf{v},1\rangle})_A^{\wedge p}$, and more generally any G-module map, one can exploit representation theory as follows: one decomposes the source and target as G-modules in its isotypic decomposition (in our case this will be a decomposition into irreducible modules, which makes the task much easier). Then, thanks to Schur's lemma, the map decomposes into a direct sum of maps between distinct spaces. In particular, any module appearing in the source that does not appear in the target must be in the kernel. Then, in the case where all isotypic components are irreducible, for each irreducible module appearing in the source and target, we will simply check the map on a single vector. Moreover, representation theory will distinguish an "easiest" vector to compute with (called a *highest weight vector*). So our immediate goals are to develop enough representation theory to decompose $V \otimes \Lambda^p(U^* \otimes V)$ and $\Lambda^{p+1}(U^* \otimes V) \otimes U$, and to be able to write down a vector from each irreducible component.

6.9. **Koszul flattenings in coordinates.** It will be useful to view $T_A^{\wedge p}$ in coordinates. Let $\dim A = 2p+1$. Write $T = a_0 \otimes X_0 + \cdots + a_{2p} \otimes X_{2p}$. The expression of $T_A^{\wedge p}$ in bases is as follows: write $a_I := a_{i_1} \wedge \cdots \wedge a_{i_p}$ for $\Lambda^p A$, require that the first $\binom{2p}{p-1}$ basis vectors 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, only we write $0K$ since a zero index is included. Then the resulting matrix is of the form

$$\begin{pmatrix} 0 & Q \\ \tilde{Q} & R \end{pmatrix} \tag{10}$$

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 $X_1, \cdots, X_{2p}$ 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 (6). If X_0 is the identity matrix, when $p = 1$ we have $Q\tilde{Q} = [X_1, X_2]$ and when $p = 2$

$$Q\tilde{Q} = \begin{pmatrix} 0 & [X_1, X_2] & [X_1, X_3] & [X_1, X_4] \\ [X_2, X_1] & 0 & [X_2, X_3] & [X_2, X_4] \\ [X_3, X_1] & [X_3, X_2] & 0 & [X_3, X_4] \\ [X_4, X_1] & [X_4, X_2] & [X_4, X_3] & 0 \end{pmatrix}. \tag{11}$$

In general, when X_0 is the identity matrix, $Q\tilde{Q}$ is a block $\binom{2p}{p-1}\mathbf{b} \times \binom{2p}{p-1}\mathbf{b}$ matrix whose block entries are either zero or commutators $[X_i, X_j]$.

6.10. **The kernel as a module.**

Remark 6.10. Before covering this section, in lecture I will present an introduction to elementary representation theory of the general linear group.

Assume $\mathbf{b} \le \mathbf{c}$, so $\mathbf{n} \le \mathbf{m}$. Let M, N, L be vector spaces of dimensions $\mathbf{m}, \mathbf{n}, \mathbf{l}$. Write $A = M \otimes N^*$, $B = N \otimes L^*$, $C = L \otimes M^*$, so $\mathbf{a} = \mathbf{mn}$, $\mathbf{b} = \mathbf{nl}$, $\mathbf{c} = \mathbf{ml}$. Recall that the matrix multiplication operator $M_{<\mathbf{m},\mathbf{n},\mathbf{l}>}$ is $M_{<\mathbf{m},\mathbf{n},\mathbf{l}>} = Id_M \otimes Id_N \otimes Id_L \in A \otimes B \otimes C$. Let $U = N^*$.

Recall the map $M^{\wedge p}_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}$ factors as $M^{\wedge p}_{\langle 1,\mathbf{m},\mathbf{n}\rangle} \otimes Id_L$, so $\mathrm{rank}(M^{\wedge p}_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}) = \mathbf{l}\mathrm{rank}(M^{\wedge p}_{\langle 1,\mathbf{m},\mathbf{n}\rangle})$.

Example 6.11. Consider the case $\mathbf{m} = \mathbf{n} = 3$, take $p = 4$, so

$$M^{\wedge 4}_{\langle 1,3,3\rangle} : \Lambda^4(M \otimes U) \otimes M \to \Lambda^5(M \otimes U) \otimes U^*$$

Since $\mathbf{m} = \mathbf{u} = 3$, by Exercise **??**, $\Lambda^4(M \otimes U) = S_{211}M \otimes S_{31}U \oplus S_{22}M \otimes S_{22}U \oplus S_{31}M \otimes S_{211}U$, and decomposing the $GL(M)$ factors in $\Lambda^4(M \otimes U) \otimes M$ gives, by the Pieri rule (see Theorem **??** and the pictures below it),

$$\begin{aligned}\Lambda^4(M \otimes U) \otimes M &= (S_{211}M \otimes S_{31}U \oplus S_{22}M \otimes S_{22}U \oplus S_{31}M \otimes S_{211}U) \otimes M \\ &= (S_{311}M \oplus S_{221}M) \otimes S_{31}U \oplus (S_{32}M \oplus S_{221}M) \otimes S_{22}U \oplus (S_{41}M \oplus S_{32}M \oplus S_{311}M) \otimes S_{211}U.\end{aligned}$$

The kernel must contain all modules that do not appear in

$$\begin{aligned}\Lambda^5(M \otimes U) \otimes U^* &= (S_{32}M \otimes S_{221}U \oplus S_{221}M \otimes S_{32}U) \otimes U^* \\ &= S_{32}M \otimes S_{221}U \oplus S_{32}M \otimes S_{22}U \oplus S_{221}M \otimes S_{31}U \oplus \oplus S_{221}M \otimes S_{22}U)\end{aligned}$$

So $S_{41}M \otimes S_{211}U$ must be in the kernel.

Hence the rank of $M^{\wedge 4}_{\langle 1,3,3\rangle}$ is at most $3 \cdot \binom{9}{4} - 24$ which would give the lower bound $\underline{\mathbf{R}}(M_{\langle 3,3,\mathbf{l}\rangle}) \ge \lceil \frac{306\mathbf{l}}{\binom{8}{4}} \rceil = \lceil \frac{306\mathbf{l}}{70} \rceil$ which is 14 when $\mathbf{l} = 3$. By Exercise 6.8.2, the rank of $M^{\wedge 4}_{\langle 1,3,3\rangle}$ is 102, so all other modules map injectively.

More generally one can show:

Proposition 6.12. [15] *$\ker(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle})^{\wedge p}_A = \oplus_\pi S_{\pi'} M \otimes S_{\pi+(1)} U \otimes L$ where the summation is over partitions $\pi = (\mathbf{m}, \nu_1, \cdots, \nu_{\mathbf{n}-1})$ where $\nu = (\nu_1, \cdots, \nu_{\mathbf{n}-1})$ is a partition of $p - \mathbf{m}$, $\nu_1 \le \mathbf{m}$ and $\pi + (1) = (\mathbf{m}+1, \nu_1, \cdots, \nu_{\mathbf{n}-1})$.*

Proposition 6.12 can be used to prove a lower bound $\underline{\mathbf{R}}(M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}) \ge 2\mathbf{n}^2 - O(\mathbf{n})$, but a better bound can be obtained by the choice of a good subspace of A, as I explain in the next section.

6.11. **A $2\mathbf{n}^2 - \mathbf{n}$ lower bound for the border rank of matrix multiplication.**

Theorem 6.13. [15] *Let $\mathbf{n} \le \mathbf{m}$. Then*

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

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

Proof. The essential idea is to choose a subspace $A' \subset M \otimes U$ on which the "restriction" of $M^{\wedge p}_{\langle 1,\mathbf{m},\mathbf{n}\rangle}$ becomes injective for $p = \mathbf{n} - 1$. Take a vector space W of dimension 2, and fix isomorphisms $U \simeq S^{\mathbf{n}-1}W^*$, $M \simeq S^{\mathbf{m}-1}W^*$. Let A' be the $SL(W)$-direct summand $S^{\mathbf{m}+\mathbf{n}-2}W^* \subset S^{\mathbf{n}-1}W^* \otimes S^{\mathbf{n}-1}W^* = M \otimes U$.

If $f \in S^\alpha W$ and $g \in S^\beta W^*$ (with $\beta \le \alpha$) then we can perform the contraction $g \lrcorner f \in S^{\alpha-\beta}W$ (see Definition 3.19). In the case $f = l^\alpha$ is the power of a linear form l, then $g \lrcorner l^\alpha = g(l) l^{\alpha-\beta}$ (in writing $g(l)$, g is being considered as a polynomial of degree β on W^*), so that $g \lrcorner l^\alpha = 0$ if and only if l is a root of g.

Since $SL(W)$ is reductive, there is a unique $SL(W)$-complement A'' to A' in A, so the projection $M\otimes U \to A'$ is well defined as is $\pi : A\otimes B\otimes C \to A'\otimes B\otimes C$. Let $T' = \pi(M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle})$. Write $T'^{\wedge 4}_{A'} = \tilde{T}\otimes Id_L$, where

$$\tilde{T} : U \otimes \Lambda^{\mathbf{n}-1} A' \longrightarrow M^* \otimes \Lambda^{\mathbf{n}} A' \tag{12}$$

is the factored map. I claim (12) is injective. (Note that when $\mathbf{n} = \mathbf{m}$ the source and target space of (12) are dual to each other.)

Consider the transposed map $S^{\mathbf{m}-1}W^*\otimes\Lambda^{\mathbf{n}} S^{\mathbf{m}+\mathbf{n}-2}W \to S^{\mathbf{n}-1}W\otimes\Lambda^{\mathbf{n}-1}S^{\mathbf{m}+\mathbf{n}-2}W$. It is defined as follows on decomposable elements (and then extended by linearity):

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

This dual map is surjective: Let $l^{\mathbf{n}-1}\otimes(l_1^{\mathbf{m}+\mathbf{n}-2} \wedge \cdots \wedge l_{\mathbf{n}-1}^{\mathbf{m}+\mathbf{n}-2}) \in S^{\mathbf{n}-1}W\otimes\Lambda^{\mathbf{n}-1}S^{\mathbf{m}+\mathbf{n}-2}W$ with $l_i \in W$. 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 $\mathbf{n} \le \mathbf{m}$, there is a polynomial $g \in S^{\mathbf{m}-1}W^*$ 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.

Since the image is closed (being a linear space), the condition that l is distinct from the l_i may be removed by taking limits.

Observe that an element of rank one in $A'\otimes B\otimes C$ induces a map $B^*\otimes\Lambda^{n-1}A' \to C\otimes\Lambda^n A'$ of rank $\binom{\mathbf{n}+\mathbf{m}-2}{\mathbf{n}-1}$.

By Exercise 6.4, the border rank of $M_{\langle \mathbf{m},\mathbf{n},\mathbf{l}\rangle}$ must be at least the border rank of $T' \in A'\otimes B\otimes C$, and by Exercise 6.4 we conclude

$$\underline{\mathbf{R}}(T') \ge \frac{\dim B^* \otimes \Lambda^{\mathbf{n}-1} A'}{\binom{\mathbf{n}+\mathbf{m}-2}{\mathbf{n}-1}} = \mathbf{nl}\frac{\binom{\mathbf{n}+\mathbf{m}-1}{\mathbf{n}-1}}{\binom{\mathbf{n}+\mathbf{m}-2}{\mathbf{n}-1}} = \frac{\mathbf{nl}\,(\mathbf{n}+\mathbf{m}-1)}{\mathbf{m}}.$$

$\square$

Remark 6.14. The Koszul flattenings potentially give equations for border rank in $\mathbb{C}^N\otimes\mathbb{C}^N\otimes\mathbb{C}^N$ up to $2N-1$. In fact the equations are indeed nontrivial in this range (see [11]). This shows that there are nontrivial equations for border rank $\le 2\mathbf{n}^2 - 1$ that *are* satisfied by $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$. This result had the effect of changing my perspective on the conjecture that the exponent of matrix multiplication is two. (I had previously thought the conjecture was ridiculous.)

At this writing, the $2N-1$ bound for equations for border rank is a wall for future progress. There are other equations in this range - defined first in [8] and later proved nontrivial in [11]. I am aware of many conjectural modules of equations in the ranges up to $2N-1$ (conjectural in the sense that they have not yet been proven to be not identically zero, which is usually the hardest part), but I am not aware of even any conjectural equations for border rank $2N$ or above. In the next section, I describe some of these equations.

7. The equations of [8]

Given $T = \sum_{j=0}^{\mathbf{a}-1} a_j \otimes X_j$ with a_j a basis of A, $X_j \in B\otimes C$, $\dim A = \mathbf{a}$ and $\dim B = \dim C = m$, assume X_0 is of full rank and use it to identify C with B^*. The equations defined by B. Griesser in [8] are stated as: if the border rank of T is at most r, with $m+1 \le r \le 2m-1$, then the space of endomorphisms $\langle [X_1, X_2], \cdots, [X_1, X_{\mathbf{a}-1}]\rangle \subset \mathfrak{sl}(B)$ is such that there exists $E \in G(2m-r, B)$, with $\dim(\langle [X_1, X_2], \cdots, [X_1, X_{\mathbf{a}-1}]\rangle(E)) \le r-m$. Here $\langle \dots \rangle$ denotes the linear span and $G(k, B)$ the Grassmannian of k planes in B. Compared with the minors of $T_A^{\wedge p}$, here one is just examining the last block column of the matrix appearing in its coordinate expression, but one is apparently extracting more refined information from it.

Assuming T is sufficiently generic, we may choose X_1 to be diagonal with distinct entries on the diagonal (a general element of $\mathfrak{sl}(B)$, the space of traceless endomorphisms, is diagonalizable with distinct eigenvalues), and this is a generic choice of X_1. Let $\mathfrak{sl}(B)_R$ denote the matrices with zero on the diagonal (the sum of the root spaces). Then

$$\begin{aligned} ad(X_1) : \mathfrak{sl}(B)_R &\to \mathfrak{sl}(B)_R, \\ Y &\mapsto [X, Y], \end{aligned}$$

is a linear isomorphism, and $ad(X_1)$ kills the diagonal matrices. Write $U_j = [X_1, X_j]$, so the U_j will be matrices with zero on the diagonal, and by picking T generically we can have any such matrices, and this is the most general choice of T possible, so if the equations vanish for a generic choice of U_j, they vanish identically.

Because of their indirect nature, it is difficult to write down these equations as polynomials.

Proposition 7.1. *Let $\dim A = \mathbf{a}$, $\dim B = \dim C = m$. Then B. Griesser's equations of [8] for $\hat{\sigma}_r$ have the following properties:*

1. *They are trivial for $r = 2m-1$ and all $\mathbf{a}$.*
2. *They are trivial for $r = 2m-2$, $\mathbf{a} = m$ and $m \le 4$.*
3. *Setting $m = \mathbf{n}^2$, matrix multiplication $M_{\langle \mathbf{n},\mathbf{n},\mathbf{n}\rangle}$ fails to satisfy the equations for $r \le \frac{3}{2}\mathbf{n}^2 - 1$ when $\mathbf{n}$ is even and $r \le \frac{3}{2}\mathbf{n}^2 + \frac{\mathbf{n}}{2} - 2$ when $\mathbf{n}$ is odd, and satisfies the equations for all larger r.*

I was unable to determine whether or not the equations are trivial for $r = 2m-2$, $\mathbf{a} = m$ and $m > 4$. If they are nontrivial for even m, they would give equations beyond the maximal minors of $T_A^{\wedge p}$.

Proof. Proof of (1): In the case $r = 2m-1$, so $r - m = m-1$ and $\mathbf{a} \le m+1$ the equations are trivial as we only have $\mathbf{a}-2 \le m-1$ linear maps. When $\mathbf{a} \ge m+2$ a naïve dimension count makes it possible for the equations to be non-trivial, the equations are that $\dim\langle U_2 v, \cdots, U_{\mathbf{a}-1} v\rangle \le m-1$. However, with our normalizations of $X_0 = Id$ and X_1 diagonal with distinct entries on the diagonal, taking $v = (1, 0, \cdots, 0)^T$ (the superscript T denotes transpose), the $U_j v$ will be contained in the hyperplane of vectors with their first entry zero. Since we only made genericity assumptions, we conclude.

Proof of (2): In the case $r = 2m-2$, the equations will be nontrivial if and only if there exist $U_2, \cdots, U_{\mathbf{a}-1}$ such that for all linearly independent v, w $\dim\langle U_2 v, \cdots, U_{\mathbf{a}-1} v, U_2 w, \cdots, U_{\mathbf{a}-1} w\rangle \ge m-1$. For $\mathbf{a} = m$, we saw we could have $U_2 v, \cdots, U_{m-1} v$ linearly independent, so the nontriviality condition is that for some j, $U_j w \notin \langle U_2 v, \cdots, U_{m-1} v\rangle$.

First observe that $U_j w \in \langle U_2 v, \cdots, U_{m-1} v\rangle \bmod U_j \hat{v}$ (where $\hat{v}$ is the line determined by v) means $w = \sum_{k\neq j} a_{j,k} U_j^{-1} U_k v$ for some constants $a_{j,k}$. (We are working with generic U_j so we may assume they are invertible.) Thus we must have v and constants $s_{i,j}, t_{i,j}$, such that $s_{i,j} \sum_{k\neq j} a_{j,k} U_j^{-1} U_k v =$

$t_{i,j}\sum_{l\neq i} a_{l,i}{U_l}^{-1}U_i v$, i.e., $U_2,\cdots,U_{m-1}$ must be such that there exist constants $s_{i,j},t_{i,j}$ for $i<j$, and $a_{j,k}$ for $j\neq k$ such that

$$\det(s_{i,j}\sum_{k\neq j} a_{j,k}{U_j}^{-1}U_k - t_{i,j}\sum_{l\neq i} a_{l,i}{U_l}^{-1}U_i)=0.$$

When $m=4$, the s,t are irrelevant and we need $a_{2,3}{U_2}^{-1}U_3v = a_{3,2}{U_3}^{-1}U_2v$, i.e., that for some choice of $[a_{2,3},a_{3,2}]\in\mathbb{P}^1$, the linear map $a_{2,3}{U_2}^{-1}U_3 - a_{3,2}{U_3}^{-1}U_2$ has a kernel. But every $\mathbb{P}^1$ of matrices intersects the hypersurface $\det_m=0$ so we conclude.

Remark 7.2. I expect the equations are non-trivial for $m\geq 5$ but I was unable to show this, even for $m=5$. The $r=2m-1$ case shows that one should be cautious. Consider the $m=5$ case. The equations would be trivial if for all $U_2,U_3,U_4\in\mathfrak{sl}(B)_R$, one could choose $([a_{2,3},a_{2,4}],[a_{2,3},a_{2,4}],[s_{2,3},t_{2,3}],[s_{2,4},t_{2,4}])\in\mathbb{P}^1\times\mathbb{P}^1\times\mathbb{P}^1\times\mathbb{P}^1$ such that the linear maps $s_{2,3}(a_{2,3}{U_2}^{-1}U_3 + a_{2,4}{U_2}^{-1}U_4) - t_{2,3}(a_{3,2}{U_3}^{-1}U_2 + a_{3,4}{U_3}^{-1}U_4)$ and $s_{2,4}(a_{2,3}{U_2}^{-1}U_3 + a_{2,4}{U_2}^{-1}U_4) - t_{2,4}((a_{4,2}{U_4}^{-1}U_2 + a_{4,3}{U_4}^{-1}U_3))$ have a common kernel. If we consider the variety $\Sigma_m\subset G(m-3,\mathbb{C}^{m^2})$ defined by

$$\Sigma_m := \{E\in G(m-3,\mathbb{C}^{m^2}) \mid \exists v\in V\backslash 0 \text{ such that } e.v=0\forall e\in E\},$$

then $\dim\Sigma_m=(m-1)+(m-3)(m^2-m-(m-3))$ (as for each point in $\mathbb{P}V$ there is an m^2-m dimensional space of endomorphisms with the line in the kernel, and we have the Grasssmannian of $m-3$ planes in that space of endomorphisms). So in the $m=5$ case a general four dimensional subvariety of the Grassmanian will fail to intersect Σ_5, but our four dimensional subvariety is not general.

Problem 7.3. Determine if Griesser's equations are non-trivial in this range.

Proof of (3): Consider matrix multiplication $M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}\in\mathbb{C}^{\mathbf{n}^2}\otimes\mathbb{C}^{\mathbf{n}^2}\otimes\mathbb{C}^{\mathbf{n}^2}=A\otimes B\otimes C$. With a judicious choice of bases, $M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle}(A)$ is block diagonal

$$\begin{pmatrix} x & & \\ & \ddots & \\ & & x\end{pmatrix} \tag{13}$$

where $x=(x^i_j)$ is $\mathbf{n}\times\mathbf{n}$. In particular, the image is closed under brackets. Choose X_0 so it is the identity. We may not have X_1 diagonal with distinct entries on the diagonal, the best we can do is for X_1 to be block diagonal with each block having the same n distinct entries. For a subspace E of dimension $2m-r=d\mathbf{n}+e$ (recall $m=\mathbf{n}^2$) with $0\leq e\leq n-1$, the image of a generic choice of $[X_1,X_2],\cdots,[X_1,X_{\mathbf{n}^2-1}]$ applied to E is of dimension at least $(d+1)\mathbf{n}$ if $e\geq 2$, at least $(d+1)n-1$ if $e=1$ and dn if $e=0$, and equality will hold if we choose E to be, e.g., the span of the first $2m-r$ basis vectors of B. (This is because the $[X_1,X_j]$ will span the entries of type (13) with zeros on the diagonal.) If n is even, taking $2m-r=\frac{n^2}{2}+1$, so $r=\frac{3\mathbf{n}^2}{2}-1$, the image occupies a space of dimension $\frac{\mathbf{n}^2}{2}+n-1>\frac{\mathbf{n}^2}{2}-1=r-m$. If one takes $2m-r=\frac{\mathbf{n}^2}{2}$, so $r=\frac{3\mathbf{n}^2}{2}$, the image occupies a space of dimension $\frac{\mathbf{n}^2}{2}=r-m$, showing Griesser's equations cannot do better for $\mathbf{n}$ even. If $\mathbf{n}$ is odd, taking $2m-r=\frac{\mathbf{n}^2}{2}-\frac{\mathbf{n}}{2}+2$, so $r=\frac{3\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}-2$, the image will have dimension $\frac{\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}>r-m=\frac{\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}-1$, and taking $2m-r=\frac{\mathbf{n}^2}{2}-\frac{\mathbf{n}}{2}+1$ the image can have dimension $\frac{n^2}{2}-\frac{n}{2}+(n-1)=r-m$, so the equations vanish for this and all larger r. Thus Griesser's equations for $\mathbf{n}$ odd give Lickteig's bound $\underline{\mathbf{R}}(M_{\langle\mathbf{n},\mathbf{n},\mathbf{n}\rangle})\geq\frac{3\mathbf{n}^2}{2}+\frac{\mathbf{n}}{2}-1$. $\square$

8. Equations via representation theory alone

8.1. **The brute force method.** One can simply write down the decomposition of $S^d(A{\otimes}B{\otimes}C)$ and test highest weight vectors on a general (i.e. sufficiently general) point of σ_r. This method is limited to small r and small dimension. Nevertheless, one can recover Strassen's degree 4 modules this way. Rather than studying all modules, one can combine this method with others, such as J. Hauenstein's method to determine the degrees and number of generators of ideals of varieties given parametrically.

The following modules were found in this way, the first three by a systematic search, and the last two in combination with Hauenstein's methods.

module	degree	in ideal of	where appeared
$S_{222}\mathbb{C}^3{\otimes}S_{222}\mathbb{C}^3{\otimes}S_{3111}\mathbb{C}^4$	6	$\sigma_4(Seg(\mathbb{P}^2\times\mathbb{P}^2\times\mathbb{P}^3))$	[13]
$S_{5111}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^4$	8	$\sigma_5(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$	[13]
$S_{3311}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^4$	8	$\sigma_5(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$	[13]
$S_{5554}\mathbb{C}^4{\otimes}S_{5554}\mathbb{C}^4{\otimes}S_{5554}\mathbb{C}^4$	19	$\sigma_6(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$	[10]
$S_{5555}\mathbb{C}^4{\otimes}S_{5555}\mathbb{C}^4{\otimes}S_{5555}\mathbb{C}^4$	20	$\sigma_6(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$	[10]

The degree 19 module was only shown to be in the ideal "with extremely high probability" as it was verified by choosing "random" points on the secant variety $\sigma_6(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$ and checking vanishing on them.

add modules found by Ikenmeyer, especially occurance obstruction hook*

The modules $S_{5111}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^4$ and $S_{3311}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^3{\otimes}S_{2222}\mathbb{C}^4$ were proven in the ideal of $\sigma_5(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$ by graph-theoretic methods. These methods were substantially generalized by Ikenmeyer in his PhD thesis, using hypergraphs and what he calls "obstruction designs", to show ******* are in the ideal.

9. Friedland's degree 16 equations

Secant varieties of Segre varieties appear in the study of algebraic statistical models corresponding to bifurcating phylogenetic trees, see [1], in particular $\sigma_4(Seg(\mathbb{P}^3\times\mathbb{P}^3\times\mathbb{P}^3))$ plays a central role. In 2008, equations for this variety were just beyond the state of the art, so E. Allmann, who resides in Alaska, offered to hand-catch, smoke and send an Alaskan salmon to anyone who could find the generators of the ideal of this variety. In [14] the problem was reduced to finding generators of the ideal for $\sigma_4(Seg(\mathbb{P}^2\times\mathbb{P}^2\times\mathbb{P}^3))$. The first major breakthrough to this conjecture was by S. Friedland [6]. The breakthrough had two essential steps: finding new equations and proving the known equations plus the new ones were sufficient to cut out the variety set-theoretically. I explain the new equations in this section. For a proof of the second step, see [6] or [**?**, §7.7.3].

Let $T\in A{\otimes}B{\otimes}C$ and assume $T_A^{\wedge 1}$ and $T_B^{\wedge 1}$ have nontrivial kernels. Let $\psi_{AB}\in\ker(T_A^{\wedge 1}: A{\otimes}B^*\to\Lambda^2A{\otimes}C)$ and $\psi_{BA}\in\ker(T_B^{\wedge 1}: B{\otimes}A^*\to\Lambda^2B{\otimes}C)$.

Proposition 9.1. *If* $\mathbf{a}=\mathbf{b}$ *and* $\psi_{AB}\in\ker(T_{AB}^{\wedge 1})$ *is of maximal rank, then* T *is equivalent to a tensor in* $S^2A{\otimes}C$ *and if furthermore* $\mathbf{c}=\mathbf{a}$ *and there exists* $\psi_{AC}\in\ker(T_{AC}^{\wedge 1})$ *of maximal rank, then* T *is equivalent to a tensor in* S^3A.

Proof. Consider $(Id_A{\otimes}\psi_{AB}{\otimes}Id_C)(T)\in A{\otimes}A{\otimes}C$. Since $\psi_{AB}\in\ker(T_A^{\wedge 1})$, we actually have $(Id_A{\otimes}\psi_{AB}{\otimes}Id_C)(T)\in S^2A{\otimes}C$, and since ψ_{AB} is injective, it is $GL(A)\times GL(B)\times GL(C)$-equivalent to T. The second assertion is similar. □

Proposition 9.2. *If* $T\in\mathbb{C}^4{\otimes}\mathbb{C}^4{\otimes}\mathbb{C}^4$ *satisfies Strassen's equations for border rank* 4 *(resp.* 5*), and* $T_{AB}^{\wedge 1}$ *and* $T_{AC}^{\wedge 1}$ *both contain elements of maximal rank, then* $\underline{\mathbf{R}}(T)\le 4$ *(resp.* $\underline{\mathbf{R}}(T)\le 5$*).*

Proof. We have $\sigma_4(v_3(\mathbb{P}^3)) = \mathbb{P}S^3\mathbb{C}^4$. □

If T satisfies the degree nine Strassen equations (i.e., $T_A^{\wedge 1}$ has a kernel) and $\mathrm{rank}(\psi_{AB}) = \mathrm{rank}(\psi_{BA}) = 3$, then $\psi_{AB}\psi_{BA} = \lambda Id$. To see this, use the normal form for a general point of $\hat{\sigma}_4$, namely, $T = a_1 \otimes b_1 \otimes c_1 + a_2 \otimes b_2 \otimes c_2 + a_3 \otimes b_3 \otimes c_3 + (a_1 + a_2 + a_3) \otimes (b_1 + b_2 + b_3) \otimes c_4$.

Thus if $T \in \hat{\sigma}_4$,

$$proj_{\mathfrak{sl}(A)}(\psi_{AB}\psi_{BA}) = 0, \ proj_{\mathfrak{sl}(B)}(\psi_{BA}\psi_{AB}) = 0. \tag{14}$$

These are equations of degree 16 for $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^3))$. I do not know what they are as a module.

****this paragraph needs fixing**** These can be generalized to equations for $\sigma_5(Seg(\mathbb{P}^5 \times \mathbb{P}^3 \times \mathbb{P}^3))$. If $T \in \sigma_5(Seg(\mathbb{P}^5 \times \mathbb{P}^3 \times \mathbb{P}^3))$ is a general point, we can write it in the normal form $T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_4 \otimes b_4 \otimes c_4 + (a_1 + a_2 + a_3 + a_4) \otimes (b_1 + b_2 + b_3 + b_4) \otimes (c_1 + c_2 + c_3 + c_4)$, where the a_j are a basis of A with dual basis α^j etc. Then $\psi_{AB} := a_1 \otimes \beta^1 + \cdots + a_4 \otimes \beta^4 \in \ker(T_{AB}^{\wedge 1} : A \otimes B^* \to \Lambda^2 A \otimes C)$ and $\psi_{BA} := b_1 \otimes \alpha^1 + \cdots + b_4 \otimes \alpha^4 \in \ker(T_{BA})$, and $[\psi_{AB}\psi_{BA}] = [Id_A]$. (This is the same proof as for the smaller dimensional case.)

10. Ideas for new equations

10.1. **The image of $T : B^* \to A \otimes C$.** The following is classical:

Theorem 10.1. *Let $T \in A \otimes B \otimes C$. Then $\mathbf{R}(T)$ equals the numberof rank one matrices needed to span a space containing $T(A^*) \subset B \otimes C$ (and similarly for permuted statements).*

Theorem 10.2. *Let $T \in A \otimes B \otimes C$, Then $\mathbf{R}(T)$ equals the number of rank one matrices needed to span (a space containing) $T(A^*) \subset B \otimes C$ (and similarly for the permuted statements).*

Proof. Let T have rank r so there is an expression $T = \sum_{i=1}^r a_i \otimes b_i \otimes c_i$. (I remind the reader that the vectors a_i need not be linearly independent, and similarly for the b_i and c_i.) Then $T(A^*) \subseteq \langle b_1 \otimes c_1, \cdots, b_r \otimes c_r \rangle$ shows that the number of rank one matrices needed to span $T(A^*) \subset B \otimes C$ is at most $\mathbf{R}(T)$.

On the other hand, say $T(A^*)$ is spanned by rank one elements $b_1 \otimes c_1, \cdots, b_r \otimes c_r$. Let $a^1, \cdots, a^{\mathbf{a}}$ be a basis of A^*, with dual basis $a_1, \cdots, a_{\mathbf{a}}$ of A. Then $T(a^i) = \sum_{s=1}^r x_s^i b_s \otimes c_s$ for some constants x_s^i. But then $T = \sum_{s,i} a_i \otimes (x_s^i b_s \otimes c_s) = \sum_{s=1}^r (\sum_i x_s^i a_i) \otimes b_s \otimes c_s$ proving $\mathbf{R}(T)$ is at most the number of rank one matrices needed to span $T(A^*) \subset B \otimes C$. □

Exercise 10.3: State and prove a border rank version of Theorem 10.2.

Thus if $\underline{\mathbf{R}}(T)$ is small, then $T(B^*)$ is a "non-generic" linear subspace of $A \otimes C$. To find new equations, one could fine ways of testing for this non-genericity. One could also study the non-genericity of the kernel of $T : A^* \otimes C^* \to B$, or the non-genericity of the images and kernels of the Koszul flattenings. What follows are a few ideas in this direction that have not yet borne fruit.

Assume $\mathbf{a} = \mathbf{c} = m$. The only $GL(A) \times GL(C)$-orbit closures inside $\mathbb{P}(A \otimes C)$ are the $\sigma_{m-q}(Seg(\mathbb{P}A \times \mathbb{P}C))$ which has codimension q^2. In particular, if $q^2 \leq \mathbf{b} - 1$, the intersection with $\sigma_q(Seg(\mathbb{P}A \times \mathbb{P}C))$ will be non-empty for any T, assuming, which we do, that $T : B^* \to A \otimes C$ is injective.

If $T = \sum_{j=1}^r a_j \otimes b_j \otimes c_j$ and we take $\beta \in b_1{}^{\perp} \cap \cdots \cap b_{\mathbf{b}-1}{}^{\perp}$, then $T(\beta) \in \sigma_{r-\mathbf{b}+1}(Seg(\mathbb{P}A \times \mathbb{P}C))$.

10.1.1. *Darboux-Luroth type equations.* Giorgio and I discussed these around 2011. For simplicity assume $\mathbf{b} = m$. Say $\mathbf{R}(T) = 2m - 1$. Then there exists a configuration of $2m - 1$ hyperplanes such $\mathbb{P}T(B^*) \cap \{\det = 0\}$ contains the $\binom{2m-1}{m-1}$ points of intersection of subsets of $m - 1$ of them. The existence of such for sufficiently large m is easily seen to be a non-generic condition (and Lüroth even showed that when $m = 3$ and it naïvely looks generic, it is still non-generic). The difficulty here is that we do not know how to test for the condition. One gets similar conditions for all $m+1 \leq r \leq 2m-1$. In particular, I don't know how to check if matrix multiplication is or is not in the zero set of these equations.

10.1.2. *Equations via intersections with orbit closures.* We saw above that if $T \in \sigma_r$, then $\mathbb{P}T(B^*) \cap \sigma_{r-\mathbf{b}-1}(Seg(\mathbb{P}A \times \mathbb{P}C)) \neq \emptyset$. This is a non-trivial condition if $(m+\mathbf{b}-r-1)^2 > m-1$, i.e., if $r < m+\mathbf{b}-\sqrt{m-1}-1$. For example, when $\mathbf{b} = m = 4$, we get nontrivial equations if $r < 7-\sqrt{3}$ i.e., if $r \leq 5$.

Note also that when $m = n^2$ and $T = M_{\langle n \rangle}$, we have $\mathbb{P}T(B^*) \cap \sigma_n(Seg(\mathbb{P}A \times \mathbb{P}C)) \neq \emptyset$, so matrix multiplication tends to satisfy these equations.

In summary:

Proposition 10.4. *Let $\mathbf{a} = \mathbf{c} = m$. There are non-trivial equations for $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ for $r < m+\mathbf{b}-\sqrt{m-1}-1$ obtained by requiring that $\mathbb{P}T(B^*) \cap \sigma_{r-\mathbf{b}-1}(Seg(\mathbb{P}A \times \mathbb{P}C)) \neq \emptyset$.*

When $m = n^2$ and $T = M_{\langle n \rangle}$ is matrix multiplication, these equations are satisfied for all $r \geq n^2+n+1$, despite the fact that $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq 2n^2-n$.

10.1.3. *Combining the two above.*

Theorem 10.5. *Let $\mathbf{b} \leq \mathbf{a} = \mathbf{c} = m$ and let $r \leq 2m-1$. Then if $T \in \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)_{general}$, then there exists a configuration of r hyperplanes $H_1, \cdots, H_r \subset \mathbb{P}B$, such that for all $q \leq m-1+\mathbf{b}-r$, the configuration of $\mathbb{P}^{q-1}$'s obtained by intersecting subsets of $\mathbf{b}-q$ of the hyperplanes in the configuration is contained in $\sigma_{r-\mathbf{b}+q}(Seg(\mathbb{P}A \times \mathbb{P}C)) \cap \mathbb{P}T(B^*)$. In general, $T \in \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ lies in the variety defined by taking the closure of such configurations.*

In particular, for every point of $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$, there exist $\mathbb{P}^{q-1}$'s in $\mathbb{P}T(B^*) \cap \sigma_{r-\mathbf{b}+q}(Seg(\mathbb{P}A \times \mathbb{P}C))$.

Remark 10.6. For all these equations, the way matrix multiplication satisfies them is very different from the way a general element of σ_r satisfies them. For example, when considering the kernel of $T_A^{\wedge p} : \Lambda^p A \otimes B^* \to \Lambda^{p+1} A \otimes C$, I expect the kernel to be a general linear subspace of $\Lambda^p A \otimes B^*$, but for matrix multiplication, it is spanned by elements of $Seg(G(p,A) \times \mathbb{P}B^*)$. The hope is to do a Friedland type trick to gain additional equations. In order to do this, we'll need a better understanding of the "shape" of a general element of the kernel.

10.2. **Equations via the kernel of $(T_A^{\wedge p})^T : \Lambda^{p+1} A^* \otimes C^* \to \Lambda^p A^* \otimes B$.** The idea here is to get new equations by requiring $\mathbb{P}\ker(T_A^{\wedge p})^T \cap Seg(G(p+1,A^*) \times \mathbb{C}^*) \neq \emptyset$. Let $A, B, C \simeq \mathbb{C}^m$ and let $e_1, \cdots, e_m$, $f_1, \cdots, f_m$, $g_1, \cdots, g_m$ be bases of A, B, C with dual bases e^s, f^s, g^s. Write $T = T^{stu} e_s \otimes f_t \otimes g_u$ (summation convention in force!). Write $\hat{\alpha}^\sigma = (-1)^{\sigma-1} \alpha^1 \wedge \cdots \wedge \alpha^{\sigma-1} \wedge \alpha^{\sigma+1} \wedge \cdots \wedge \alpha^{p+1}$. We have

$$(T_A^{\wedge p})^T (\alpha^1 \wedge \cdots \wedge \alpha^{p+1} \otimes \gamma) = T^{stu} \gamma(g_u) (\sum_{\sigma=1}^{p+1} \alpha^\sigma(e_s) \hat{\alpha}^\sigma) \otimes f_t \tag{15}$$

If we want (15) to be zero, for each t we must have $T^{stu} \gamma(g_u) (\sum_{\sigma=1}^{p+1} \alpha^\sigma(e_s) \hat{\alpha}^\sigma) = 0$. Since the vectors $\hat{\alpha}^\sigma$ are all linearly independent, we see (15) can be zero if and only if the system of equations

$$\begin{pmatrix} T^{11u}\gamma(g_u) & \cdots & T^{n1u}\gamma(g_u) \\ & \vdots & \\ T^{1nu}\gamma(g_u) & \cdots & T^{nnu}\gamma(g_u) \end{pmatrix}$$

has a $(p+1)$-dimensional space of solutions for some choice of γ. (In this case, if the solutions are z_σ, we take

$$\begin{pmatrix}\alpha^\sigma(e_1)\\ \vdots \\ \alpha^\sigma(e_m)\end{pmatrix} = z_\sigma$$

But this is just the condition of §10.1.2, with the roles of B and C exchanged!

If we look at more complicated elements in the kernel, we recover other parts of the configuration discussed above, e.g., general points of $\tau(Seg(G(p+1,A^*)\times\mathbb{P}C^*))$ are of the form $\alpha^1\wedge\alpha^2\otimes\gamma_3+\alpha^1\wedge\alpha^3\otimes\gamma_2+\alpha^2\wedge\alpha^3\otimes\gamma_1$, and to have one of these in the kernel reproduces the configuration of 3 $\mathbb{P}^1$'s in $\sigma_{m-2}(Seg(\mathbb{P}A\times\mathbb{P}B))$, each of which contains two points in $\sigma_{m-3}(Seg(\mathbb{P}A\times\mathbb{P}B))$, and these points are the points of intersection of the 3 $\mathbb{P}^1$'s (all of which lie in a plane).

This should not come as a surprise, as after all, $T(B^*)$ contains all the information of T (modulo the action of $GL(B)$ which we don't care about). However looking at things from different perspectives will give rise to different insights - at the end of the day we might want to translate back to $T(B^*)$ to facilitate comparisions of potential methods.

We can see from this perspective that matrix multiplication does satisfy these equations: fix $\mu\in U^*$, $u\in\mu^\perp\subset U$, $v_1,v_2\in V$, and $\omega\in W^*$, then

$$(\mu\otimes v_1)\wedge(\mu\otimes v_2)\otimes(\omega\otimes u)\mapsto 0.$$

Thus we have a copy of $Flag_{1,n-1}(U)\times G(2,V)\times\mathbb{P}W^*$ mapping to zero, which is very pathological. Even for larger p there are elements of the Segre mapping to zero. Take the other extreme, $p=\frac{n-1}{2}$ (assume for simplicity that $n=2q+1$ is odd), then taking $u_1,\cdots,u_q\in U$ independent, $\nu^1,\cdots,\nu^n$ a basis of V and $\mu\in\{u^1,\cdots,u^q\}^\perp\subset U^*$ and $\omega\in W^*$, we see

$$[(u_1\otimes\nu^1)\wedge\cdots\wedge(u_1\otimes\nu^n)\wedge(u_2\otimes\nu^1)\wedge\cdots\wedge(u_2\otimes\nu^n)\wedge\cdots(u_q\otimes\nu^1)\wedge\cdots\wedge(u_q\otimes\nu^n)]\otimes(\omega\otimes u)\mapsto 0.$$

So there is at least a $Flag_{q,n-1}(U)\times\mathbb{P}W$'s worth of points on the Segre mapping to zero.

10.3. **Duality.** In fact, we do not need to consider $(T_A^{\wedge p})^T$. Write $T_{AB}^{\wedge p}:\Lambda^pA\otimes B^*\to\Lambda^{p+1}A\otimes C$ to be more precise.

Proposition 10.7. *If we choose a volume element for A, then*

$$T_{AB}^{\wedge p}=(T_{AC}^{\wedge\mathbf{a}-p-1})^T.$$

Proof. We have $(T_{AC}^{\wedge\mathbf{a}-p-1})^T:\Lambda^{\mathbf{a}-p}A^*\otimes B^*\to\Lambda^{\mathbf{a}-p-1}A^*\otimes C$, but using the volume form, we may identify $\Lambda^{\mathbf{a}-p}A^*\simeq\Lambda^pA$ and $\Lambda^{\mathbf{a}-p-1}A^*\simeq\Lambda^{p+1}A$. To see that the maps are the same up to scale one can argue via Schur's lemma applied to the module maps $A\otimes B\otimes C\times\Lambda^{\mathbf{a}-p}A^*\otimes B^*\to\Lambda^{\mathbf{a}-p-1}A^*\otimes C$, or write $T=\sum_j a_j\otimes b_j\otimes c_j$ and compute with basis vectors. □

10.4. **Equations via $T_A^{\wedge p}(G(p+1,A)\times\mathbb{P}B^*)$ being in special position.** Let

$$\begin{aligned}\Sigma_{p,q}:=&\overline{\{[E_1\otimes c_1+\cdots E_q\otimes c_q]\mid c_u\in C,\ E_u\in\hat{G}(p+1,A),\exists F\in G(p,A)\text{ such that }F\subset E_u\ \forall u\}}\\ &\subset\sigma_q(Seg(G(p+1,A)\times\mathbb{P}C))).\end{aligned}$$

Then for all tensors T, $(T_A^{\wedge p})(G(p,A)\times\mathbb{P}C)\subset\Sigma_{p,\mathbf{c}}$.

Note

$$\begin{aligned}\dim\Sigma_{p,q}&=\dim G(p,A)+\dim\sigma_q(Seg(\mathbb{P}(A/F)\times\mathbb{P}C))\\ &=p(\mathbf{a}-p)+q(\mathbf{a}-p+m-q)-1.\end{aligned}$$

Let's specialize to $p = 1$ and $\mathbf{a} = \mathbf{b} = \mathbf{c}$ for the moment. If $\mathbf{R}(T) = r$, then $\mathbb{P}T_A^{\wedge 1}(A \otimes B^*) \cap \Sigma_{1,r-m} \neq \emptyset$. To test if we actually have an intersection, we need equations for $\Sigma_{1,r-m}$. By Weyman's method (thanks to help from Jerzy), the ideal of $\Sigma_{1,m}$ is generated in degrees 2 and 3 respectively by the modules $\Lambda^4 A^* \otimes S^4 B$ and $S_{222}A^* \otimes \Lambda^3 B$. Then the ideal of $\Sigma_{1,q}$ is generated by these equations plus $\Lambda^{q+1}(\Lambda^2 A^*) \otimes \Lambda^{q+1} B$ (possibly with redundancy).

The situation is nontrivial only if $r-m < m-1$, so at best we can get equations up to $r = 2m-2$.

Now consider the general case. If $\mathbf{R}(T) = r$, then $\mathbb{P}T_A^{\wedge p}(G(p,A) \otimes B^*) \cap \Sigma_{p,r-m-p+1} \neq \emptyset$. To test if we actually have an intersection, we would need equations for $\Sigma_{p,r-m-p+1}$. The situation is nontrivial only if $r - m - p + 1 < m - p$, so at best we can get equations up to $r = 2m - 2$ in general. So if we do examine these equations further, we should just stick to the $p = 1$ case (at least initially).

Note that matrix multiplication is extremely degenerate for these equations as well: take $v_1, \cdots, v_q \in V$ and $\nu \in \{v_1, \cdots, v_q\}^\perp$, and $w \in W$. Then

$$[(\mu^1 \otimes v_1) \wedge (\mu^2 \otimes v_1) \wedge \cdots (\mu^n \otimes v_1) \wedge \cdots \wedge (\mu^n \otimes v_q)] \otimes (\nu \otimes w) \mapsto 0.$$

So we have at least a $Flag(q, n-1)(V) \times \mathbb{P}W$ in the Segre that maps to zero.

If we do not restrict to $G(p,A) \times \mathbb{P}B^*$, then we can look for degenerate images subject to less restrictions. Still need to work out the numbers.

11. Barriers

We seem to have the $2m$ barrier for even conjectural equations for border rank. Similarly, the computer scientists (especially Wigderson) see a big barrier to prove lower bounds given by the "input size" which in our case is $3m$. (In fact they were initially suprised by the Mignon-Ressayre result because it appeared to cross the input size, until they "realized" that the information of the determinant hypersurface could be stored in its dual variety, which has dimension $2n-2$, which resolved their apparent paradox.)

References

1. Elizabeth S. Allman and John A. Rhodes, *Phylogenetic ideals and varieties for the general Markov model*, Adv. in Appl. Math. **40** (2008), no. 2, 127–148. MR 2388607 (2008m:60145)
2. 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
3. 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
4. Luca Chiantini, Jon Hauenstein, Christian Ikenmeyer, J.M. Landsberg, and Giorgio Ottaviani, *Towards algorithms for matrix multiplication*, in preparation.
5. P. Comon and G. Ottaviani, *On the typical rank of real binary forms*, arXiv:0909.4865, preprint.
6. Shmuel Friedland, *On tensors of border rank l in* $\mathbb{C}^{m\times n\times l}$, Linear Algebra Appl. **438** (2013), no. 2, 713–737. MR 2996364
7. Franois Le Gall, *Powers of tensors and fast matrix multiplication*, arXiv:1401.7714.
8. B. Griesser, *A lower bound for the border rank of a bilinear map*, Calcolo **23** (1986), no. 2, 105–114 (1987). MR 88g:15021
9. Wolfgang Hackbusch, *Tensor spaces and numerical tensor calculus*, Springer series in computational mathematics, vol. 42, Springer, Heidelberg, 2012.
10. 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
11. J. M. Landsberg, *Explicit tensors of border rank at least* $2n-1$, arXiv:1209.1664.
12. ———, *Tensors: geometry and applications*, Graduate Studies in Mathematics, vol. 128, American Mathematical Society, Providence, RI, 2012. MR 2865915

13. J. M. Landsberg and L. Manivel, *On the ideals of secant varieties of Segre varieties*, Found. Comput. Math. **4** (2004), no. 4, 397–422. MR 2097214 (2005m:14101)
14. J. M. Landsberg and Laurent Manivel, *Generalizations of Strassen's equations for secant varieties of Segre varieties*, Comm. Algebra **36** (2008), no. 2, 405–422. MR MR2387532
15. J.M. Landsberg and Giorgio Ottaviani, *New lower bounds for the border rank of matrix multiplication*, preprint, arXiv:1112.6007.
16. Joseph M. Landsberg, Yang Qi, and Ke Ye, *On the geometry of tensor network states*, Quantum Inf. Comput. **12** (2012), no. 3-4, 346–354. MR 2933533
17. Thomas Lickteig, *A note on border rank*, Inform. Process. Lett. **18** (1984), no. 3, 173–178. MR 86c:68040
18. ———, *Typical tensorial rank*, Linear Algebra Appl. **69** (1985), 95–120. MR 87f:15017
19. David Mumford, *Algebraic geometry. I*, Classics in Mathematics, Springer-Verlag, Berlin, 1995, Complex projective varieties, Reprint of the 1976 edition. MR 1344216 (96d:14001)
20. A. Stothers, *On the complexity of matrix multiplication*, PhD thesis, University of Edinburgh, 2010.
21. V. Strassen, *Rank and optimal computation of generic tensors*, Linear Algebra Appl. **52/53** (1983), 645–685. MR 85b:15039
22. Volker Strassen, *Gaussian elimination is not optimal*, Numer. Math. **13** (1969), 354–356. MR 40 #2223
23. 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.
24. Virginia Williams, *Breaking the coppersimith-winograd barrier*, preprint.

E-mail address: `jml@math.tamu.edu`