GEOMETRY AND THE COMPLEXITY OF MATRIX MULTIPLICATION

J. M. LANDSBERG

ABSTRACT. We survey results in algebraic complexity theory, focusing on matrix multiplication. Our goals are (i) to show how open questions in algebraic complexity theory are naturally posed as questions in geometry and representation theory, (ii) to motivate researchers to work on these questions, and (iii) to point out relations with more general problems in geometry. The key geometric objects for our study are the secant varieties of Segre varieties. We explain how these varieties are also useful for algebraic statistics, the study of phylogenetic invariants, and quantum computing.

1. INTRODUCTION

1.1. **Strassen's algorithm.** Let A and B be 2×2 matrices

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}, \quad B = \begin{pmatrix} b_1^1 & b_2^1 \\ b_1^2 & b_2^2 \end{pmatrix}.$$

Recall the usual algorithm to calculate the matrix product $C = AB$:

(1.1.1)
$$\begin{aligned} c_1^1 &= a_1^1 b_1^1 + a_2^1 b_1^2, \\ c_2^1 &= a_1^1 b_2^1 + a_2^1 b_2^2, \\ c_1^2 &= a_1^2 b_1^1 + a_2^2 b_1^2, \\ c_2^2 &= a_1^2 b_2^1 + a_2^2 b_2^2. \end{aligned}$$

This algorithm uses eight multiplications and for $n \times n$ matrices it uses n^3.
Question: Is there a "better" algorithm for multiplying matrices? By "better" one could mean an algorithm that uses fewer arithmetic operations $(+, -, *)$ or simply fewer multiplications. The number of multiplications needed governs the total number of arithmetic operations in such a way that asymptotic results depend primarily on the number of multiplications used. (See Definition 1.2.1 for a precise statement.) In this article we focus exclusively on minimizing multiplications. (In actual implementations memory cost is also an important factor.)

In 1969 V. Strassen [54] made the following discovery. Set

$$I = (a_1^1 + a_2^2)(b_1^1 + b_2^2),$$
$$II = (a_1^2 + a_2^2)b_1^1,$$
$$III = a_1^1(b_2^1 - b_2^2),$$
(1.1.2)
$$IV = a_2^2(-b_1^1 + b_1^2),$$
$$V = (a_1^1 + a_2^1)b_2^2,$$
$$VI = (-a_1^1 + a_1^2)(b_1^1 + b_2^1),$$
$$VII = (a_2^1 - a_2^2)(b_1^2 + b_2^2).$$

Now check for yourself that if $C = AB$, then

$$c_1^1 = I + IV - V + VII,$$
$$c_1^2 = II + IV,$$
(1.1.3)
$$c_2^1 = III + V,$$
$$c_2^2 = I + III - II + VI.$$

Thus (1.1.2), (1.1.3) provides an algorithm for multiplying 2×2 matrices that uses only seven multiplications.

Remark 1.1.1. Strassen was attempting to prove, by process of elimination, that such an algorithm did not exist when he arrived at it. We will see in §3 why the result could have been anticipated using elementary algebraic geometry.

1.2. **The exponent of matrix multiplication.** In Strassen's algorithm the entries of the matrices need not be scalars - they could be elements of an algebra. Let A, B be 4×4 matrices, and write

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}, \quad B = \begin{pmatrix} b_1^1 & b_2^1 \\ b_1^2 & b_2^2 \end{pmatrix},$$

where a_j^i, b_j^i are 2×2 matrices. We may apply Strassen's algorithm to get the blocks of $C = AB$ in terms of the blocks of A, B performing seven multiplications of 2×2 matrices. Since we can apply Strassen's algorithm to each block, we can multiply 4×4 matrices using $7^2 = 49$ multiplications instead of the usual $4^3 = 64$. In fact, if A, B are $2^k \times 2^k$ matrices, we may multiply them using 7^k multiplications rather than the usual $(2^k)^3$. Even if n is not a power of 2, we can still save multiplications asymptotically by enlarging the dimensions of our matrices, placing zeros in the new entries, to obtain matrices whose size is a power of 2. Asymptotically we can multiply $n \times n$ matrices using $O(n^{log_2(7)}) \simeq O(n^{2.81})$ operations, as if we let $n = 2^k$ and write $7^k = (2^k)^a$ so $k(log_2 7) = ak(log_2 2)$ and we obtain $a = log_2 7$.

Definition 1.2.1. The *exponent* ω of matrix multiplication is

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

Strassen's algorithm shows $\omega \le log_2(7) < 2.81$.

Remark 1.2.2. If one replaces the phrase "scalar multiplications" with the phrase "arithmetic operations" in the definition, ω is unchanged; see [14], Proposition 15.1.

1.3. **Matrix multiplication as a bilinear map.** Matrix multiplication of square matrices is a bilinear map that we denote

$$(1.3.1) \qquad\qquad M_{n,n,n} : \mathbb{C}^{n^2} \times \mathbb{C}^{n^2} \to \mathbb{C}^{n^2}.$$

(In this article we restrict our attention to the complex numbers, so, e.g., all vector spaces are finite dimensional vector spaces over $\mathbb{C}$.) When discussing a minimal number of arithmetic operations (or multiplications) for executing a bilinear map, it is usually within the context of a class of algorithms. A natural class of algorithms for executing a bilinear map is as follows: let A, B, C be vector spaces, let $A^* := \{f : A \to \mathbb{C} \mid f \text{ is linear}\}$ denote the dual vector space (and similarly for B), and let $T : A \times B \to C$ be a bilinear map. Choose $\alpha^i \in A^*$, $\beta^i \in B^*$, $c_i \in C$ such that $T(v, w) = \sum_{i=1}^{r} \alpha^i(v)\beta^i(w)c_i$. The minimal number r over all such presentations of T is called the *rank* of T and is denoted $\mathbf{R}(T)$. A related notion, more natural to geometry and defined in §2, is that of *border rank*, denoted $\underline{\mathbf{R}}(T)$. Another concept that comes into play when discussing the space of all bilinear maps $A \times B \to C$ is the *typical rank*, which is the rank of a generic bilinear map $A \times B \to C$.

Thus $\mathbf{R}(M_{n,n,n})$ and $\underline{\mathbf{R}}(M_{n,n,n})$ measure the number of multiplications that are needed in order to compute the product of two $n \times n$ matrices. Strassen's algorithm shows that the rank of the multiplication of 2×2 matrices is at most seven, and S. Winograd [58] proved that in fact it equals seven.

1.4. **Overview.** To examine the complexity of matrix multiplication more geometrically, we first, in §2, rephrase it using tensors. Next, in §3, we introduce algebraic varieties which stratify the space of tensors, the *secant varieties of Segre varieties*. (The above-mentioned border rank of a tensor describes its location with respect to this stratification.) This is done in two steps: first introducing secant varieties to any algebraic variety in §3.3; then specializing to Segre varieties in §3.5. We also rephrase the main open problems in the complexity of matrix multiplication in terms of secant varieties of Segre varieties. In §3.8 we summarize the known results.

Before discussing those results in detail, we take two detours. In the first, we describe two problems from algebraic geometry where secant varieties arise: the *polynomial Waring problem* and *Hartshorne's conjecture on linear normality*. These are described in §4. In the second, we describe other applications of secant varieties of Segre varieties - to *algebraic statistics* (especially the study of *phylogenetic invariants*) and *quantum computing*, which is done in §5. These detours will allow the reader to place the topics discussed in the remainder of the paper in a larger mathematical context.

In §6 we describe *Strassen's equations* for secant varieties of Segre varieties and their use in establishing lower bounds for rank and border rank. In particular, we present a new proof of *Bläser's $\frac{5}{2}$-Theorem*. We rephrase Strassen's equations invariantly in §10 and describe generalizations.

While it is well known that the limit of a family of secant lines is a tangent line (or a secant line itself), exactly what can be in the limit of a secant k-plane is *not* known. We discuss what is known about this problem in §7 and show how to use this knowledge to establish upper bounds for the complexity of matrix multiplication in §8.1. (We explain how to use such limits to establish lower bounds in the discussion below Theorem 3.8.6.) A group-theoretic approach to upper bounds is described briefly in §8.2.

We discuss dimensions of secant varieties of Segre varieties in §9, focusing on the use of *Terracini's Lemma*.

Any proper study of varieties invariant under a group action, e.g., the secant varieties of Segre varieties, should exploit *representation theory*. The representation theory relevant to this study is discussed in §11. Representation theory is the most important tool discussed in this article.

A common technique in geometry is to understand a complicated geometric object via the construction of auxiliary objects that are more tractable, and the problem at hand is no exception. We describe two such objects in §12.

In §13, we describe a collection of techniques developed by J. Weyman for the study of G-varieties and their application to secant varieties of Segre varieties. (A G-variety is a variety invariant under the action of an algebraic group G.) These techniques help one to find the entire minimal free resolution of the ideal of a variety and describe the nature of its singularities.

Finally, in an appendix §14, we give nontraditional and more invariant presentations of two standard notions in complexity theory - *multiplicative complexity* and *separations*.

2. TENSOR FORMULATION

Let V, V_j, be vector spaces. The *dual space* V^* and the *tensor product* $V_1 \otimes \cdots \otimes V_n$ are defined as follows:

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

$$V_1 \otimes \cdots \otimes V_n := \{f : V_1^* \times \cdots \times V_n^* \to \mathbb{C} \mid f \text{ is linear in each factor}\}.$$

Given $v_j \in V_j$, $\alpha_j \in V_j^*$, define $v_1 \otimes \cdots \otimes v_n \in V_1 \otimes \cdots \otimes V_n$ by

$$v_1 \otimes \cdots \otimes v_n(\alpha_1, \ldots, \alpha_n) = \alpha_1(v_1) \cdots \alpha_n(v_n).$$

An element $f \in V_1 \otimes V_2$, that is, a bilinear map $f : V_1^* \times V_2^* \to \mathbb{R}$, may also be considered as a linear map

$$f : V_1^* \to V_2$$
$$\alpha \mapsto f(\alpha, \cdot)$$

where $f(\alpha, \cdot) \in (V_2^*)^* = V_2$; that is, for $\beta \in V_2^*$, $f(\alpha, \cdot)(\beta) = f(\alpha, \beta)$.

Definition 2.0.1. Let $V_1, \ldots, V_k$ be vector spaces. An element $z \in V_1 \otimes \cdots \otimes V_k$ is called *decomposable* if there exist $v_i \in V_i$ such that $z = v_1 \otimes \cdots \otimes v_k$. Define the *rank* of an element $T \in V_1 \otimes V_2 \otimes \ldots \otimes V_k$ to be the minimal number r such that $T = \sum_{u=1}^{r} z_u$ with each z_u decomposable. We refer to an explicit expression for a tensor T as a sum of r monomials as a *computation of T of length r*. This terminology is consistent with the definition of the rank of a linear map $T : V_1^* \to V_2$ (that is, an element $T \in V_1 \otimes V_2$) and the rank of a bilinear map $T : V_1^* \times V_2^* \to V_3$ given in §1.2 (that is, an element of $T \in V_1 \otimes V_2 \otimes V_3 = A^* \otimes B^* \otimes C$). Note that the length of a computation of a tensor is unchanged if we make changes of bases in the vector spaces V_i.

2.1. Tensor expressions for the standard algorithm and Strassen's algorithm. The standard algorithm (1.1.1) for the multiplication of 2×2 matrices $M_{2,2,2}$ (1.3.1) in terms of tensors is as follows: let A, B, C each denote the space of 2×2 matrices (that is, $A, B, C \simeq \mathbb{C}^4$); give A the standard basis a_j^i for the matrix

with a 1 in the (i, j)-th slot and zeros elsewhere, and let α^i_j denote the corresponding elements of the dual basis of A^*, and similarly for B, C. Then the standard algorithm for $M_{2,2,2} : \mathbb{C}^{4*} \times \mathbb{C}^{4*} \to \mathbb{C}^4$ is:

$$(2.1.1) \quad \begin{aligned} M_{2,2,2} = {} & \alpha^1_1 {\otimes} \beta^1_1 {\otimes} c^1_1 + \alpha^1_2 {\otimes} \beta^2_1 {\otimes} c^1_1 + \alpha^2_1 {\otimes} \beta^1_1 {\otimes} c^2_1 + \alpha^2_2 {\otimes} \beta^2_1 {\otimes} \ldots {\otimes} c^2_1 \\ & + \alpha^1_1 {\otimes} \beta^1_2 {\otimes} c^1_2 + \alpha^1_2 {\otimes} \beta^2_2 {\otimes} c^1_2 + \alpha^2_1 {\otimes} \beta^1_2 {\otimes} c^2_2 + \alpha^2_2 {\otimes} \beta^2_2 {\otimes} \ldots {\otimes} c^2_2 \end{aligned}$$

and Strassen's algorithm (1.1.2), (1.1.3) is

$$(2.1.2) \quad \begin{aligned} M_{2,2,2} = {} & (\alpha^1_1 + \alpha^2_2) {\otimes} (\beta^1_1 + \beta^2_2) {\otimes} (c^1_1 + c^2_2) + (\alpha^1_2 + \alpha^2_2) {\otimes} \beta^1_1 {\otimes} (c^2_1 - c^2_2) \\ & + \alpha^1_1 {\otimes} (\beta^1_2 - \beta^2_2) {\otimes} (c^1_2 + c^2_2) + \alpha^2_2 {\otimes} (-\beta^1_1 + \beta^2_1) {\otimes} (c^2_1 + c^1_1) \\ & + (\alpha^1_1 + \alpha^1_2) {\otimes} \beta^2_2 {\otimes} (-c^1_1 + c^1_2) + (-\alpha^1_1 + \alpha^2_1) {\otimes} (\beta^1_1 + \beta^1_2) {\otimes} c^2_2 \\ & + (\alpha^1_2 - \alpha^2_2) {\otimes} (\beta^2_1 + \beta^2_2) {\otimes} c^1_1. \end{aligned}$$

2.2. Approximate algorithms.
An *approximate algorithm* for a tensor T is a sequence of algorithms, usually of lower rank tensors, that converges to an algorithm for T. The *border rank* of a tensor T is the lowest rank of tensors in such sequences and is denoted $\underline{\mathbf{R}}(T)$. Note that rank and border rank can indeed be different - consider the following example:

$$(2.2.1) \quad T = a_1 {\otimes} b_1 {\otimes} c_1 + a_1 {\otimes} b_1 {\otimes} c_2 + a_1 {\otimes} b_2 {\otimes} c_1 + a_2 {\otimes} b_1 {\otimes} c_1.$$

One can show that $\mathbf{R}(T) = 3$, but we can approximate T as closely as we like by tensors of rank two as follows. Let

$$(2.2.2) \quad T(\epsilon) = \frac{1}{\epsilon} [(\epsilon - 1) a_1 {\otimes} b_1 {\otimes} c_1 + (a_1 + \epsilon a_2) {\otimes} (b_1 + \epsilon b_2) {\otimes} (c_1 + \epsilon c_2)]$$

and allow $\epsilon \to 0$, so $\underline{\mathbf{R}}(T) \le 2$ (in fact equality holds). The geometry of this limit is discussed in §3.5.

3. Geometric formulation

3.1. Varieties.
Let V be a vector space and let $\mathbb{P}V$ be the associated projective space of lines through the origin in V, so we have a map $\pi : V \backslash 0 \to \mathbb{P}V$. We use the following notation: for $v \in V \backslash 0$, let $[v] = \pi(v) \in \mathbb{P}V$ and for $Z \subset \mathbb{P}V$, let $\hat{Z} = \pi^{-1}(Z) \subset V$. For scale invariant sets $U \subset V \backslash 0$, write $\mathbb{P}U$ for $\pi(U)$. We use projective space in addition to vector spaces because the properties we are interested in (rank, border rank) are scale invariant. Because we go back and forth between vector and projective spaces, many objects end up being decorated with hats and "$\mathbb{P}$"s.

For our purposes, a *variety* $X \subset \mathbb{P}V$ is the common zero locus in $\mathbb{P}V$ of a collection of homogeneous polynomials on V. Given a variety X, we will construct a sequence of auxiliary varieties $X \subset \sigma_2(X) \subset \cdots \subset \sigma_f(X) = \mathbb{P}V$, called the *secant varieties of* X, which determine a stratification of $\mathbb{P}V$. This stratification will generalize the stratification of the space of $m \times n$ matrices by rank. When $V = A_1 {\otimes} \cdots {\otimes} A_n$ and X is the projectivization of the set of decomposable tensors, the stratification will coincide with the stratification of tensors by their border rank, and f is the *typical rank* mentioned in §1 and defined below.

3.2. **Joins.** For readers not accustomed to secant varieties, we begin with several special cases to help visualize them. Projective space $\mathbb{P}V$ has the property that, given any two distinct points $p, q \in \mathbb{P}V$, there is a unique *line*, that is, a linearly embedded $\mathbb{P}^1 \subset \mathbb{P}V$, containing p and q, which we denote $\mathbb{P}^1_{p,q}$. Let $C \subset \mathbb{P}V$ be a smooth curve (one-dimensional variety) and $q \in \mathbb{P}V$ a point. Let $J(q, C) \subset \mathbb{P}V$ denote the *cone* over C with vertex q, which by definition contains the union of all points on all lines containing q and a point of C. More precisely, $J(q, C)$ denotes the closure of the set of such points. It is necessary to take the closure only when $q \in C$, as in this case one also includes the points on the tangent line to C at q, because, as anyone who has ever taught calculus knows, the tangent line is the limit of secant lines $\mathbb{P}^1_{q,x_j}$ as $x_j \to q$. Define $J(q, Z)$ similarly for $Z \subset \mathbb{P}V$, a variety of any dimension. Unless Z is a linear space and $q \in Z$, $\dim J(q, Z) = \dim Z + 1$.

The *join* of $Y, Z \subset \mathbb{P}V$ is

$$J(Y, Z) = \overline{\bigcup_{x \in Y, y \in Z, x \neq y} \mathbb{P}^1_{xy}}.$$

Here the overline denotes Zariski closure; that is, if $U \subset \mathbb{P}V$ is a subset, then $\overline{U}$ is the common zero set of all homogeneous polynomials vanishing on U. The same set is obtained if one takes the closure in the usual topology, but the Zariski closure is more useful when dealing with polynomials. We may think of $J(Y, Z)$ as the union of the cones $\cup_{q \in Y} J(q, Z)$ (or as the union of the cones over Y with vertices points of Z).

3.3. **Secant varieties.** If $Y = Z$, we call $\sigma_2(Y) = J(Y, Y)$ the *secant variety* of Y. By the discussion above, $\sigma_2(Y)$ contains all points of all secant and tangent lines to Y. Similarly, define the join of k varieties to be the closure of the union of the corresponding $\mathbb{P}^{k-1}$'s or by induction as $J(Y_1, \ldots, Y_k) = J(Y_1, J(Y_2, \ldots, Y_k))$. Define *k-th secant variety of Y* to be $\sigma_k(Y) = J(Y, \ldots, Y)$, the join of k copies of Y. For smooth varieties $Y \subset \mathbb{P}V$, let $\tau(Y)$ denote the union of all points on all embedded tangent lines to Y. Usually $\tau(Y)$ is a hypersurface in $\sigma_2(Y)$.

For a variety $X \subset \mathbb{P}V$, and point $p \in \mathbb{P}V$, the *X-rank of p* is the smallest number r such that p is in the linear span of r points of X. Thus $\sigma_r(X)$ is the Zariski closure of the set of points of X-rank r. The *X-border rank of p* is the smallest r such that $p \in \sigma_r(X)$. The *typical X-rank* of $\mathbb{P}V$ is the smallest r such that $\sigma_r(X) = \mathbb{P}V$.

3.4. **Expected dimensions.** The expected dimension of $J(Y, Z)$ is $\min\{\dim Y + \dim Z + 1, \dim \mathbb{P}V\}$ because a point $x \in J(Y, Z)$ is obtained by picking a point of Y, a point of Z, and a point on the line joining the two points. This expectation fails if and only if a general point of $J(Y, Z)$ lies on a family of lines intersecting Y and Z, since when this happens one can vary the points on Y and Z used to form the secant line without varying the point x.

Similarly, the expected dimension of $\sigma_r(Y)$ is $r(\dim Y) + r - 1$, which fails if and only if a general point of $\sigma_r(Y)$ lies on a family of secant $\mathbb{P}^{r-1}$'s to Y.

3.5. **The Segre variety.** A linear map $f : V_1^* \to V_2$, that is, an element $f \in V_1 \otimes V_2$, is of rank one if and only if there exists $v_1 \in V_1$, $v_2 \in V_2$ such that, expressed as a tensor, $f = v_1 \otimes v_2$. Geometrically, v_1 is the annihilator of $\ker(f) \subset V_1^*$ and v_2 spans the image of f. In bases, v_1 corresponds to a column vector, v_2 a row vector; and the matrix of any rank one linear map is obtained by multiplying a column vector with a row vector. We want to study linear maps and tensors by their rank

(which is unchanged if we multiply the map by a nonzero constant), so we utilize the following geometric object:

Define $Seg(\mathbb{P}V_1 \times \mathbb{P}V_2) \subset \mathbb{P}(V_1 \otimes V_2)$, the *(two-factor) Segre variety*, to be the projectivization of all the rank one elements of $V_1 \otimes V_2$. Here Seg is the injective map

$$Seg : \mathbb{P}V_1 \times \mathbb{P}V_2 \to \mathbb{P}(V_1 \otimes V_2)$$

$$([v_1], [v_2]) \mapsto [v_1 \otimes v_2],$$

which, in bases, corresponds to multiplying a column vector (defined up to scale) with a row vector (defined up to scale) to get a rank one rectangular matrix (defined up to scale). Note that $\hat{\sigma}_r(Seg(\mathbb{P}V_1 \times \mathbb{P}V_2))$ is isomorphic to the set of $(\dim V_1 \times \dim V_2)$-matrices of rank at most r, as the rank at most r matrices are exactly those that can be written as the sum of r matrices of rank one.

More generally, the projectivization of the set of decomposable tensors in $V_1 \otimes \cdots \otimes V_n$, $\mathbb{P}\{T \in V_1 \otimes \cdots \otimes V_n \mid \exists v_j \in V_j, \ T = v_1 \otimes \cdots \otimes v_n\}$, may be identified with the product $\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_n$. Let $Seg(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_n) \subset \mathbb{P}(V_1 \otimes \cdots \otimes V_n)$ denote the corresponding variety, the *(n-factor) Segre variety*.

3.6. **The Segre variety and border rank.** For any variety $X \subset \mathbb{P}V$, a point of $\sigma_2(X)$ is a point on a secant line or a point on a line that is a limit of secant lines, so, if X is smooth, a point of $\sigma_2(X)$ is either on X, on a secant line to X, or on a tangent line to X. Equation (2.2.2), when projectivized, describes a curve of points on secant lines of $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$ limiting to a point on a tangent line to $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$, that is, a point of $\hat{\tau}(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$. In other words, equation (2.2.2) exhibits an explicit tensor whose rank is different from its border rank. (Aside: $Seg(\mathbb{P}V_1 \times \mathbb{P}V_2)$ has the unusual property that any point on a tangent line is also on a secant line, and, more generally, for $X = Seg(\mathbb{P}V_1 \times \mathbb{P}V_2)$, X-rank coincides with X-border rank, but this property fails to hold for n factor Segre varieties with $n > 2$.)

3.7. **Geometric formulations of notions from complexity.** We can now give geometric formulations of the concepts introduced in §1 and §2:

- The *border rank* of a tensor $T \in V_1 \otimes \cdots \otimes V_n$, $\underline{\mathbf{R}}(T)$, defined in §2 above, is the smallest r such that $[T] \in \sigma_r(Seg(\mathbb{P}V_1 \times \cdots \times \mathbb{P}V_n))$.
- The *border rank of matrix multiplication*

$$M_{m,n,p} : (\mathbb{C}^{m*} \otimes \mathbb{C}^n) \times (\mathbb{C}^{n*} \otimes \mathbb{C}^p) \to (\mathbb{C}^{m*} \otimes \mathbb{C}^p)$$

  is the smallest r such that

  $$[M_{m,n,p}] \in \sigma_r(Seg(\mathbb{P}(\mathbb{C}^m \otimes \mathbb{C}^{n*}) \times \mathbb{P}(\mathbb{C}^n \otimes \mathbb{C}^{p*}) \times \mathbb{P}(\mathbb{C}^{m*} \otimes \mathbb{C}^p))).$$

- The *exponent of matrix multiplication* is

  $$log(\varliminf_{n \to \infty}\{\min_r\{[M_{n,n,n}] \in \sigma_r(Seg(\mathbb{P}^{n^2-1} \times \mathbb{P}^{n^2-1} \times \mathbb{P}^{n^2-1}))\}\}).$$

- *Upper bounds for border rank* for a given n can be established by finding values of r such that $[M_{n,n,n}] \in \sigma_r(Seg(\mathbb{P}^{n^2-1} \times \mathbb{P}^{n^2-1} \times \mathbb{P}^{n^2-1}))$ and lower bounds by finding values of r such that $[M_{n,n,n}] \notin \sigma_r(Seg(\mathbb{P}^{n^2-1} \times \mathbb{P}^{n^2-1} \times \mathbb{P}^{n^2-1}))$.
- The *typical rank* of an element of $\mathbb{C}^a \otimes \mathbb{C}^b \otimes \mathbb{C}^c$ is the smallest r such that $\sigma_r(Seg(\mathbb{P}^{a-1} \times \mathbb{P}^{b-1} \times \mathbb{P}^{c-1})) = \mathbb{P}(\mathbb{C}^a \otimes \mathbb{C}^b \otimes \mathbb{C}^c)$.

3.8. **What is known regarding matrix multiplication.** The problem of determining the typical rank for the spaces of tensors that include the multiplication of square matrices has been completely solved:

Theorem 3.8.1 (T. Lickteig [44]). *For all $n \neq 3$,*

$$\dim \sigma_r(Seg(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \times \mathbb{P}^{n-1})) = \min\{r(3n-2)-1, n^3-1\}.$$

Theorem 3.8.1 shows that Strassen's algorithm for 2×2 matrices could have been anticipated, as $\sigma_7(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)) = \mathbb{P}(\mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \mathbb{C}^4)$. (This observation is enough to recover the bound $\omega < 2.81$ mentioned in §1.2.) We outline the proof of Theorem 3.8.1 and discuss what is known about typical rank in §9.

For the $n = 3$ case we have:

Theorem 3.8.2 (Strassen [52]). $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2))$ *is a hypersurface of degree 9.*

This case was solved by finding an explicit equation vanishing on $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2))$. In §6 we discuss this equation and its consequences for matrix multiplication.

The best lower bound on the border rank of matrix multiplication is:

Theorem 3.8.3 (Lickteig [43]). $\underline{\mathbf{R}}(M_{m,m,m}) \geq \frac{3m^2}{2} + \frac{m}{2} - 1$.

While we do not provide Lickteig's proof here, we remark that implicit in his proof is the presence of auxiliary varieties that we believe will play a central role in future work. In §12 we describe some of these varieties, including the *subspace variety*, which is implicit in his proof.

The best lower bound on the rank of matrix multiplication is:

Theorem 3.8.4 (M. Bläser [10]). $\mathbf{R}(M_{m,m,m}) \geq \frac{5}{2}m^2 - 3m$.

A new proof of Bläser's theorem is presented in §6.2. Bläser has also proved that $\mathbf{R}(M_{3,3,3}) \geq 19$ [11], and we discuss the main tool in the proof of Bläser's 19-theorem in §14.2.

The best upper bound for the exponent of matrix multiplication is $\omega < 2.38$, due to D. Coppersmith and S. Winograd [24]. They use methods of Strassen [53]. We do not discuss these asymptotic bounds, as we have no geometric interpretation for them. However, an earlier asymptotic bound due to A. Schönhage [48] does have relations with geometry. We discuss the geometric aspect of Schönhage's argument in §8.1 and present his explicit approximate algorithm for multiplying 3×3 matrices using 21 multiplications.

There is also an algorithm for multiplying 3×3 matrices using 23 multiplications due to J. Laderman [34], which we do not discuss.

The only case where the exact rank and border rank are known for the multiplication of square matrices is 2×2 matrices:

Theorem 3.8.5 (Winograd [58]). $\mathbf{R}(M_{2,2,2}) = 7$.

J. Hopcroft and L. Kerr [31] proved Theorem 3.8.5 in the case of algorithms with integer coefficients.

We do not discuss the proof of Theorem 3.8.5. We do mention here that an alternative proof is a consequence of a theorem of R. Brockett and D. Dobkin [13] that the rank of the multiplication in any simple algebra is at least twice the dimension of the algebra minus one. A proof of the Brockett-Dobkin theorem,

due to W. Baur and presented in [14], proceeds by splitting any putative simpler algorithm several times to eventually obtain a contradiction by producing a right ideal that is contained in a left ideal.

Theorem 3.8.6 ([36]). $\underline{\mathbf{R}}(M_{2,2,2}) = 7$.

To prove Theorem 3.8.6 we first decomposed $\sigma_6(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ into various components based on how the limiting $\mathbb{P}^5$ was obtained from a family of secant $\mathbb{P}^5$'s. (By Theorem 3.8.5 one needs to examine only limiting planes.) For each possible limiting type we wrote down normal forms for the limit. Then we applied variants of Baur's proof of the Brockett-Dobkin theorem in each case to obtain a contradiction. In §7 we outline how to determine such limiting planes, which are also used in the construction of upper bounds.

3.9. **What is not known.** A central conjecture in algebraic complexity theory is that the exponent of matrix multiplication is two. It is also of importance to find good upper and lower bounds for matrix multiplication for small and human scale values of n. Already for $n = 3$ all that is known is $14 \le \underline{\mathbf{R}}(M_{3,3,3}) \le 21$ and $19 \le \mathbf{R}(M_{3,3,3}) \le 23$. While the problem of finding the defining equations for secant varieties of Segre varieties is a means to an end as far as matrix multiplication is concerned, for the purposes of *algebraic statistics*, it is essential to develop techniques for finding these equations and the equations of related varieties. For the area of phylogenetic invariants, an important open problem is to find the defining equations for $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$, as explained in §5.2. Other open questions are discussed in the remaining sections.

4. SECANT VARIETIES IN ALGEBRAIC GEOMETRY

In this section we take a detour from our main subject to discuss two other situations where secant varieties arise: the solution of the polynomial Waring problem and the resolution of Hartshorne's conjecture on linear normality. Nothing in this section is used in the following sections (with the exception of the definition of the Grassmannian, used in §7).

4.1. **The polynomial Waring problem and variants.** Waring asked if there exists a function $k(s)$ such that every natural number n is expressible as at most s k-th powers of integers. Hilbert answered Waring's question affirmatively, and this function has essentially been determined; see [55] for a survey of results and related open questions.

The polynomial Waring problem is as follows:

What is the smallest $r_0 = r_0(d, n)$ such that a general homogeneous polynomial $P(x^1, \ldots, x^n)$ of degree d in n variables is expressible as the sum of r_0 d-th powers of linear forms?

Let $V = \mathbb{C}^n$, and let S^dV^* denote the space of homogeneous polynomials of degree d on V. Let

$$v_d : \mathbb{P}V^* \to \mathbb{P}S^dV^*$$

$$[\alpha] \mapsto [\alpha \circ \cdots \circ \alpha]$$

denote the *Veronese map* that sends the projectivization of a linear form to the projectivization of its d-th power. Thus the image is the set of (projectivized) d-th powers of linear forms. Similarly $\sigma_p(v_d(\mathbb{P}V))$ is the Zariski closure of the set

of homogeneous polynomials that are expressible as the sum of p d-th powers of linear forms. So the polynomial Waring problem may be re-expressed as:

Let $V = \mathbb{C}^n$ and let $X = v_d(\mathbb{P}V^*)$. What is the typical X-rank of an element of $\mathbb{P}S^d V^*$; that is, what is the smallest $r_0 = r_0(d, n)$ such that $\sigma_{r_0}(v_d(\mathbb{P}V^*)) = \mathbb{P}S^d V^*$?

This problem was solved by Alexander and Hirshowitz [4]: all $\sigma_r(v_d(\mathbb{P}^n))$ are of the expected dimension except $\sigma_7(v_3(\mathbb{P}^4)), \sigma_5(v_4(\mathbb{P}^2)), \sigma_9(v_4(\mathbb{P}^3)), \sigma_{14}(v_4(\mathbb{P}^4))$ (which are all hypersurfaces), and $\sigma_r(v_2(\mathbb{P}^n))$, $2 \leq r \leq n$ (where $\dim \sigma_r(v_2(\mathbb{P}^n)) = rn - \frac{r^2 - 3r}{2} - 1$) . In other words,

Theorem 4.1.1 ([4]). *A general homogeneous polynomial of degree d in n variables is expressible as the sum of*

$$r_0(d, n) = \lceil \frac{\binom{n+d-1}{d} + 1}{n} \rceil$$

d-th powers with the exception of the cases $r_0(3, 5) = 8$, $r_0(4, 3) = 6$, $r_0(4, 4) = 10$, $r_0(4, 5) = 15$, and $d = 2$, where $r_0(2, n) = n$.

For a beautiful discussion of this problem and its history, including a self-contained proof, see [12].

A variant of the polynomial Waring problem is to find the typical rank of alternating tensors. Let $\Lambda^k V \subset V^{\otimes k}$ be the space of alternating tensors. Let $G(k, V) \subset \mathbb{P}(\Lambda^k V)$ denote the projectivization of the set of minimal rank alternating tensors. This variety is called the *Grassmannian* of k-planes through the origin in V (that is, we have a bijection, for linearly independent sets of vectors $v_1, \ldots, v_k$, $\mathrm{Span}\{v_k, \ldots, v_k\} \simeq [v_1 \wedge \cdots \wedge v_k]$). In [18] they show that for $3 \leq k \leq \frac{n}{2}$, $\sigma_r(G(k, n))$ has the expected dimension provided that $r \leq \frac{n}{k}$. Previous to that, it was known that $G(2, n)$ had all secant varieties defective and $G(3, 7), G(4, 8)$, and $G(3, 9)$ all had their "last" secant variety before filling defective. (The examples $G(2, n)$ are just the skew symmetric matrices of minimal rank; the examples $G(3, 7)$ and $G(4, 8)$ can be understood in terms of the geometry of the exceptional groups G_2 and $Spin_7$.)

Further generalizations of the polynomial Waring problem and their uses are discussed in [21].

The main tool for proving that secant varieties are of the expected dimension is *Terracini's Lemma* 9.2.1. Proving they are degenerate, other than in cases when it is obvious, is more subtle. For these Waring problems, there are often interpretations of the exceptional cases in terms of the geometry of Veronese varieties. The most interesting exception in the case of secant varieties of Segre varieties is $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2))$, which is discussed in detail in §6. In the proof of Lemma 3.16 of [1], a geometric explanation of the degeneracy is given: any four points on $Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)$ lie in some $v_3(\mathbb{P}^2) \subset \mathbb{P}(S^3\mathbb{C}^2) \subset \mathbb{P}(\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2)$. Thus when one applies Terracini's Lemma, each of the four embedded tangent spaces to the Segre must have at least a two-dimensional subspace in the $\mathbb{P}(S^3\mathbb{C}^2) = \mathbb{P}^9$, forcing a degeneracy. It would be interesting to have a systematic understanding of the Veronese varieties that unirule these exceptional cases, e.g., in terms of representation-theoretic data.

4.2. **Zak's theorems.** Smooth projective varieties $X^n \subset \mathbb{P}^{n+a}$ of small codimension were shown by W. Barth and M. Larsen (see, e.g., [7]) to behave topologically

as if they were *complete intersections*, i.e, the zero set of a homogeneous polynomials. This motivated R. Hartshorne's famous conjecture on complete intersections [30], which says that if $a < \frac{n}{2}$, then X must indeed be a complete intersection. A first approximation to this difficult conjecture was also made by Hartshorne - his conjecture on *linear normality*, which was proved by F. Zak [59] (see [60] for an exposition). The linear normality conjecture was equivalent to showing that if $a < \frac{n}{2} + 2$ and X is not contained in a hyperplane, then $\sigma_2(X) = \mathbb{P}^{n+a}$. Zak went on to classify the exceptions in the equality case $a = \frac{n}{2} + 2$. There are exactly four, which Zak called *Severi varieties* (after Severi, who solved the $n = 2$ case [50]). The first three Severi varieties have already been introduced: $v_2(\mathbb{P}^2) \subset \mathbb{P}^5$, $Seg(\mathbb{P}^2 \times \mathbb{P}^2) \subset \mathbb{P}^7$, and $G(2,6) \subset \mathbb{P}^{13}$. The last is the complexified Cayley plane $\mathbb{OP}^2 \subset \mathbb{P}^{15}$. These four varieties admit uniform interpretations as the set of rank one elements in a rank three Jordan algebra over a composition algebra.

An interesting open question is the *secant defect problem*. For a smooth projective variety $X^n \subset \mathbb{P}V$ not contained in a hyperplane, with $\sigma_2(X) \neq \mathbb{P}V$, let $\delta(X^n) = 2n + 1 - \dim \sigma_2(X)$, the *secant defect of X*. The largest known secant defect is 8, which occurs for the complexified Cayley plane. Problem: Is a secant defect larger than 8 possible? If we do not assume the variety is smooth, the defect is unbounded. (This question was posed originally in [42].)

5. OTHER USES OF SECANT VARIETIES OF SEGRE VARIETIES AND RELATED OBJECTS

5.1. Algebraic statistics. A *probability distribution* is a point in $V := \mathbb{R}^{a_1} \otimes \cdots \otimes \mathbb{R}^{a_n}$ where the sums of coordinate elements add to one. For example, say we have two biased coins. Then $V = \mathbb{R}^2 \otimes \mathbb{R}^2$ and a point corresponds to a matrix

$$\begin{pmatrix} p_{h,h} & p_{h,t} \\ p_{t,h} & p_{t,t} \end{pmatrix}$$

where $p_{h,h}$ is the probability that both coins, when tossed, come up heads, etc.

A *statistical model* is a family of probability distributions given by a set of constraints that these distributions must satisfy, that is, a subset of V. An *algebraic statistical model* consists of all joint probability distributions that are the common zeros of a set of polynomials on V.

To continue our example, assume the outcome of the two coin tosses do not effect each other (as is the case with actual coins). Then the resulting matrix must have rank one. The set of all rank one, 2×2 matrices in the positive coordinate simplex is the corresponding algebraic statistical model, but it is almost equivalent to work with $\hat{S}eg(\mathbb{RP}^1 \times \mathbb{RP}^1)$.

Now assume we can measure the outcome of two of the events (tosses) but there may be a third event whose outcome influences the outcome of the other two, although the outcomes of the two events we can measure are independent of one another (e.g. someone may be cheating by using magnets). Naïvely we should have a point of $\mathbb{R}^{a_1} \otimes \mathbb{R}^{a_2} \otimes \mathbb{R}^{a_3}$, but we can't measure the possible third; in fact we don't even know what a_3 should be.

Let's posit that some fixed a_3 parametrizes the third outcome (if we posit there is no third event, then one takes $a_3 = 1$). Then we sum up over all possibilities for the third factor to get a 2×2 matrix whose entries are

(5.1.1) $p_{i,j} = p_{i,j,1} + \cdots + p_{i,j,a_3}, \quad 1 \leq i \leq a_1,\ 1 \leq j \leq a_2$

The algebraic statistical model here is the set of rank at most a_3 matrices in the space of $a_1 \times a_2$ matrices, $\hat{\sigma}_{a_3}(Seg(\mathbb{RP}^{a_1-1} \times \mathbb{RP}^{a_2-1}))$. Thus, given a particular model, e.g. a fixed value of a_3, to test if our data (as points of $\mathbb{R}^{a_1} \otimes \mathbb{R}^{a_2}$) fits the model, we can check if it (mostly) lies inside $\hat{\sigma}_{a_3}(Seg(\mathbb{RP}^{a_1-1} \times \mathbb{RP}^{a_2-1}))$.

In algebraic statistics one wants to test if a given model is applicable to a particular collection of data sets. Thus in particular, one needs a way of testing if a point $p \in \mathbb{R}^{a_1} \otimes \cdots \otimes \mathbb{R}^{a_n}$ is a sum of at most r decomposable elements.

It is easier to solve this problem first over the complex numbers and then return to the real situation later. Thus to test models of the type discussed above, one needs defining equations for secant varieties of Segre varieties. In sections §§6 - 13 we discuss methods for finding such equations. These methods are applicable to finding equations for more general algebraic statistical models as well. They all rely on exploiting the group under which the model is invariant.

For more on algebraic statistics see [32, 47].

5.2. **Phylogenetic invariants.** This is a special case of algebraic statistics but is sufficiently important to merit its own subsection. In order to determine a tree that describes the evolutionary descent of a family of extant species, J. Lake [35], J. Cavender and J. Felsenstein [20] proposed the use of what is now called algebraic statistics by viewing the four bases composing DNA as the possible outcomes of a random variable.

Given a collection of extant species, one would like to assess the likelihood of each of the possible evolutionary trees that could have led to them. To do this, one can test the various DNA sequences that arise to see which algebraic statistical model fits best. More than that, the invariants discussed below identify the trees (nearly) uniquely.

In what follows, contrary to some of the literature, we ignore time.

The simplest situation is where one species gives rise to two new species. This can be pictured by a tree of the form
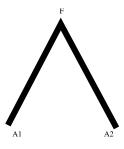

FIGURE 1

There are three species involved—the parent F and the two offspring, $A1$, $A2$—so the DNA occupies a point of the positive coordinate simplex in $\mathbb{R}^4 \otimes \mathbb{R}^4 \otimes \mathbb{R}^4$, and we make our lives easier by working with $\mathbb{P}(\mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \mathbb{C}^4)$. We can measure the DNA of the two new species but not the ancestor, so the relevant algebraic statistical model is $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3))$, which is well understood. Here $a_1 = a_2 = a_3 = 4$ in the analogue of equation (5.1.1), and we sum over the third factor. In this case there is nothing new to be learned from the model.

The next case is where a parent F gives rise to three new species, $A1, A2, A3$. Assuming species bifurcate, one might think that this gives rise to three distinct algebraic statistical models, as we could have F giving rise to A_1 and G, then G splitting to A_2 and A_3 or two other possibilities. However, all three scenarios give rise to the same algebraic statistical model: $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$; see [6]. In other words, the following pictures all give rise to the same algebraic statistical models.
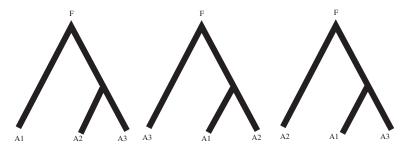

FIGURE 2

The defining equations of $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$ are *not known*, and for reasons we explain below, it is a central question for the study of phylogenetic invariants to find them.

Now consider the case where there are four new species—$A1, A2, A3, A4$—all from a common ancestor F. Here finally there are three different scenarios that give rise to distinct algebraic statistical models.
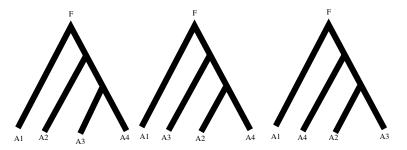

FIGURE 3

Note that there are no pictures like

FIGURE 4

because such pictures give rise to equivalent algebraic statistical models to the exhibited trees.

We consider that parent F first gives rise to A_1 and E, and then E gives rise to A_2 and G, and G gives rise to A_3 and A_4, as well as the equivalent (by the discussion above) scenarios. The resulting algebraic statistical model is

$$\Sigma_{12,34} := \sigma_4(Seg(\mathbb{P}A_1 \times \mathbb{P}A_2 \times \mathbb{P}(A_3 \otimes A_4))) \cap \sigma_4(Seg(\mathbb{P}(A_1 \otimes A_2) \times \mathbb{P}A_3 \times \mathbb{P}A_4)).$$

Similarly we get the other two possibilities:

$$\Sigma_{13,24} := \sigma_4(Seg(\mathbb{P}A_1 \times \mathbb{P}A_3 \times \mathbb{P}(A_2 \otimes A_4))) \cap \sigma_4(Seg(\mathbb{P}(A_1 \otimes A_3) \times \mathbb{P}A_2 \times \mathbb{P}A_4))$$

and

$$\Sigma_{14,23} := \sigma_4(Seg(\mathbb{P}A_1 \times \mathbb{P}A_4 \times \mathbb{P}(A_2 \otimes A_3))) \cap \sigma_4(Seg(\mathbb{P}(A_1 \otimes A_4) \times \mathbb{P}A_2 \times \mathbb{P}A_3)).$$

Note that these three models are isomorphic as projective varieties but are situated differently in $\mathbb{P}(A_1 \otimes A_2 \otimes A_3 \otimes A_4)$; thus having defining equations for them would enable one to test between different evolutionary possibilities. An essential result of [6] is:

Once one has defining equations for $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$, one can obtain defining equations for all algebraic statistical models corresponding to bifurcating phylogenetic trees.

The proof relies on two results. First, no matter how many species one observes, because of the structure of the evolutionary trees, the resulting algebraic statistical model is an intersection of fourth secant varieties of Segre varieties corresponding to summing over the four outcomes on a hidden variable. The second ([6], Theorem 11) is equivalent to (and arrived at independently of) Proposition 12.2.2 below, which in particular reduces the study of the fourth secant variety of any triple Segre product to the study of $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$.

5.3. **Entanglement and quantum computing.** In *quantum computing* (see, e.g., [8] and the numerous references therein) a *pure state* corresponds to a point of $\mathbb{P}(\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2)$ where there are N copies of $\mathbb{C}^2$. A *product state* corresponds to a point of $Seg(\mathbb{P}^1 \times \cdots \times \mathbb{P}^1) \subset \mathbb{P}(\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2)$. A pure state is *entangled* if it is not a product state, and quantum computing is based on exploiting entangled states. A perhaps overly optimistic program is to classify the $U(2) \times \cdots \times U(2)$ and/or $SL(2, \mathbb{C}) \times \cdots \times SL(2, \mathbb{C})$ orbits in $\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$, which would give a complete classification of entangled states. Failing that, one is interested in finding specific measures of entanglement. One measure of entanglement is called the *Schmidt measure*, introduced in [26]. In the language of this paper, the Schmidt measure of a tensor is the base two log of its rank. In [25] J. Eisert and D. Gross observe that a tensor of a given Schmidt measure might be a limit of tensors of a lower Schmidt measure; in fact they give the explicit example of (2.2.1) in their equation (19), where their $|1, 0, 0 >$ corresponds to $a_1 \otimes b_1 \otimes c_1$ in (2.2.1). In [25] they decompose $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \backslash 0$ into the union of four disjoint components which they label S, B, W, GHZ. In the language of this paper, the components are

$$S = \hat{S}eg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)\backslash 0 = \hat{S}eg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)\backslash 0,$$

$$B = \{\hat{S}eg(\mathbb{P}A \times \mathbb{P}(B \otimes C)) \cup \hat{S}eg(\mathbb{P}(A \otimes B) \times \mathbb{P}C) \cup \hat{S}eg(\mathbb{P}(A \otimes C) \times \mathbb{P}B)\}\backslash\{0 \cup S\},$$

$$W = \hat{\tau}(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))\backslash\{0 \cup S \cup B\},$$

$$GHZ = \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2 \backslash\{0 \cup S \cup B \cup W\}.$$

Compare B with the discussion of *flattenings* in §12.

There is a vast literature regarding entanglement, and there does not appear yet to be a consensus regarding the best way to measure entanglement, but it is clear that secant varieties of Segre varieties and related auxiliary varieties are relevant to the problem.

6. Strassen's equations and lower bounds for rank and border rank

In this section we introduce Strassen's equations and use them to give a new proof of Bläser's $\frac{5}{2}$-theorem. In §10 we rephrase the equations invariantly and give generalizations.

6.1. **Strassen's equations.** Let A, B, C be vector spaces respectively of dimensions $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

Theorem 6.1.1 (Strassen [52]). *Let $3 \leq \mathbf{a} \leq \mathbf{b} = \mathbf{c} \leq r$. Let $T \in \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ and $\alpha \in A^*$ be such that $T_\alpha := T(\alpha) \in B \otimes C$, considered as a map $T_\alpha : C^* \to B$, is of full rank. For each $\alpha^1, \alpha^2 \in A^*$, define the linear map $T_{\alpha,\alpha^j} : B \to B$ by $T_{\alpha,\alpha^j} = T_{\alpha^j} T_\alpha{}^{-1}$. Then*

$$\mathrm{Rank}[T_{\alpha,\alpha^1}, T_{\alpha,\alpha^2}] \leq 2(r - \mathbf{b})$$

where $[S, T] = ST - TS$ is the commutator of endomorphisms.

Theorem 6.1.1 enables us to prove part of Theorem 3.8.2:

Corollary 6.1.2 (Strassen [52]). $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)) \neq \mathbb{P}(\mathbb{C}^3 \otimes \mathbb{C}^3 \otimes \mathbb{C}^3)$.

Proof of corollary. For generic $T \in A \otimes B \otimes C = \mathbb{C}^3 \otimes \mathbb{C}^3 \otimes \mathbb{C}^3$ and $\alpha, \alpha^1, \alpha^2 \in A^*$, one has $\mathrm{Rank}([T_{\alpha,\alpha^1}, T_{\alpha,\alpha^2}]) = 3$, but for points in $\sigma_4(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$, the rank is at most two. $\square$

An easy calculation with Terracini's Lemma (9.2.1) shows that $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2))$ is at least a hypersurface, so the above corollary shows it is exactly a hypersurface. Strassen's equations are not presented as polynomials above. In §10 we describe them as polynomials and give generalizations.

Recall that matrix multiplication $M_{m,m,m}$ corresponds to a point in $A \otimes B \otimes C$ when $A^*, B^*, C \simeq \mathbb{C}^{m^2}$.

Corollary 6.1.3 (Strassen [52]). $\underline{\mathbf{R}}(M_{m,m,m}) \geq \frac{3m^2}{2}$.

Proof. Write out $M_{m,m,m}$ explicitly in a good basis and take a generic $\alpha \in A^* = Mat_{m \times m}$. Then the corresponding linear map T_α will be a block diagonal matrix with blocks of size m, each block identical and the entries of the block arbitrary. So we have $\mathrm{Rank}([T_{\alpha,\alpha^1}, T_{\alpha,\alpha^2}]) = m^2$. Hence $m^2 \leq 2(r - m^2)$, and the result follows. $\square$

6.2. **Proof of Bläser's lower bound.** Here is a proof of Theorem 3.8.4 that uses Theorem 6.1.1, which is implicit, but hidden, in the original proof.

Lemma 6.2.1. *Let U be a vector space, and let $P \in S^d U^* \backslash 0$. Let $u_1, \ldots, u_n$ be a basis of U. Then there exists a subset $u_{i_1}, \ldots, u_{i_s}$ of cardinality $s \leq d$ such that $P \mid_{\langle u_{i_1}, \ldots, u_{i_s} \rangle}$ is not identically zero.*

The proof is an easy exercise. Let $Mat_{m \times m}$ denote the vector space of $m \times m$ matrices.

Lemma 6.2.2. *Given any basis of $Mat^*_{m \times m}$, there exists a subset of at least $m^2 - 3m$ basis vectors that annhilate elements $Id, x, y \in Mat_{m \times m}$ such that $[x, y] := xy - yx$ has maximal rank m.*

Proof. Let $A = Mat_{m \times m} \simeq U^* \otimes W$. Fixing a basis of A^* is equivalent to fixing its dual basis of A. By Lemma 6.2.1 with $P = \det$, we may find a subset S_1 of at most m elements of our basis of A with some $z \in Span(S_1)$ with $\det(z) \neq 0$. We use $z : U \to W$ to identify $U \simeq W$, which enables us to now consider A as an algebra with z playing the role of the identity element.

Now let $a \in A$ be generic. Then the map $ad(a) : A \to A$, $x \mapsto [a, x]$ will have a one-dimensional kernel. By letting $P = ad(a)^*(det)$ and applying Lemma 6.2.1 again, we may find a subset S_2 of our basis of cardinality at most m such that there is an element $x \in A$ such that $ad(a)(x)$ is invertible. Note that $ad(x) : A \to A$ also is such that there are elements y with $ad(x)y$ invertible. Thus we may apply Lemma 6.2.1 a third time to find a cardinality at most m subset S_3 of our basis such that $ad(x)y$ is invertible. Now in the worst possible case our three subsets are of maximal cardinality and do not intersect, in which case we have a cardinality $m^2 - 3m$ subset of our dual basis that annihilates $z = Id, x, y$ with $\text{Rank}([x, y]) = m$. □

Proof of Theorem 3.8.4. Let ϕ denote a computation of $M = M_{m,m,m}$ of length r. Since $\text{Lker}(M) = 0$ (that is, $\forall a \in A \backslash 0$, $\exists b \in B$ such that $M(a, b) \neq 0$), we may write $\phi = \psi_1 + \psi_2$ with $\mathbf{R}(\psi_1) = m^2$, $\mathbf{R}(\psi_2) = r - m^2$ and $\text{Lker}(\psi_1) = 0$. Now consider the m^2 elements of A^* appearing in ψ_1. Since they span A^*, by Lemma 6.2.2 we may choose a subset of $m^2 - 3m$ of them that annihilate Id, x and y, where x, y are such that $[x, y]$ has full rank. Let ϕ_1 denote the sum of all monomials in ψ_1 whose A^* terms annihilate Id, x, y, so $\mathbf{R}(\phi_1) \geq m^2 - 3m$. Let $\phi_2 = \psi_1 - \phi_1 + \psi_2$.

Now apply Theorem 6.1.1 with $T = \phi_2$, $\alpha = Id$, $\alpha_1 = x$, $\alpha_2 = y$ to get $\underline{\mathbf{R}}(\phi_2) \geq \frac{1}{2}\text{rank}[x, y] + m^2 = \frac{3}{2}m^2$ and thus $\mathbf{R}(\phi_1 + \phi_2) \geq \frac{5}{2}m^2 - 3m$. □

7. Limits of secant planes

There are several reasons for studying points on $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ that are not on secant $\mathbb{P}^{r-1}$'s. First, in order to prove that a set of equations E is a set of defining equations for $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$, one must prove that any point in the zero set of E is either a point on a secant $\mathbb{P}^{r-1}$ or on a limit $\mathbb{P}^{r-1}$. For example, the proof of the set-theoretic GSS conjecture (see §12) in [37] proceeded in this fashion. Second, to establish lower bounds for the border rank of a given tensor, e.g., matrix multiplication, one could try to first prove that it cannot lie on any secant $\mathbb{P}^{r-1}$ and then that it cannot lie on any limiting $\mathbb{P}^{r-1}$ either. This was the technique of proving that $\underline{\mathbf{R}}(M_{2,2,2}) = 7$ in [36]. Finally, a central ingredient for writing explicit approximate algorithms for matrix multiplication is to exploit certain limiting $\mathbb{P}^{r-1}$'s discussed below.

This section is used only in §8, which is not used in the remainder of the article, so both this section and §8 can be skipped by readers primarily interested in the equations of secant varieties of Segre varieties.

7.1. Limits for arbitrary projective varieties. Let $X \subset \mathbb{P}V$ be a projective variety. In this subsection we give a coarse description of the different types of points in $\sigma_r(X)$. Let $\sigma_r^0(X)$ denote the set of points on $\sigma_r(X)$ that lie on a

secant $\mathbb{P}^{r-1}$, that is $[p] \in \sigma_r(X)$ for which there exists $[x_1], \ldots, [x_r] \in X$ such that $p \in \langle x_1, \ldots, x_r \rangle$, where $\langle x_1, \ldots, x_r \rangle$ denotes the span of $x_1, \ldots, x_r \in V$. We work inductively, so we assume we know the nature of points on $\sigma_{r-1}(X)$ and study points on $\sigma_r(X) \backslash (\sigma_r^0(X) \cup \sigma_{r-1}(X))$.

It is convenient to study the limiting r-planes as points on the cone over the Grassmannian in its Plucker embedding, $G(r, V) \subset \mathbb{P}(\Lambda^r V)$ (see the end of §4.1). That is, we consider the curve of r planes as being represented by $x_1(t) \wedge \cdots \wedge x_r(t)$ and examine the limiting plane as $t \to 0$. (There must be a unique such plane, as the Grassmannian is compact.) Points whose X-rank is larger than their X-border rank must belong to such limiting planes. The approximate algorithms in §8 utilize these special points in $\sigma_r(X)$ when $X = Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$. The algorithm of Example 8.1.1 below uses a point of the form (7.1.1) below, the algorithm of Example 8.1.3 uses a point of the form (7.1.3) below, and Algorithm 8.2 in [48] uses a point that is obtained after taking twenty derivatives!

Let $[p] \in \sigma_r(X)$. Then there exist curves $x_1(t), \ldots, x_r(t) \subset \hat{X}$ with $p \in \lim_{t \to 0} \langle x_1(t), \ldots, x_r(t) \rangle$. We are interested in the case when $\dim \langle x_1(0), \ldots, x_r(0) \rangle < r$. Use the notation $x_j = x_j(0)$. Assume for the moment that $x_1, \ldots, x_{r-1}$ are linearly independent. Then we may write $x_r = c_1 x_1 + \cdots + c_{r-1} x_{r-1}$ for some constants $c_1, \ldots, c_{r-1}$. Write each curve $x_j(t) = x_j + t x_j' + t^2 x_j'' + \cdots$ where derivatives are taken at $t = 0$.

Consider the Taylor series

$$x_1(t) \wedge \cdots \wedge x_r(t) = (x_1 + t x_1' + t^2 x_1'' + \cdots) \wedge \cdots \wedge (x_{r-1} + t x_{r-1}' + t^2 x_{r-1}'' + \cdots)$$
$$\wedge (x_r + t x_r' + t^2 x_r'' + \cdots)$$
$$= t((-1)^r (c_1 x_1' + \cdots + c_{r-1} x_{r-1}' - x_r') \wedge x_1 \wedge \cdots \wedge x_{r-1}) + t^2(\ldots) + \cdots.$$

If the t coefficient is nonzero, then

$$(7.1.1) \qquad p \in \langle x_1, \ldots, x_{r-1}, (c_1 x_1' + \cdots + c_{r-1} x_{r-1}' - x_r') \rangle.$$

If the t coefficient is zero, then $c_1 x_1' + \cdots + c_{r-1} x_{r-1}' - x_r' = e_1 x_1 + \cdots + e_{r-1} x_{r-1}$ for some constants $e_1, \ldots, e_{r-1}$. In this case we must examine the t^2 coefficient of the expansion. It is

$$(7.1.2) \qquad (\sum_{k=1}^{r-1} e_k x_k' + \sum_{j=1}^{r-1} c_j x_j'' - x_r'') \wedge x_1 \wedge \cdots \wedge x_{r-1},$$

and in this case

$$(7.1.3) \qquad p \in \langle x_1, \ldots, x_{r-1}, (\sum_{k=1}^{r-1} e_k x_k' + \sum_{j=1}^{r-1} c_j x_j'' - x_r'') \rangle.$$

One continues to higher order terms if this is zero.

7.2. Limits for Segre varieties.
A general curve on $\hat{S}eg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is of the form $x(t) = a(t) \otimes b(t) \otimes c(t)$ where $a(t), b(t), c(t)$ are respectively arbitrary curves in $A \backslash 0, B \backslash 0, C \backslash 0$ with $a(0) = a$, etc. We have $x' = a' \otimes b \otimes c + a \otimes b' \otimes c + a \otimes b \otimes c'$ where a', b', c' are respectively arbitrary elements of A, B, C, and higher order derivatives are obtained similarly.

While the easiest way to obtain r points that are linearly dependent in the limit is to have two points limit to the same point, this turns out to be not as useful for upper bound algorithms as more subtle limits. On the other hand, when r is

sufficiently small, any other type of limit involves exploiting the geometry of the
Segre variety as we now explain.

To simplify the situation, we work inductively and just look at "primitive" cases,
that is, require that the points on the limiting $\mathbb{P}^{r-1}$ do not lie on $\sigma_r(Seg(\mathbb{P}A' \times \mathbb{P}B' \times \mathbb{P}C'))$ where $\dim A' \leq \dim A$, etc. (with at least one inequality strict), and
moreover that the points do not lie on $\sigma_{r-1}(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$.

For example, for the two factor Segre $Seg(\mathbb{P}A \times \mathbb{P}B)$ (which, if we are work-
ing by induction, must be studied for the three factor case, as it corresponds to
the case $\dim C' = 1$), in order to have $x_1, \ldots, x_r \in Seg(\mathbb{P}A \times \mathbb{P}B)$ such that
$\dim\langle x_1, \ldots, x_r \rangle < r - 1$ and the points are not contained in some $Seg(\mathbb{P}A' \times \mathbb{P}B')$,
we must have $\dim A + \dim B \leq r$ (see the erratum to [36]). In the erratum to [36]
we determine all possible $x_1, \ldots, x_6 \in Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)$ with $\dim\langle x_1, \ldots, x_6 \rangle < 6$.
The only possible cases where the points fail to lie in some $Seg(\mathbb{P}^0 \times \mathbb{P}B \times \mathbb{P}C)$
occur when they all lie in some $Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)$.

A basic property of projective space is that if $X^n \subset \mathbb{P}^{n+a}$ is a subvariety, then
a general $\mathbb{P}^a$ will intersect X in $\deg(X)$ points. (In fact this is the definition of
the degree of X.) One can calculate that $\deg(Seg(\mathbb{P}^2 \times \mathbb{P}^2)) = 6$ (see, e.g., [29],
lecture 18) and $\operatorname{codim}(Seg(\mathbb{P}^2 \times \mathbb{P}^2)) = 4$. Therefore, for any set of five points on
$Seg(\mathbb{P}^2 \times \mathbb{P}^2)$ that are linearly independent, that is, that span a $\mathbb{P}^4$, there is a sixth
point in the $\mathbb{P}^4$ that also lies on the Segre. Taking the span of these six points as
our $x_i(0)$, we get a limit set that allows the use of derivatives. This type of limit set
is used several times in Example 8.1.3 to build Schönhage's approximate algorithm
for multiplying 3×3 matrices using 21 multiplications.

Similarly, $\deg(Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)) = 6$ and $\operatorname{codim}(Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)) = 4$, which
is exploited in Example 8.1.1.

8. Upper bounds

We now discuss how to use the geometry discussed above to find explicit approx-
imate algorithms for executing a bilinear map.

8.1. **Schönhage's results.** Schönhage [48] isolated a common aspect to certain
approximate algorithms for matrix multiplication, which enabled him to generalize
them and establish upper bounds for the exponent of matrix multiplication without
even having explicit approximate algorithms. The essence of his idea is as follows:

*Say we have two bilinear maps $f : U^* \times V^* \to W$ and $g : \tilde{U}^* \times \tilde{V}^* \to \tilde{W}$. Under
certain conditions, $\underline{\mathbf{R}}(f \oplus g) < \underline{\mathbf{R}}(f) + \underline{\mathbf{R}}(g)$, where $f \oplus g : (U \oplus \tilde{U})^* \times (V \oplus \tilde{V})^* \to (W \oplus \tilde{W})$.*

We explain how to obtain an approximate algorithm for $f \oplus g$ that occurs as a
point p as in (7.1.3). Let $A = U \oplus \tilde{U}$, $B = V \oplus \tilde{V}$, $C = W \oplus \tilde{W}$, and recall from
§7.2 that curves $x_j(t)$ on $\hat{Seg}(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ are of the form $a_j(t) \otimes b_j(t) \otimes c_j(t)$.

We first require that the $a_j(0)$ be the U vectors needed to execute f, the $a_j(0)' = 0$, and the $a_j(0)''$ be the $\tilde{U}$ vectors needed to execute g. For the vectors of B, we
take the $b_j(0)$ to be the V vectors needed for f and the $b_j(0)'$ the $\tilde{V}$ vectors needed
for g, and the C limits are of the same nature as the B limits. Then the sum of
the second derivatives will be $f \oplus g$.

The only problem is, as explained in §7, we need the zero-th and first order
terms to be linearly dependent so that we are allowed to take the sum of the second
derivatives. To obtain linear dependence, the points must lie in some degenerate

position with respect to the Segre, but this is difficult to arrange. Schönhage's solution is to have these limit points in a two factor Segre (where it is easier to have degenerate limits), but this forces one of each U, V, W and $\tilde{U}, \tilde{V}, \tilde{W}$ to be one-dimensional. Moreover, these restrictions only take care of the zero-th order term. To get the first order term killed, two of e.g., $\tilde{U}, \tilde{V}, \tilde{W}$ are taken to be of dimension one and the third, say $\tilde{W}$, to be of dimension roughly $\dim U \dim V$ (assuming $\dim W = 1$). Even so, we still must add in a few extra terms to insure linear dependence, but they are small in number. Schönhage points out that in this situation it is known that neither of the f, g admits an approximate algorithm better than the standard algorithm. A more geometric understanding of this "trick" could lead to better upper bounds. What follows are two examples for matrix multiplication, the second of which follows the above scheme.

Example 8.1.1 (Bini et al.)**.** An approximate algorithm for multiplying 2×2 matrices where the first matrix has a zero in the $(2, 2)$ slot is presented in [9]. In what follows we show how the algorithm corresponds to a point of $\sigma_5(Seg(\mathbb{P}^2 \times \mathbb{P}^3 \times \mathbb{P}^3))$. (It is relatively simple to pass back and forth between the algorithms and the description of the limiting $\mathbb{P}^4$ that lies in $\sigma_5(Seg(\mathbb{P}^2 \times \mathbb{P}^3 \times \mathbb{P}^3))$ that the tensor lies on. But the description of the $\mathbb{P}^4$ shows the non-uniqueness of the algorithm and the salient geometric facts that are used more transparently.) In this case we have five points that are linearly dependent. In fact only four are needed; one can take any 5-th point in the span of the four and ignore it, as its derivatives are not needed for the algorithm. We take

$$x_1 = \alpha_2^1 \otimes \beta_2^1 \otimes c_2^1, \ x_2 = \alpha_1^2 \otimes \beta_1^1 \otimes c_1^1, \ x_3 = \alpha_2^1 \otimes \beta_2^1 \otimes (c_1^1 + c_2^1), \ x_4 = \alpha_1^2 \otimes (\beta_1^1 + \beta_2^1) \otimes c_1^1.$$

Note that all these points lie on a $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$. Because $\text{codim}(Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)) = 4$, we are assured there is a fifth point of $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$ in the span of these four. (A general $\mathbb{P}^3$ will intersect $Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1))$ in $\deg(Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)) = 6$ points.) Moreover, the 5-th point will not be in the span of any three of $x_1, \ldots, x_4$. Then taking

$$x_1' = \alpha_1^1 \otimes \beta_2^1 \otimes c_2^1 + \alpha_2^1 \otimes \beta_2^2 \otimes c_2^1 - \alpha_2^1 \otimes \beta_1^2 \otimes c_2^1,$$
$$x_2' = \alpha_1^1 \otimes \beta_1^1 \otimes c_1^1 + \alpha_1^2 \otimes \beta_1^1 \otimes c_1^2 - \alpha_1^2 \otimes \beta_1^1 \otimes c_2^1,$$
$$x_3' = \alpha_2^1 \otimes \beta_1^2 \otimes (c_1^1 + c_2^1), \ x_4' = \alpha_1^2 \otimes (\beta_1^1 + \beta_2^1) \otimes c_2^2,$$

our matrix multiplication operator M for the partially filled matrices is $M = x_1' + x_2' + x_3' + x_4'$. The fact that we didn't use any of the initial points is not surprising, as the derivatives can always be altered to incorporate the initial points.

A splitting of the computation is the key to the reduction here as well. Split the calculation of M into two pieces, the terms involving α_1^1 and the rest. Those terms involving α_1^1 can be accomplished using two multiplications, and the rest can be accomplished using six. We change notation slightly and write $x_j = a_j \otimes b_j \otimes c_j$ and $x_j' = a_j' \otimes b_j \otimes c_j + a_j \otimes b_j' \otimes c_j + a_j \otimes b_j \otimes c_j'$ as we did before we began this example. The elements of $B \otimes C$ appearing with α_1^1 each appear in the original x_1, x_2, so in order to have them appear in the final tensor we just need to take $a_1', a_2' = \alpha_1^1$. Now to have the terms involving α_2^1, α_1^2 appear in the final tensor, we need to differentiate the terms on the B and C factors. We can obtain two of these by setting $b_1' = \beta_2^2$ and $c_2' = c_1^2$. We can get the remaining terms using x_3' and x_4', but we must introduce an error, which can then be absorbed by modifying b_1' and c_2'.

The result is that $a'_1 = a'_2 = \alpha_1^1$, $b'_1 = \beta_2^2 - \beta_1^2$, $c'_2 = c_1^2 - c_2^2$, $b'_3 = \beta_1^2$, $c'_4 = c_2^2$, and all the other first derivatives are zero.

Remark 8.1.2. There is a similarity between this example and the algorithms using multiplicative complexity discussed in §14.1.1.

Example 8.1.3 (Schönhage). Consider matrix multiplication of 3×3 matrices where in the first matrix $\alpha_1^2 = \alpha_1^3 = 0$, in the second matrix $\beta_2^2 = \beta_3^2 = \beta_2^3 = \beta_3^3 = 0$, and thus in the third, $c_2^2 = c_3^2 = c_2^3 = c_3^3 = 0$. We again split the computation into terms involving α_1^1 and those that do not. (It might be useful to think of this multiplication as $B \times C \to A$ to make it look more symmetric.) Those that do not involve α_1^1 use six multiplications in the naïve algorithm and those involving α_1^1 use four.

As explained in §7.2, $\mathbb{P}^4 \cap (Seg(\mathbb{P}^2 \times \mathbb{P}^2))$ will generally consist of $6 = \deg(Seg(\mathbb{P}^2 \times \mathbb{P}^2))$ points. Now the principle described above is used. That is, the initial six terms contain the correct six monomials in the B, C factors for the terms without α_1^1, and the second derivatives of the A factor in these terms are used to provide the correct A terms; while the original A factor term is always α_1^1, and it is paired with the derivatives in the B, C factors of the original terms. In this example, the spaces in B, C where the two different pieces live are nearly disjoint, so we need to differentiate twice to be able to get both the B and C coefficients new (which is why we used second rather than first derivatives in the A-factor).

What is interesting about this example is that taking three such blockings, one can "cover" the space of 3×3 matrices and by adding them together obtain an approximate algorithm for $M_{3,3,3}$ using 21 multiplications.

8.2. **Finite group approach to upper bounds.** H. Cohn and C. Umans [23] have proposed a different approach to constructing algorithms for matrix multiplication using the discrete Fourier transform and the representation theory of finite groups.

Let G be a finite group and $\mathbb{C}[G]$ its group algebra. (See, e.g., [49] for definitions and properties of the group algebra.) The discrete Fourier transform (DFT) $D : \mathbb{C}[G] \to \mathbb{C}^{|G|}$ is an invertible linear map that actualizes Wedderburn's theorem that $\mathbb{C}[G] \simeq Mat_{d_1 \times d_1}(\mathbb{C}) \times \cdots \times Mat_{d_k \times d_k}(\mathbb{C})$, where G has k irreducible representations and the dimension (character) of the j-th is d_j. (See, e.g., [14] for an exposition.) Thus multiplication in the group ring is converted to multiplication of $d_1 \times d_1, \ldots, d_r \times d_r$ matrices.

The idea is that to multiply $Mat_{n \times m} \times Mat_{m \times p} \to Mat_{n \times p}$ one first bijectively maps bases of each of these three spaces into subsets of some finite group G. The subsets are themselves formed from three subsets, S_1, S_2, S_3, of cardinalities n, m, p which have a disjointness property, called the *triple product property* in [23]: if $s_1 s_2 s_3 = Id$, with $s_i \in S_i^{-1} S_i$, then each $s_i = Id$. Then the maps are to the three subsets $S_1^{-1} S_2$, $S_2^{-1} S_3$, $S_1^{-1} S_3$. The triple product property enables one to read off matrix multiplication from multiplication in the group ring. They then show, if ω is the exponent of matrix multiplication, that if one can find such a group and subsets, then

$$(nmp)^{\frac{\omega}{3}} \leq d^{\omega - 2} |G|$$

where d is the largest character of G. So one needs to find groups that are big enough to support triples satisfying the triple product property but as small as possible and with largest character as small as possible.

In [22] Cohn et al. give explicit examples that recover $\omega < 2.41$ and state several combinatorial and group theoretic conjectures that, if true, would imply $\omega = 2$.

9. DIMENSIONS OF SECANT VARIETIES OF SEGRE VARIETIES

The most basic invariant of an algebraic variety is its dimension. In this section we discuss the standard tool for computing dimensions of secant varieties of projective varieties and its application to secant varieties of Segre varieties. The results of this section are not used in the subsequent sections.

9.1. **Dimensions of secant varieties of Segre varieties and matrix multiplication.** Let A, B, C be vector spaces of dimensions $\mathbf{a}, \mathbf{b}, \mathbf{c}$. By subsection 3.4, the expected dimension of $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ is $r(\mathbf{a}-1+\mathbf{b}-1+\mathbf{c}-1)+r-1 = r(\mathbf{a}+\mathbf{b}+\mathbf{c}-2)-1$. The dimension of the ambient space is $\mathbf{abc}-1$, so we expect $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ to fill $\mathbb{P}(A \otimes B \otimes C)$ as soon as $r(\mathbf{a}+\mathbf{b}+\mathbf{c}-2)-1 \geq \mathbf{abc}-1$, that is,

$$(9.1.1) \qquad\qquad r \geq \frac{\mathbf{abc}}{\mathbf{a}+\mathbf{b}+\mathbf{c}-2}.$$

Note that in the case $\mathbf{a} = \mathbf{b} = \mathbf{c}$ equation (9.1.1) becomes $r \geq \mathbf{a}^3/(3\mathbf{a}-2) \simeq \mathbf{a}^2/3$. Taking $\mathbf{a} = n^2$, the right-hand side of (9.1.1) is roughly $n^4/3$, showing already that matrix multiplication is far from being a generic bilinear map, as even the standard algorithm gives $\mathbf{R}(M_{n,n,n}) \leq n^3$. (The actual typical X-rank cannot be smaller than the expected typical X-rank.) However for $n = 2$ we obtain $r \geq 64/10$, and thus $r = 7$ is expected to (and we will see below does) fill, so $M_{2,2,2}$ is generic in this sense.

9.2. **Terracini's Lemma and applications.** Recall the notations from the beginning of §3.3 and adopt the additional notation that for $Z \subset \mathbb{P}V$, $\hat{T}_{[z]}Z = T_z\hat{Z} \subset V$ is the embedded tangent space to $\hat{Z}$ at $z \in \hat{Z}$.

Lemma 9.2.1 (Terracini's Lemma (see, e.g., [21, 33, 60])). *If $[x] \in J(Y,Z)_{smooth}$ with $[x] = [y+z]$ such that $[y] \in Y_{smooth}, [z] \in Z_{smooth}$, then*

$$\hat{T}_{[x]}J(Y,Z) = \hat{T}_{[y]}Y + \hat{T}_{[z]}Z.$$

Thus, if $[p] = [x_1 + \cdots + x_r] \in \sigma_r(X)_{smooth}$ with $[x_j] \in X_{smooth}$, then

$$\hat{T}_{[p]}\sigma_r(X) = \hat{T}_{[x_1]}X + \cdots + \hat{T}_{[x_r]}X.$$

Terracini's Lemma implies that for a variety $X \subset \mathbb{P}V$, if any given $\sigma_r(X)$ is of the expected dimension $r \dim X + r - 1$, then all $\sigma_{r'}(X)$ for $r' < r$ are of the expected dimension.

Thus if $\dim \mathbb{P}V = p(n-1) + p - 1$ and one shows $\sigma_p(X) = \mathbb{P}V$, then all secant varieties of X are of the expected dimension.

The following trick occurs frequently in the literature: let $Y_1, \ldots, Y_p \subset X$, so $\hat{T}_{y_1}Y_1 + \cdots + \hat{T}_{y_p}Y_p \subseteq \hat{T}_{[y_1+\cdots+y_p]}\sigma_p(X)$. If one can show $\hat{T}_{y_1}Y_1 + \cdots + \hat{T}_{y_p}Y_p = V$, one has shown $\sigma_p(X) = \mathbb{P}V$. Lickteig and Strassen show that for $X = Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$, remarkably just taking the Y_i to be the Segre itself at most three times and taking other than Y_i to be linear spaces in it is sufficient for certain cases:

Lemma 9.2.2 (Lickteig [44]). *Adopt the notation $\mathbb{P}A_i = \mathbb{P}(A \otimes b_i \otimes c_i) \subset Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$, $\mathbb{P}B_j = \mathbb{P}(a_j \otimes B \otimes c'_j) \subset Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$.*

(1) *We may choose points $a_1, \ldots, a_s \in A$, $b_1, \ldots, b_q \in B$, $c_1, \ldots, c_q, c'_1, \ldots, c'_s \in$
    C, such that*

$$\hat{J}(\mathbb{P}A_1, \ldots, \mathbb{P}A_q, \mathbb{P}B_1, \ldots, \mathbb{P}B_s) = A \otimes B \otimes C$$

*when $q = \mathbf{b}l_1$, $s = \mathbf{a}l_2$ and $\mathbf{c} = l_1 + l_2$ and when $\mathbf{a} = \mathbf{b} = 2$, $q + s = 2\mathbf{c}$,
$s, q \geq 2$.*

(2) *We may choose points $a_1, \ldots, a_s \in A$, $b_1, \ldots, b_q \in B$, $c_1, \ldots, c_q, c'_1, \ldots, c'_s \in$
    C, such that*

$$\hat{J}(\sigma_2(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)), \mathbb{P}A_1, \ldots, \mathbb{P}A_q, \mathbb{P}B_1, \ldots, \mathbb{P}B_s) = A \otimes B \otimes C$$

when $q + s + 2 = \mathbf{c}$ and $\mathbf{a} = \mathbf{b} = 2$.

(3) *We may choose points $a_1, \ldots, a_s \in A$, $b_1, \ldots, b_q \in B$, $c_1, \ldots, c_q, c'_1, \ldots, c'_s \in$
    C, such that*

$$\hat{J}(\sigma_3(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)), \mathbb{P}A_1, \ldots, \mathbb{P}A_q, \mathbb{P}B_1, \ldots, \mathbb{P}B_s) = A \otimes B \otimes C$$

when $q = s = \mathbf{c} - 2 \geq 2$ and $\mathbf{a} = \mathbf{b} = 3$.

Using Lemma 9.2.2, Lickteig shows

Theorem 9.2.3 (Lickteig [44]). *$\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$ is non-degenerate for all
r whenever $\mathbf{a} \leq \mathbf{b} \leq \mathbf{c}$, $\mathbf{b}, \mathbf{c}$ are even and $\mathbf{abc}/(\mathbf{a} + \mathbf{b} + \mathbf{c} - 2)$ is an integer.*

With a little more work Lickteig obtains Theorem 3.8.1.

A classical technique for showing a secant variety of any variety $X \subset \mathbb{P}V$ is
degenerate is to find a variety $Y \subset \mathbb{P}V$, with $X \subset Y$, with $\sigma_k(Y)$ very degenerate.
Then, if X "catches up", that is, if there exists r such that $\sigma_r(X) = \sigma_r(Y)$, then
$\sigma_t(X) = \sigma_t(Y)$ for all $t > r$ as well. (See, e.g., [19] for a recent application.) To see
this, first note that for $u < r$, $\sigma_r(X) = J(\sigma_{r-u}(X), \sigma_u(X)) \subseteq J(\sigma_{r-u}(Y), \sigma_u(X)) \subseteq$
$\sigma_r(Y)$, so $\sigma_r(Y) = J(\sigma_{r-u}(Y), \sigma_u(X))$. Now write $t = mr + u$,

$$\begin{aligned}
\sigma_t(X) &= J(\sigma_{mr}(X), \sigma_u(X)) \\
&= J(\sigma_{(m-1)r}(Y), \sigma_u(Y), J(\sigma_{r-u}(Y), \sigma_u(X))) \\
&= \sigma_{mr+u}(Y).
\end{aligned}$$

In particular, since $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B))$ is very degenerate, if we have a three
factor case that is "unbalanced" in the sense that one space is much smaller than the
others, it can catch up to a corresponding two factor case. For example $\sigma_2(Seg(\mathbb{P}^1 \times
\mathbb{P}^1 \times \mathbb{P}^3)) = \sigma_2(Seg(\mathbb{P}(\mathbb{C}^2 \otimes \mathbb{C}^2) \times \mathbb{P}^3))$. Note that when this catching up occurs, if one
knows the ideal of the *a priori* larger variety, one can obtain the ideals of the secant
varieties of the smaller variety. Other uses of auxiliary varieties to understand the
secant varieties of Segre varieties are discussed in §12.

In the past few years there have been several papers on the dimensions of secant
varieties of Segre varieties, e.g., [17, 16, 15, 19, 1]. These papers use methods
similar to those of Strassen and Lickteig, but the language is more geometric (fat
points, degeneration arguments). Some explanation of the relation between the
algebreo-geometric and tensor language is given in [1].

With such steady progress, it seems reasonable to hope for a complete solution
for the secant defectivity of Segre varieties in the near future, at least in the three
factor case.

10. INVARIANT DESCRIPTION OF STRASSEN'S EQUATIONS AND GENERALIZATIONS

In this section we first rephrase Strassen's equations as the image of a $GL(A) \times GL(B) \times GL(C)$-equivariant map. We use this rephrasing to describe how to explicitly write a basis of his equations in a "good" basis and to generalize his equations. To ease the reader into this perspective, we begin with a familiar case.

10.1. **Warm up: Invariant description of generators of the ideal of** $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B))$**.** The set of $\mathbf{a} \times \mathbf{b}$ matrices of rank at most r is the zero set of the $(r + 1) \times (r + 1)$ minors, in fact these minors generate the ideal of $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B))$. To understand this space of equations invariantly, we begin with 2×2 minors. Choose bases $\{a_i\}$ of A, $\{b_s\}$ of B, and write our resulting matrix representing a point of $A \otimes B$ as $x = (x_s^i)$; that is, $x = \sum_{i,s} x_s^i a_i \otimes b_s$. Consider the minor $P_{ij,st} := x_s^i x_t^j - x_t^i x_s^j \in S^2(A \otimes B)^*$. Note that $P_{ij,st} = -P_{ji,st}$ and $P_{ij,st} = -P_{ij,ts}$. Hence $P_{ij,st} \in \Lambda^2 A^* \otimes \Lambda^2 B^*$, and in fact we have an injective map

$$\Lambda^2 A^* \otimes \Lambda^2 B^* \to S^2(A \otimes B)^*$$

whose image is the space of 2×2 minors. By the same reasoning, there is an injective map $\Lambda^d A^* \otimes \Lambda^d B^* \to S^d(A \otimes B)^*$ with image the $d \times d$ minors. We conclude:

The ideal of $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B))$ is generated by $\Lambda^{r+1}A^* \otimes \Lambda^{r+1}B^* \subset S^{r+1}(A \otimes B)^*$.

We will see in §11 that $\Lambda^{r+1}A^* \otimes \Lambda^{r+1}B^*$ is an irreducible $GL(A) \times GL(B)$-submodule of $S^{r+1}(A \otimes B)^*$. A more precise goal than "finding equations for secant varieties of Segre varieties" is to find the irreducible modules generating their ideals. When we discuss finding invariant descriptions of sets of equations, ultimately we will mean as modules, but in the interim we can simply mean "without reference to choices of bases", such as we have done here for the $(r + 1) \times (r + 1)$ minors.

10.2. **Strassen's equations reconsidered.** In order to understand Strassen's equations invariantly, we would like to get rid of the choices of $\alpha, \alpha^1, \alpha^2$, and the requirement that α is such that $T(\alpha)$ is invertible in Theorem 6.1.1. In what follows we will deal with tensors instead of endomorphisms, composition of endomorphisms will correspond to contractions of tensors, and the commutator of two endomorphisms will correspond to contracting a tensor in two different ways and taking the difference of the two results. Note that matrix multiplication $M : (U^* \otimes V) \times (V^* \otimes W) \to U^* \otimes W$ itself is simply the contraction of V with V^*.

A linear map $f : V \to W$ induces linear maps $f^{\wedge k} : \Lambda^k V \to \Lambda^k W$. If $\dim V = \dim W = n$, then, letting $\det(f) := f^{\wedge n}$, we have $f^{\wedge n-1} = f^{-1} \otimes \det(f)$, which follows from the canonical identification $\Lambda^{n-1}V \simeq V^* \otimes \Lambda^n V$.

The punchline of this section is:

Strassen's equations correspond to the image of the composition of the inclusion

$$\Lambda^2 A \otimes S^{\mathbf{b}-1} A \otimes \Lambda^{\mathbf{b}} B \otimes B \otimes \Lambda^{\mathbf{b}} C \otimes C \to (A \otimes B \otimes C)^{\mathbf{b}+1}$$

with the projection

$$(A \otimes B \otimes C)^{\mathbf{b}+1} \to S^{\mathbf{b}+1}(A \otimes B \otimes C).$$

We remark that the composition of these two maps is not injective. In §11.2 we describe the image precisely. We emphasize this perspective because it leads to vast generalizations of Strassen's equations discussed in §10.4.

Given $T \in A \otimes B \otimes C$, recall our notation $T_\alpha \in B \otimes C$ from §6.1. We have $T_\alpha^{\wedge \mathbf{b}-1} \in \Lambda^{\mathbf{b}-1} B \otimes \Lambda^{\mathbf{b}-1} C = \Lambda^{\mathbf{b}-1} B \otimes C^* \otimes \Lambda^{\mathbf{b}} C$. We may wedge the $\Lambda^{\mathbf{b}-1} B$ and B factors in

$$T_\alpha^{\wedge \mathbf{b}-1} \otimes T_{\alpha^j} \in \Lambda^{\mathbf{b}-1} B \otimes C^* \otimes \Lambda^{\mathbf{b}} C \otimes B \otimes C$$

together to obtain an element

$$T_{\alpha^j}^\alpha \in \Lambda^{\mathbf{b}} B \otimes C^* \otimes \Lambda^{\mathbf{b}} C \otimes C = C^* \otimes C \otimes \Lambda^{\mathbf{b}} B \otimes \Lambda^{\mathbf{b}} C.$$

That is, up to tensoring with a one-dimensional vector space, we have linear maps $C \to C$ and can now take their commutators. Consider

$$T_{\alpha^1}^\alpha \otimes T_{\alpha^2}^\alpha \in (\Lambda^{\mathbf{b}} B \otimes C^* \otimes \Lambda^{\mathbf{b}} C \otimes C)^{\otimes 2} = C^* \otimes C \otimes C^* \otimes C \otimes (\Lambda^{\mathbf{b}} B)^{\otimes 2} \otimes (\Lambda^{\mathbf{b}} C)^{\otimes 2}$$

and contract a copy of C from $T_{\alpha^1}^\alpha$ with a copy of C^* from $T_{\alpha^2}^\alpha$ to obtain an element of $C^* \otimes C \otimes (\Lambda^{\mathbf{b}} B)^{\otimes 2} \otimes (\Lambda^{\mathbf{b}} C)^{\otimes 2}$. This contraction corresponds to the matrix multiplication of $T_{\alpha^1}^\alpha$ with $T_{\alpha^2}^\alpha$. Reversing the roles of $T_{\alpha^1}^\alpha, T_{\alpha^2}^\alpha$ reverses the order of the matrix multiplication. Thus the difference of these two contractions is

$$[T_{\alpha^1}^\alpha, T_{\alpha^2}^\alpha] \in C^* \otimes C \otimes (\Lambda^{\mathbf{b}} B)^{\otimes 2} \otimes (\Lambda^{\mathbf{b}} C)^{\otimes 2},$$

and Strassen's theorem states that the rank of $[T_{\alpha^1}^\alpha, T_{\alpha^2}^\alpha]$ is at most $2(r - \mathbf{b})$.

With a little more care, one obtains a lower degree tensor; see [39] for details.

Remark 10.2.1. Strassen's equations were rediscovered in [6], guided by the geometry of phylogenetic trees, which also enabled a nice presentation of them. The recent preprint [45] gives an even simpler description of Strassen's equations. Unfortunately the generalizations discussed below are not yet evident from either of these presentations.

10.3. **Explicit polynomials in bases.** Here are polynomials corresponding to Strassen's commutator being of rank at most w: Let $\alpha^1, \alpha^2, \alpha^3$ be a basis of A^*, $\beta_1, \ldots, \beta_{\mathbf{b}}, \xi_1, \ldots, \xi_{\mathbf{b}}$ bases of B^*, C^*. Consider the element

$$P = \alpha^2 \wedge \alpha^3 \otimes (\alpha^1)^{\mathbf{b}-1} \otimes \beta_1 \wedge \cdots \wedge \beta_{\mathbf{b}} \otimes \beta_s \otimes \xi_1 \wedge \cdots \wedge \xi_{\mathbf{b}} \otimes \xi_t.$$

This expands to (ignoring scalars)

$$(\alpha_2 \otimes \alpha_3 - \alpha_3 \otimes \alpha_2) \otimes (\alpha_1)^{\mathbf{b}-1} \otimes (\sum_j (-1)^{j+1} \beta_{\hat{j}} \otimes \beta_j \otimes \beta_s) \otimes (\sum_k (-1)^{k+1} \xi_{\hat{k}} \otimes \xi_k \otimes \beta_t)$$

$$= (-1)^{j+k} [((\alpha_1)^{b-1} \otimes \beta_{\hat{j}} \otimes \xi_{\hat{k}}) \otimes (\alpha_2 \otimes \beta_j \otimes \xi_t) \otimes (\alpha_3 \otimes \beta_s \otimes \xi_k)$$

$$- ((\alpha_1)^{b-1} \otimes \beta_{\hat{j}} \otimes \xi_{\hat{k}}) \otimes (\alpha_3 \otimes \beta_j \otimes \xi_t) \otimes (\alpha_2 \otimes \beta_s \otimes \xi_k)].$$

A hat over an index indicates the wedge product of all vectors in that index range except the hatted one. If we choose dual bases for A, B, C and write $T = a_1 \otimes X + a_2 \otimes Y + a_3 \otimes Z$ where the a_j are dual to the α_j and X, Y, Z are represented as $b \times b$ matrices with respect to the dual bases of B, C, then let $P(T)$ be the matrix with

$$P(T)_t^s = \sum_{j,k} (-1)^{j+k} (\det X_{\hat{k}}^{\hat{j}})(Y_t^j Z_k^s - Y_k^s Z_t^j)$$

where $X_{\hat{k}}^{\hat{j}}$ is X with its j-th row and k-th column removed. Strassen's commutator has rank at most w if and only if all the $(w+1) \times (w+1)$ minors of $P(T)$ are zero. It turns out that when one takes the determinant of $P(T)$, one gets a reducible polynomial that is divisible by the determinant of X, so, e.g., when $b = 3$ one obtains an irreducible polynomial of degree 9 (as opposed to 12).

10.4. **Generalizations of Strassen's conditions.** The key point in the discussion above was that contracting T in two different ways yielded tensors that commute if T is in $\sigma_r(Seg(\mathbb{P}A^* \times \mathbb{P}B^* \times \mathbb{P}C^*))$. Consider for s, t such that $s + t \leq \mathbf{b}$ and $\alpha, \alpha_j \in A^*$ the tensors

$$T_{\alpha_j}^{\wedge s} \in \Lambda^s B \otimes \Lambda^s C, \ T_\alpha^{\wedge t} \in \Lambda^t B \otimes \Lambda^t C$$

(in §10.2 we had $s = 1, t = \mathbf{b} - 1$). We contract $T_\alpha^{\wedge t} \otimes T_{\alpha_1}^{\wedge s} \otimes T_{\alpha_2}^{\wedge s}$ to obtain elements of $\Lambda^{s+t} B \otimes \Lambda^{s+t} C \otimes \Lambda^s B \otimes \Lambda^s C$ in two different ways; call these contractions $\psi_{\alpha,\alpha_1,\alpha_2}^{s,t}(T)$ and $\psi_{\alpha,\alpha_2,\alpha_1}^{s,t}(T)$.

Now say $\mathbf{R}(T) = r$, so we may write $T = a_1 \otimes b_1 \otimes c_1 + \cdots + a_r \otimes b_r \otimes c_r$ for elements $a_i \in A$, $b_i \in B$, $c_i \in C$. We have

$$\psi_{\alpha,\alpha_1,\alpha_2}^{s,t}(T) = \sum_{|I|=s,|J|=t,|K|=s} \langle a_I, \alpha_1 \rangle \langle a_J, \alpha \rangle \langle a_K, \alpha_2 \rangle (b_{I+J} \otimes b_K) \otimes (c_I \otimes c_{J+K}),$$

where $a_I = a_{i_1} \wedge \cdots \wedge a_{i_s} \in \Lambda^s A$, $\langle A_I, \alpha \rangle \in \Lambda^{s-1} A$ and $a_{I+J} = a_I \wedge a_J$, etc. For this to be non-zero, we need I and J to be disjoint subsets of $\{1, \ldots, r\}$. Similarly, J and K must be disjoint. If $s + t = r$, this implies $J = K$. In summary:

Theorem 10.4.1 ([39]). *For* $T \in \sigma_{s+t}(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$, *for all* $\alpha, \alpha_1, \alpha_2 \in A^*$

$$\psi_{\alpha,\alpha^1,\alpha^2}^{s,t}(T) - \psi_{\alpha,\alpha^2,\alpha^1}^{s,t}(T) = 0.$$

We have the bilinear map

$$(\Lambda^2(S^s A) \otimes S^t A)^* \times (A \otimes B \otimes C)^{\otimes 2s+t} \to \Lambda^{s+t} B \otimes \Lambda^{s+t} C \otimes \Lambda^s B \otimes \Lambda^s C$$

whose image is $\psi_{\alpha,\alpha^1,\alpha^2}^{s,t}(T) - \psi_{\alpha,\alpha^2,\alpha^1}^{s,t}(T)$. We rewrite it as a polynomial map

$$\Psi^{s,t} : A \otimes B \otimes C \to (\Lambda^2(S^s A) \otimes S^t A) \otimes \Lambda^{s+t} B \otimes \Lambda^{s+t} C \otimes \Lambda^s B \otimes \Lambda^s C.$$

So just as with Strassen's equations, we no longer need to make choices of elements of A^*.

The only catch is we don't know whether or not $\Psi^{s,t}$ is identically zero. In [39] we show many of the $\Psi^{s,t}$ are indeed non-zero and give independent subspaces (in fact independent $GL(A) \times GL(B) \times GL(C)$-submodules; see §11) of the ideal of $\sigma_{s+t}(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C))$.

In [39], Corollary 5.6, using the above methods, we show that set-theoretic defining equations for $\sigma_4(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3))$, the case of interest for phylogenetic invariants, could be explicitly determined if one had a complete set of defining equations for $\sigma_4(Seg(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^3))$.

11. Representation theory and equations for secant varieties of Segre varieties

As mentioned in the introduction, the most important tool for studying varieties invariant under a group action is representation theory. In this section we develop the necessary representation theory for studying secant varieties of Segre varieties. The theory developed in this section is also what is needed in the more general study of algebraic statistical models. We first describe how to decompose the space of polynomials on $A_1 \otimes \cdots \otimes A_n$ into subspaces invariant under the action of the group of changes of bases in the vector spaces $GL(A_1) \times \cdots \times GL(A_n)$. We then describe Strassen's equations from this perspective and how to find preferred polynomials in each irreducible submodule. We also describe two notions, *inheritance*

and *prolongation*, which facilitate our study. Once one has an explicit description of a space of polynomials as modules, it is algorithmic to write down an explicit basis of the module as we did in §10.1. See [37, 40] for more details.

11.1. Polynomials come in modules. Since $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ is invariant under the action of $G = GL(A_1) \times \cdots \times GL(A_n)$ acting on $A_1 \otimes \cdots \otimes A_n = V$, its ideal, which is a subset of the module $\oplus_d S^d V^*$, must be as well. Thus we should study the equations of $\sigma_r(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$ as G-modules.

Given any G-module W, the first thing to do when studying W is to try to decompose it into *isotypic* components (which is always possible when G is reductive, as is our situation). That is, one can decompose W into a direct sum of irreducible modules, but this is not canonical. The isotypic decomposition (which is canonical) is obtained from the decomposition into irreducible submodules by grouping together all copies of isomorphic irreducible submodules.

To decompose $S^d V^*$ into G-isotypic components we use the *Shur-Weyl duality* between representations of the symmetric group on d letters $\mathfrak{S}_d$ and the representations of the general linear group $GL(W)$. Both groups act on $W^{\otimes d}$: for $A \in GL(W)$ and $\sigma \in \mathfrak{S}_d$ we respectively have

$$A.(v_1 \otimes \cdots \otimes v_d) = (A.v_1) \otimes \cdots \otimes (A.v_d)$$
$$\sigma.(v_1 \otimes \cdots \otimes v_d) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)}.$$

Schur-Weyl duality is the statement that each group is the commuting subgroup of the other; that is

$$\mathfrak{S}_d = \{g \in GL(W^{\otimes d}) \mid g.A.(v_1 \otimes \cdots \otimes v_d) = A.g.(v_1 \otimes \cdots \otimes v_d)$$
$$\forall A \in GL(W), \forall v_1, \ldots, v_d \in W\}$$

and

$$GL(W) = \{g \in GL(W^{\otimes d}) \mid g.\sigma.(v_1 \otimes \cdots \otimes v_d) = \sigma.g.(v_1 \otimes \cdots \otimes v_d)$$
$$\forall \sigma \in \mathfrak{S}_d, \forall v_1, \ldots, v_d \in W\}.$$

Thus we can use the action of $\mathfrak{S}_d$ to obtain projection operators $W^{\otimes d} \to W^{\otimes d}$, whose images are necessarily $GL(W)$-submodules. Moreover, the duality assures us that all $GL(W)$-submodules may be obtained this way. For example

$$S^d W = \{T \in W^{\otimes d} \mid \sigma(T) = T \ \forall \sigma \in \mathfrak{S}_d\}$$
$$= \operatorname{Im} \pi_S : W^{\otimes d} \to W^{\otimes d} \text{ where } \pi_S(w_1 \otimes \cdots \otimes w_d) = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} w_{\sigma(1)} \otimes \cdots \otimes w_{\sigma(d)}$$

$$\Lambda^d W = \{T \in W^{\otimes d} \mid \sigma(T) = sgn(\sigma)T \ \forall \sigma \in \mathfrak{S}_d\}$$
$$= \operatorname{Im} \pi_\Lambda : W^{\otimes d} \to W^{\otimes d}$$
$$\text{where } \pi_\Lambda(w_1 \otimes \cdots \otimes w_d) = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} sgn(\sigma) w_{\sigma(1)} \otimes \cdots \otimes w_{\sigma(d)}.$$

Let $\pi = (p_1, \ldots, p_f)$ be a partition of d; that is, $p_1 \geq \cdots \geq p_f$ and $p_1 + \cdots + p_f = d$. We use the notations $|\pi| = d$ and $l(\pi) = f$.

The irreducible representations of $\mathfrak{S}_d$ are indexed by partitions of d; we let $[\pi]$ denote the module induced by π. Here $[\pi]$ may be obtained by a choice of Young symmetrizer c_λ corresponding to a choice of a Young tableau associated to π and

applying the projection operator c_λ to the group algebra $\mathbb{C}[\mathfrak{S}_d]$. (For proofs and more details regarding these assertions, see, e.g., [27], Chapter 4.)

Define $S_\pi W := \mathrm{Hom}_{\mathfrak{S}_d}([\pi], W^{\otimes d})$, which is an irreducible $GL(W)$-module. The $GL(W)$-isotypic decomposition of $W^{\otimes d}$ is $W^{\otimes d} = \oplus_{|\pi|=d}[\pi] \otimes S_\pi W$. The first factor is a trivial $GL(W)$-module, so it serves only to tell us the multiplicity of the second, which is $\dim[\pi]$.

We now return to the space we are interested in, $V = A_1 \otimes \cdots \otimes A_n$ as a $G = GL(A_1) \times \cdots \times GL(A_n)$-module:

Proposition 11.1.1 ([37]). *The $G = GL(A_1) \times \cdots \times GL(A_n)$ isotypic decomposition of $S^d(A_1 \otimes \cdots \otimes A_n)$ is*

$$S^d(A_1 \otimes \cdots \otimes A_n) = \bigoplus_{|\pi_1|=\cdots=|\pi_k|=d} ([\pi_1] \otimes \cdots \otimes [\pi_n])^{\mathfrak{S}_d} \otimes S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_k} A_k,$$

where $([\pi_1] \otimes \cdots \otimes [\pi_k])^{\mathfrak{S}_d}$ denotes the space of $\mathfrak{S}_d$-invariants (that is, instances of the trivial representation of $\mathfrak{S}_d$) in $[\pi_1] \otimes \cdots \otimes [\pi_n]$.

The $([\pi_1] \otimes \cdots \otimes [\pi_n])^{\mathfrak{S}_d}$ factor in the tensor product just serves to tell us the multiplicity of $S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_k} A_k$ via its dimension.

Proof. We need to decompose $S^d(A_1 \otimes \cdots \otimes A_n)$ as a $G = GL(A_1) \times \cdots \times GL(A_n)$-module. We have

$$(A_1 \otimes \cdots \otimes A_n)^{\otimes d} = \bigoplus_{|\pi_j|=d} ([\pi_1] \otimes \cdots \otimes [\pi_n]) \otimes (S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_n} A_n).$$

But $S^d(A_1 \otimes \cdots \otimes A_n) \subset (A_1 \otimes \cdots \otimes A_n)^{\otimes d}$ is the set of elements invariant under the action of $\mathfrak{S}_d$. (Here $\mathfrak{S}_d$ acts only on the $[\pi_j]$; it leaves the $S_{\pi_j} A_j$'s invariant.) $\square$

Now we need a way to calculate $\dim([\pi_1] \otimes \cdots \otimes [\pi_k])^{\mathfrak{S}_d}$. This can be done using *characters* in low degrees (degrees as high as your computer is willing to tolerate). The key point is

$$\dim([\pi_1] \otimes \cdots \otimes [\pi_n])^{\mathfrak{S}_d} = \frac{1}{d!} \sum_{\sigma \in \mathfrak{S}_d} \chi_{\pi_1}(\sigma) \cdots \chi_{\pi_n}(\sigma)$$

where $\chi_{\pi_j} : \mathfrak{S}_d \to \mathbb{C}$ is the *character* of $[\pi_j]$ (see, e.g., [27, 49]). For any given d, one can compute these dimensions, but there is no known closed form formula for them when $n > 2$.

Obtaining the above decomposition is essential when dealing with explicit equations. For example, Strassen has *a priori* three sets of equations for $\sigma_3(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)$. Are they redundant or not? By examining these equations as modules we find that they are.

11.2. **Strassen's equations as modules.** Recall from §10 that Strassen's equations for $\sigma_r(Seg(\mathbb{P}^2 \times \mathbb{P}^{b-1} \times \mathbb{P}^{b-1}))$ in degree $b+1$ are obtained by composing the inclusion

$$\Lambda^2 A \otimes S^{b-1} A \otimes \Lambda^b B \otimes B \otimes C \otimes \Lambda^b C \to (A \otimes B \otimes C)^{b+1}$$

with the projection

$$(A \otimes B \otimes C)^{b+1} \to S^{b+1}(A \otimes B \otimes C).$$

Now $\Lambda^2 A \otimes S^{b-1} A \otimes \Lambda^b B \otimes B \otimes C \otimes \Lambda^b C$ is not an irreducible module. Since the maps are G-equivariant, by Shur's lemma the image is a direct sum of irreducible

submodules. We need to determine which modules in $\Lambda^2 A \otimes S^{b-1} A \otimes \Lambda^b B \otimes B \otimes C$ $\otimes \Lambda^b C$ map non-trivially into $S^{b+1}(A \otimes B \otimes C)$.

Since here $b = \dim B = \dim C$, we have, using a very special case of the Littlewood-Richardson rule (see, e.g., [27], Chapter 6),

$$(\Lambda^2 A \otimes S^{b-1} A) \otimes (\Lambda^b B \otimes B) \otimes (C \otimes \Lambda^b C) = (S_{b,1} A \oplus S_{b-1,1,1} A) \otimes \Lambda_{b,1} B \otimes \Lambda_{b,1} C$$

(where we use the notation $\Lambda_{b,1} B = S_{2,1,\dots,1} B$), so there are two possible modules. Were the first in the image, then one would be able to get equations in the case $\dim A = 2$, but $\sigma_3(\mathbb{P}^1 \times \mathbb{P}^2 \times \mathbb{P}^2) = \mathbb{P}(A \otimes B \otimes C)$, so only the second can occur (and it is easy to check that it does). We conclude:

Proposition 11.2.1 ([39]). *Strassen's equations for $\sigma_b(\mathbb{P}^2 \times \mathbb{P}^{b-1} \times \mathbb{P}^{b-1})$ expressed as a module is*

$$S_{b-1,1,1} \mathbb{C}^3 \otimes \Lambda_{b,1} \mathbb{C}^b \otimes \Lambda_{b,1} \mathbb{C}^b;$$

in particular it is an irreducible module.

When $b = 3$, we obtain $S_{211} A \otimes S_{211} B \otimes S_{211} C$, which occurs with multiplicity one in $S^4(A \otimes B \otimes C)$. Thus, despite the apparently different role of A from B and C, in this case - and *only* in this case - exchanging the role of A with B or C yields the same space of equations.

11.3. **Highest weight vectors.** When we study modules of polynomials, it will be convenient to have a "best" polynomial in the module. For example, since an irreducible G-module in $S^d V^*$ is either entirely in or out of the ideal of a G-variety $Z \subset \mathbb{P}V$, it is sufficient to check just a single polynomial in the module. In general, this "best polynomial" is provided by a choice of highest weight vector. We explain how to obtain such vectors when $G = GL(A_1) \times \cdots \times GL(A_n)$.

Fix a basis $e_1, \dots, e_n$ of a vector space V. Let W be an irreducible $GL(V)$-module occurring in $V^{\otimes d}$ for some d. We say $w \in W$ is a *highest weight vector* for W if $\rho(g).[w] = [w]$ for all upper triangular matrices $g \in GL(V)$. (It makes sense to discuss matrices because we have fixed a basis of V.) Highest weight vectors are in some sense the simplest vectors occurring in a module. (More precisely, after having fixed bases, the simplest vectors correspond to the vectors in the orbit of the highest weight vector under the action of the Weyl group.) For example, when $W = S^d V$, $(e_1)^d$ is a highest weight vector. For $W = \Lambda^d V$, $e_1 \wedge e_2 \wedge \cdots \wedge e_d$ is a highest weight vector. In general the highest weight vector of an irreducible module will not correspond to a decomposable tensor. In $c_\pi V^{\otimes d}$ ($\simeq S_\pi V$), the highest weight vector is

$$c_\pi(e_1^{\otimes p_1} \otimes e_2^{\otimes p_2} \otimes \cdots \otimes e_d^{\otimes p_d})$$

where $\pi = (p_1, \dots, p_d)$ and we allow the last few p_j to be zero in order to have a uniform expression.

In [37] we give explicit algorithms for writing down highest weight vectors of submodules of $S^d(A_1 \otimes \cdots \otimes A_n)$.

An important observation for the next section is if $v \in A^{\otimes d}$ is a highest weight vector for a submodule corresponding to a partition π and $a_1, \dots, a_n$ is a basis of A, v may be expressed using only the vectors $a_1, \dots, a_{l(\pi)}$.

11.4. **Inheritance.** By examining equations grouped into modules, the dimensions of the vector spaces involved come into play only when verifying that the dimension is large enough to support a given module. For example, let $A_1, \ldots, A_n$ be vector spaces and let $\mathbf{a}_j = \dim A_j$. We have:

Proposition 11.4.1 ([39]). *If a copy of*

$$S_{\pi_1} A_1 \otimes \cdots S_{\pi_n} A_n$$

occurs in

$$I_d(\sigma_r(Seg(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_n^*))),$$

then for all vector spaces $A_j' \supseteq A_j$, the corresponding copy of

$$S_{\pi_1} A_1' \otimes \cdots \otimes S_{\pi_n} A_n'$$

occurs in

$$I_d(\sigma_r(Seg(\mathbb{P}A_1'^* \times \cdots \times \mathbb{P}A_n'^*))).$$

Moreover, a module $S_{\pi_1} A_1' \otimes \cdots \otimes S_{\pi_n} A_n'$ where the length of each π_j is at most $\mathbf{a}_j$ is in $I_d(\sigma_r(Seg(\mathbb{P}A_1'^* \times \cdots \times \mathbb{P}A_n'^*)))$ if and only if the corresponding module is in $I_d(\sigma_r(Seg(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_n^*)))$.

Our notation is such that given a variety $Z \subset \mathbb{P}V^*$, $I(Z) \subset S^\bullet V$ denotes its ideal and $I_d(Z) = I(Z) \cap S^d V$.

Proof. A module is in the ideal if and only if its highest weight vector is. Choose ordered bases for A_j' such that the first $\mathbf{a}_j$ basis vectors form a basis of A_j. Then any highest weight vector for $S_{\pi_1} A_1' \otimes \cdots \otimes S_{\pi_n} A_n'$ is also a highest weight vector for $S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_n} A_n$ as long as $l(\pi_j) \leq \mathbf{a}_j$. $\square$

Thus a copy of a module $S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_n} A_n$ will be in $I(\sigma_r(Seg(\mathbb{P}^{r-1} \times \cdots \times \mathbb{P}^{r-1})))$ if and only if the corresponding copy of the module $S_{\pi_1} \mathbb{C}^{l(\pi_1)} \otimes \cdots \otimes S_{\pi_n} \mathbb{C}^{l(\pi_n)}$ is in the ideal of $\sigma_r(Seg(\mathbb{P}^{l(\pi_1)-1} \times \cdots \times \mathbb{P}^{l(\pi_n)-1}))$.

It is straightforward to determine $I_3(\sigma_2(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)))$ as a module:

Theorem 11.4.2 ([37], Thm. 4.7). *The space of cubics vanishing on $\sigma_2(Seg(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_k^*))$ is*

$$I_3(\sigma_2(Seg(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_k^*))) = \bigoplus_{\substack{I+J+L=\{1,\ldots,k\}, \\ j=|J|>1,\, |L|>0}} \frac{2^{j-1} - (-1)^{j-1}}{3} S_3 A_I \otimes S_{21} A_J \otimes S_{111} A_L$$

$$\oplus \bigoplus_{\substack{I+J=\{1,\ldots,k\}, \\ j=|J|>3}} \left(\frac{2^{j-1} - (-1)^{j-1}}{3} - 1\right) S_3 A_I \otimes S_{21} A_J \oplus \bigoplus_{\substack{I+L=\{1,\ldots,k\}, \\ |L|>0\,even}} S_3 A_I \otimes S_{111} A_L.$$

11.5. **Prolongation.** For $A \subset S^k V$ define

$$A^{(p)} = (A \otimes S^p V) \cap S^{p+k} V,$$

the *p-th prolongation of A.* Let $\mathrm{Zeros}(A) = \{[v] \in \mathbb{P}V^* \mid P(v) = 0 \; \forall P \in A\}$.

Ideals of secant varieties satisfy a *prolongation property*; in particular for secant varieties of intersections of quadrics we have:

Lemma 11.5.1 ([38]). *Let $A \subset S^2 V$ be a linear subspace with zero set $\mathrm{Zeros}(A) \subset \mathbb{P}V^*$. Then*

$$\mathrm{Zeros}(A^{(k-1)}) \supseteq \sigma_k(\mathrm{Zeros}(A)).$$

Moreover, if Zeros(A) *is not contained in a hyperplane, then for* $k \geq 2$, $I_k(\sigma_k(\text{Zeros}(A))) = 0$; *and if* $A = I_2(\text{Zeros}(A))$, *then* $I_{k+1}(\sigma_k(\text{Zeros}(A))) = A^{(k-1)}$.

Usually, for a variety $X \subset \mathbb{P}V$, $I(\sigma_k(X))$ is not generated in degree $k+1$. For example, consider the simplest intersection of quadrics, four points in $\mathbb{P}^2$. They generate six lines, so $\sigma(X)$ is a hypersurface of degree six.

Let G be a semi-simple Lie or algebraic group, let V_λ be the irreducible G-module of highest weight λ, and let $X = G/P \subset \mathbb{P}V_\lambda^*$ be a homogeneously embedded rational homogeneous variety, that is, the orbit of a highest weight line. ($X = Seg(\mathbb{P}A_1^* \otimes \cdots \otimes \mathbb{P}A_n^*) \subset \mathbb{P}(A_1 \otimes \cdots \otimes A_n)^* = \mathbb{P}V^*$ is one such.) By an unpublished theorem of Kostant, $I_2(X) = (V_{2\lambda}^*)^\perp \subset S^2 V_\lambda$ and $I(X)$ is generated in degree two. More generally, $I_k(X) = (V_{k\lambda}^*)^\perp \subset S^k V_\lambda$. We adopt the notation that if $V = V_\lambda$, we write $V^k = V_{k\lambda}$. In the Segre case,

$$V^k = S^k A_1 \otimes \cdots \otimes S^k A_n \subset S^k(A_1 \otimes \cdots \otimes A_n).$$

Proposition 11.5.2 ([37])**.** *Let* $X \subset \mathbb{P}V^*$ *be a variety not contained in a linear space. Then for all* $d > 0$, $I_d(\sigma_d(X)) = 0$.

If $X = G/P$ *is homogeneous, then* $I_{d+1}(\sigma_d(X))$ *is the kernel of the contraction map* $(V^2)^* \otimes S^{d+1}V \to S^{d-1}V$.

Examples illustrating Proposition 11.5.2 are given in [37]. Extensions and further applications of prolongations are given in [51].

12. Auxiliary varieties

A simple observation is that if $X \subset Y \subset \mathbb{P}V$, then any polynomial vanishing on Y also vanishes on X. We want to find polynomials in the ideal of secant varieties of Segre varieties, so it is natural to look for varieties Y that contain $X = \sigma_r(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)$ whose ideals we understand. In this section we give two examples of such varieties Y.

12.1. $Flat_r^{\overline{a}}$ **and the GSS conjecture.**
Note that $A \otimes B \otimes C = A \otimes (B \otimes C)$, which leads to the simple observation that $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)) \subseteq \sigma_r(Seg(\mathbb{P}A \times \mathbb{P}(B \otimes C)))$. Moreover, we explicitly know the generators of the ideal of $\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}(B \otimes C)))$; see §10.1.

More generally, define the *flattening* of a tensor $T \in A_1 \otimes \cdots \otimes A_n$ by letting $I = \{i_1, \ldots, i_p\} \subset \{1, \ldots, n\}$, $J = \{1, \ldots, n\} \backslash I$, $A_I = A_{i_1} \otimes \cdots \otimes A_{i_p}$, $A_J = A_{j_1} \otimes \cdots \otimes A_{j_{n-p}}$ and consider $T \in A_I \otimes A_J$.

Let $\overline{a} = A_1 \otimes \cdots \otimes A_n$ and define $I_{Flat_r^{\overline{a}}}$ to be the ideal generated by the modules $\Lambda^{r+1}A_I^* \otimes \Lambda^{r+1}A_J^* \subset S^{r+1}(A_1 \otimes \cdots \otimes A_n)^*$ as I, J range over complementary subsets of $\{1, \ldots, n\}$. We let $Flat_r^{\overline{a}}$ denote the corresponding variety; that is,

$$Flat_r^{\overline{a}} = \cap_{I,J} \sigma_r(Seg(\mathbb{P}A_I \times \mathbb{P}A_J)).$$

We have $\sigma_r(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n) \subseteq Flat_r^{\overline{a}}$.

The *GSS conjecture* [28] is that equality holds when $r = 2$. Actually the conjecture is the stronger statement that $I_{\sigma_2(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)} = I_{Flat_2^{\overline{a}}}$. The weaker statement that equality holds as sets was proven in [37]. It was also shown in [37] that the conjecture holds when $\overline{a} = A_1 \otimes A_1 \otimes A_3$. Since $\sigma_2(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_n^*)$ is reduced and irreducible and $Flat_2^{\overline{a}}$ is irreducible, to prove the conjecture it would be sufficient to show that $Flat_2^{\overline{a}}$ is reduced. Using the methods outlined in §13, it is

possible to reduce the conjecture further to showing that $Flat_2^{\overline{a}}$ is arithmetically Cohen-Macaulay; see [41].

In [28], a computer calculation is presented that gives the dimensions of the minimal space of generators of the ideals of $\sigma_2(Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1))$ and $\sigma_2(Seg(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1))$, which, as shown in [5], allows one to prove the GSS conjecture for up to five factors. The proof relies on a variant of Proposition 12.2.2 which was arrived at independently using the geometry of phylogenetic trees.

12.2. Subspace varieties.

Definition 12.2.1. Define the *s-subspace variety*

(12.2.1) $Sub_s := \mathbb{P}\{T \in A_1 \otimes \cdots \otimes A_n \mid \exists A_j' \subset A_j \dim A_j' = s, T \in A_1' \otimes \cdots \otimes A_n'\}$.

Note that $\sigma_s(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n)) \subseteq Sub_s$, so the equations of Sub_s are also equations for $\sigma_s(Seg(\mathbb{P}A_1 \times \cdots \times \mathbb{P}A_n))$.

Proposition 12.2.2 ([39]). *The ideal of $\sigma_r(Seg(\mathbb{P}A_1^* \times \cdots \times \mathbb{P}A_n^*))$, when each* $\dim A_j^* \geq r$ *is generated by the union of the modules in its ideal inherited from the modules generating the ideal of $\sigma_r(Seg(\mathbb{P}^{r-1} \times \cdots \times \mathbb{P}^{r-1}))$ and the modules generating the ideal of Sub_r.*

To see this, note that by Proposition 11.4.1, a copy of a module $S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_n} A_n$ will be in $I(\sigma_r(Seg(\mathbb{P}^{r-1} \times \cdots \times \mathbb{P}^{r-1})))$ if and only if the corresponding copy of the module $S_{\pi_1}\mathbb{C}^{l(\pi_1)} \otimes \cdots \otimes S_{\pi_n}\mathbb{C}^{l(\pi_n)}$ is in the ideal of $\sigma_r(Seg(\mathbb{P}^{l(\pi_1)-1} \times \cdots \times \mathbb{P}^{l(\pi_n)-1}))$.

The ideal of Sub_r is easy to describe:

Theorem 12.2.3 ([41]). *The ideal of Sub_r is generated in degree $r + 1$ by the modules*

(12.2.2) $$\Lambda^{r+1}A_j \otimes \Lambda^{r+1}(A_1 \otimes \cdots \otimes A_{j-1} \otimes A_{j+1} \otimes \cdots \otimes A_n)$$

for $1 \leq j \leq n$ (minus redundancies).

Proof. First note that the ideal of Sub_r consists of all modules $S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_n} A_n$ occurring in $S^d(A_1 \otimes \cdots \otimes A_n)$ where each π_j is a partition of d and at least one π_j has $l(\pi_j) > r$. We need to show that this ideal is generated by the modules (12.2.2). But for each j, the ideal consisting of representations $S_{\pi_1} A_1 \otimes \cdots \otimes S_{\pi_n} A_n$ occurring in $S^d(A_1 \otimes \cdots \otimes A_n)$ where $l(\pi_j) > r$ is generated in degree $r + 1$ by

$$\Lambda^{r+1}A_j \otimes \Lambda^{r+1}(A_1 \otimes \cdots \otimes A_{j-1} \otimes A_{j+1} \otimes \cdots \otimes A_n),$$

because it is just the ideal of $\sigma_r(\mathbb{P}A_j \times \mathbb{P}(A_1 \otimes \cdots \otimes \hat{A}_j \otimes \cdots \otimes A_n))$. $\square$

Corollary 12.2.4 ([37]). *The ideal of $\sigma_2(Seg(\mathbb{P}A^* \times \mathbb{P}B^* \times \mathbb{P}C^*))$ is generated in degree three by $\Lambda^2 A \otimes \Lambda^2(B \otimes C)$, $\Lambda^2 B \otimes \Lambda^2(A \otimes C)$ and $\Lambda^2 C \otimes \Lambda^2(A \otimes B)$.*

Proof. $\sigma_2(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C) = Sub_2$ because $\sigma_2(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1) = \mathbb{P}(\mathbb{C}^2 \times \mathbb{C}^2 \times \mathbb{C}^2)$. $\square$

We remark that the spaces $\Lambda^2 A \otimes \Lambda^2(B \otimes C)$, $\Lambda^2 B \otimes \Lambda^2(A \otimes C)$, $\Lambda^2 C \otimes \Lambda^2(A \otimes B)$ intersect, so there is redundancy in the above description. This redundancy becomes apparent if one expresses the spaces as sums of irreducible modules.

The *s*-subspace variety is a cousin of the *rank varieties* in [57]. Moreover, it has a natural desingularization explained in §13.

13. Weyman's method

In this section we describe techniques for obtaining generators of the ideals of secant varieties of Segre varieties and more generally of G-varieties $Z \subset \mathbb{P}V$, where G is a reductive group, V is an irreducible G-module and by definition Z is a variety invariant under the action of G. In addition to providing generators of the ideal, the techniques enable one to compute the entire minimal free resolution of the ideal of Z as well as precise information about the singularities of Z. These techniques require considerably more machinery from commutative algebra and representation theory than we have used up until this point. We expect they will be useful in future work.

G-varieties are often uniruled by large linear spaces, and singularities occur when the linear spaces crash into one another. To remedy this, one could try to untangle the linear spaces. This appears to be the idea underlying Kempf's desingularization by the *collapsing of a vector bundle*. The idea is, given a G-variety $Z \subset \mathbb{P}V$, to find (i) a homogeneous variety G/P, (ii) a homogeneous vector bundle $E \to G/P$ that is the subbundle of a trivial bundle $\underline{V}$ with fiber isomorphic to V, and (iii) a map $\mathbb{P}E \to Z$ that is a desingularization. (Here P is a parabolic subgroup of G.)

For example, let $G(k, A)$ denote the Grassmannian of k-planes through the origin in A. Let $G = Gl(A) \times GL(B) \times GL(C)$, let $Z = Sub_s$ be as defined in Definition 12.2.1. Then let $G/P = G(s, A) \times G(s, B) \times G(s, C)$ and let $E = \mathcal{S}_A \otimes \mathcal{S}_B \otimes \mathcal{S}_C$, where $\mathcal{S}_A|_F$ is the s-plane $F \subset A$. Then $\mathbb{P}E \to Sub_s$ gives the desired desingularization.

Weyman takes Kempf's idea a step further by observing that often one can "push down" the minimal free resolution of the total space of E as a subvariety of the total space of the trivial bundle (more precisely, of the structure sheaf of E as an $\mathcal{O}_{\underline{V}}$-module) to obtain the minimal free resolution of Z. Moreover, since the whole procedure is G-equivariant, one gets the generators as modules.

The idea is as follows: Assume that the sheaf cohomology groups $H^i(S^d(E^*))$ are all zero for $i > 0$ and for all d. Consider the exact sequence

$$0 \to (\underline{V}/E)^* \to \underline{V}^* \to E^* \to 0$$

giving rise, for each j, to a sequence

$$0 \to \Lambda^j (\underline{V}/E)^* \to \Lambda^j \underline{V}^* \to \Lambda^{j-1} \underline{V}^* \otimes E^* \to \cdots \to \underline{V}^* \otimes S^{j-1} E^* \to S^j E^* \to 0.$$

Since $\underline{V}$ is trivial and by our hypothesis all terms but the first have no cohomology in degree greater than zero, when we take the long exact sequence in cohomology, we can split it into short exact sequences that we can in turn splice together to conclude that $H^k(\Lambda^j (\underline{V}/E)^*)$ is the k-th homology of the sequence

$$0 \to H^0(\Lambda^j \underline{V}^*) \to H^0(\Lambda^{j-1} \underline{V}^* \otimes E^*) \to \cdots \to H^0(S^j E^*) \to 0.$$

We add the hypothesis that the last step is surjective.

Now consider

$$
\begin{array}{ccccccccc}
\Lambda^d \underline{V}^* & \to & \Lambda^{d-1}\underline{V}^* \otimes H^0(S^1 E^*) & \to \cdots \to & \underline{V}^* \otimes H^0(S^{d-1}E^*) & \to & H^0(S^d E^*) & \to 0 \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & \\
\Lambda^d \underline{V}^* & \to & \Lambda^{d-1}\underline{V}^* \otimes V^* & \to \cdots \to & \underline{V}^* \otimes S^{d-1}V^* & \to & S^d V^* & \to 0 \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & \\
0 & \to & \Lambda^{d-1} \otimes I_1(Z) & \to \cdots \to & \underline{V}^* \otimes I_{d-1}(Z) & \to & I_d(Z) & \to 0
\end{array}
$$

where in the middle row we have $S^d V^* = H^0(S^d \underline{V}^*)$ which justifies the top row of vertical arrows. The horizontal arrows are from the Koszul sequence. The

generators of the ideal of Z in degree d correspond to the cokernel of the lower right arrow. Now apply the snake lemma to see that it is the homology of the d-th entry in the top sequence, which by the observation above is $H^{d-1}(\Lambda^d(\underline{V}/E)^*)$. (One obtains the full minimal free resolution in a similar fashion.)

All the bundles in question are homogeneous. If they are moreover irreducible, then one can apply the Bott-Borel-Weil theorem to reduce the calculation of the cohomology to a combinatorial calculation with the Weyl group of G. Even if they are not irreducible, one can use BBW on the associated graded bundles and then apply spectral sequences. For those who prefer to avoid spectral sequences in such calculations, see [46].

Note that since we had to use the snake lemma, we have no canonical way of identifying $H^{d-1}(\Lambda^d(\underline{V}/E)^*)$ with the space of generators in degree d, but in the equivariant setup, at least they agree as modules.

Sometimes it is sufficient to work with a partial desingularization of Z or a desingularization of a G variety that contains Z as a variety of small codimension. In fact, one does not need Z to be a G-variety (although for applications it almost always is). Here is Weyman's "basic theorem".

Theorem 13.0.1 ([57]). *Let $Y \subset \mathbb{P}V$ be a variety and suppose there is a projective variety B and a vector bundle $E \to B$ that is a subbundle of a trivial bundle $\underline{V} \to B$ with $\underline{V}_z \simeq V$ for $z \in B$ such that $E \to \hat{Y}$ is a desingularization. Write $\eta = E^*$ and $\xi = (\underline{V}/E)^*$.*

If the sheaf cohomology groups $H^i(B, S^d\eta)$ are all zero for $i > 0$ and the linear maps $H^0(B, S^d\eta) \otimes V^* \to H^0(B, S^{d+1}\eta)$ are surjective for all $d \geq 0$, then

    (1) *$\hat{Y}$ is normal, with rational singularities.*
    (2) *The coordinate ring $K[\hat{Y}]$ satisfies $K[\hat{Y}]_d \simeq H^0(B, S^d\eta)$.*
    (3) *The vector space of minimal generators of the ideal of $\hat{Y}$ in degree d is isomorphic to $H^{d-1}(B, \Lambda^d\xi)$, which is also the homology of the sequence*

$$\Lambda^2 V \otimes H^0(B, S^{d-2}\eta) \to V \otimes H^0(B, S^{d-1}\eta) \to H^0(B, S^d\eta).$$

    (4) *More generally, $\oplus_j H^j(\Lambda^{i+j}\xi)$ is isomorphic to the i-th term in the minimal free resolution of Y.*

If moreover Y is a G-variety and the desingularization is G-equivariant, then the identifications above are as G-modules.

Using these methods, the minimal generators of the ideals of $\sigma_r(Seg(\mathbb{P}^1 \times \mathbb{P}^b \times \mathbb{P}^c))$, $\sigma_3(\mathbb{P}^a \times \mathbb{P}^b \times \mathbb{P}^c)$ and $\sigma_2(\mathbb{P}^a \times \mathbb{P}^b \times \mathbb{P}^c \times \mathbb{P}^d)$ have been determined; see [41]. The method also gives information about the singularities (e.g. normality, arithmetically Cohen-Macaulay-ness), which, as mentioned above, can be used to reduce problems such as the GSS conjecture.

14. Appendix: Invariant formulations of two definitions from complexity theory

The purpose of this section is to show how *multiplicative complexity* and *separations* can be viewed invariantly, and to discusses advantages of the invariant perspective. While the discussion is elementary, it is intended primarily for those already familiar with these notions and their uses.

14.1. **Multiplicative complexity and tensors.** A slightly larger class of algorithms for executing bilinear maps $f : A \times B \to C$ than those discussed in §1.2 is obtained by writing $V = A \oplus B$ and considering T as a bilinear map $V \times V \to C$. The *multiplicative complexity* of T is the rank of T considered as a bilinear map $V \times V \to C$, that is, as an element of $V^* \otimes V^* \otimes C$. This definition differs from those in the literature, e.g., [14], p. 352, but is equivalent.

The multiplicative complexity is the minimal number of multiplications needed over all algorithms expressible as *straight line programs*, which is a class of algorithms that are intended to model (classical) computer programs. See [14], Definition 4.2, for a precise definition and a proof of this statement.

Our definition of multiplicative complexity gives an immediate proof of (14.8) in [14], which says that $R(T) \geq$ multiplicative complexity$(T) \geq 2R(T)$. To see this, note that $(A \oplus B) \otimes (A \oplus B) \otimes C = A \otimes B \otimes C \oplus A \otimes B \otimes C \otimes A \otimes A \otimes C \oplus B \otimes B \otimes C$, so any expression for T in $(A \oplus B)^{\otimes 2} \otimes C$ of rank r projects to an expression for T of rank at most $2r$ in $A \otimes B \otimes C$ (and of course the projections to $A \otimes A \otimes C$ and $B \otimes B \otimes C$ must be zero).

Here is an example where the multiplicative complexity of a tensor is lower than its rank, whose presentation also illustrates our definition.

Example 14.1.1. V. Alekseyev [3], building on work of Hopcroft and Kerr [31], showed that $\mathrm{Rank}(M_{2,2,3}) = 11$, but A. Waksman [56] gives an explicit algorithm for $M_{2,2,3}$ that uses 10 multiplications. Here is such an algorithm expressed as a tensor in $(A \oplus B)^* \otimes (A \oplus B)^* \otimes C$ (with bases as in §2):

$$M_{2,2,3} = \frac{1}{2}(\alpha_1^1 + \beta_1^2) \otimes (\alpha_2^1 + \beta_1^1) \otimes (c_1^1 - c_1^2) + \frac{1}{2}(\alpha_1^1 + \beta_2^2) \otimes (\alpha_2^1 + \beta_2^1) \otimes (c_2^1 + c_1^2 + c_3^2)$$

$$+ \frac{1}{2}(\alpha_1^1 + \beta_3^2) \otimes (\alpha_2^1 + \beta_3^1) \otimes (c_3^1 - c_3^2) + (\alpha_1^2 + \beta_1^2) \otimes (\alpha_2^2 + \beta_1^1) \otimes c_1^2$$

$$+ \frac{1}{2}(\alpha_1^2 + \beta_2^2) \otimes (\alpha_2^2 + \beta_2^1) \otimes (-c_1^2 + c_2^2 - c_3^2) + (\alpha_1^2 + \beta_3^2) \otimes (\alpha_2^2 + \beta_3^1) \otimes c_3^2$$

$$+ \frac{1}{2}(\alpha_1^1 - \beta_1^2) \otimes (-\alpha_2^1 + \beta_1^1) \otimes (c_1^1 + c_1^2) + \frac{1}{2}(\alpha_1^1 - \beta_2^2) \otimes (-\alpha_2^1 + \beta_2^1) \otimes (c_2^1 - c_1^2 - c_3^2)$$

$$+ \frac{1}{2}(\alpha_1^1 - \beta_3^2) \otimes (-\alpha_2^1 + \beta_3^1) \otimes (c_3^1 + c_3^2) + \frac{1}{2}(\alpha_1^2 - \beta_2^2) \otimes (-\alpha_2^2 + \beta_2^1) \otimes (c_1^2 + c_2^2 + c_3^2).$$

Remark 14.1.2. It might also be natural to consider expressions of $T \in A \otimes B \otimes C$ in $(A \oplus B \oplus C)^{\otimes 3}$, although it is not clear how to encode such an object in a straight line program. In any case, the savings would be at best by a factor of 6 by the same reasoning as in the paragraph above.

14.2. **Separations of computations.** A standard technique for establishing lower bounds (due to A. Alder and Strassen [2]) is *separations*. The best known lower bound for $M_{3,3,3}$ is 19 (due to Bläser [11]). It is obtained by extensive use of separations. In this section we define separations in a more invariant fashion than in [2] and suggest a more geometric variant.

Definition 14.2.1. Let $\phi \in A^* \otimes B^* \otimes C$ be a computed tensor with computation of length r. Let $A_1 \subseteq A$, $B_1 \subseteq B$, $C_1 \subseteq C$ be subspaces. We say ϕ *separates* (A_1, B_1, C_1) if we may write $\phi = \phi_1 + \phi_2 + \phi_3$ where the ϕ_i's are computed tensors whose lengths sum to r with the properties that

$$\mathrm{Lker}(\phi_1|_{A_1}) = 0, \ \mathrm{Rker}(\phi_2|_{B_1}) = 0$$

and no decomposable tensor appearing in the expression $\phi_1 + \phi_2$ takes values in C_1. (This definition is equivalent to the standard one.) Here for a bilinear map $\psi : A \times B \rightarrow C$, $\mathrm{Lker}(\psi) = \{a \in A \mid \psi(a, b) = 0 \, \forall b \in B\}$ and similarly for $\mathrm{Rker}(\psi) \subset B$.

For ϕ as above, the length of ϕ is at least $\dim A_1 + \dim B_1$ plus the number of decomposable tensors appearing in ϕ_3 taking values in C_1; this is called the *Separation Lemma*. As this observation indicates, separations are useful for obtaining lower bounds for the rank of a tensor.

If $\mathrm{Lker}(\phi) = 0$, then ϕ separates $(A, 0, 0)$, and similarly for the right kernel. If $\mathrm{Image}(\phi) = C$, then ϕ separates $(0, 0, C)$. Also, if ϕ separates (A', B', C'), then for any $A'' \subseteq A'$, $B'' \subseteq B'$, $C'' \subseteq C'$, ϕ separates (A'', B'', C'').

Lemma 14.2.2 ([2, Extension lemma]). *Let $\phi \in A^* \otimes B^* \otimes C$ be a computed tensor that separates (A_1, B_1, C_1). Let $A_1 \subseteq A_2 \subseteq A$. If ϕ fails to separate (A_2, B_1, C_1), then there exists $a \in A_2 \backslash A_1$ with*

$$(14.2.1) \qquad \phi(a, B) \subseteq \phi(a, B_1) + C_1.$$

Of course the same is true with the roles of A and B interchanged.

Proof. We try to write $\phi = \tilde{\phi}_1 + \tilde{\phi}_2 + \tilde{\phi}_3$ such that the tilded splitting of ϕ separates (A_2, B_1, C_1).

Write $\phi_3 = \tilde{\phi}_3 + \phi_3'$ with $\mathrm{Image}(\tilde{\phi}_3) \subset C_1$ and $\tilde{\phi}_3$ maximal with this property. (Note that $\tilde{\phi}_3$ is unique.) Then consider $\psi = \phi_1 + \phi_2 + \phi_3'$ and say ψ has length l. Then we have the best chance of separating (A_2, B_1, C_1) if we choose $\tilde{\phi}_2$ of minimal rank such that $\mathrm{Rker}\tilde{\phi}_2 \mid_{B_1} = 0$. Thus the length of $\tilde{\phi}_2 = \dim B_1 =: b_1$. There are at most $\binom{l}{b_1}$ choices of such $\tilde{\phi}_2$. Given any admissible such choice, the resulting $\tilde{\phi}_1 := \psi - \tilde{\phi}_2$ must also have the property that $\mathrm{Lker}\tilde{\phi}_1 \mid_{A_1} = 0$. Say we have such a choice and we want to see if the separation extends to A_2, that is, that $\mathrm{Lker}\tilde{\phi}_1 \mid_{A_2} = 0$. Now suppose not; then there exists $a \in A_2 \backslash A_1$ such that $a \in \mathrm{Lker}(\tilde{\phi}_1)$, and thus for all $b \in B$

$$\phi(a, b) = \tilde{\phi}_2(a, b) + \tilde{\phi}_3(a, b).$$

Write $B = B_1 \oplus \mathrm{Rker}(\tilde{\phi}_2)$ and given $b \in B$, $b = b' + b''$ uniquely with $b' \in B_1$, $b'' \in \mathrm{Rker}(\tilde{\phi}_2)$. So

$$\phi(a, b) = \phi(a, b') + \tilde{\phi}_3(a, b'') \in \langle \phi(a, B_1) \rangle + C_1.$$

So we see if ϕ fails to separate for at least one choice of tilded splitting equation, then (14.2.1) holds. In particular equation (14.2.1) holds if it fails for all possible choices. $\square$

Here is an easy application of the extension lemma:

Proposition 14.2.3. *If A is a simple algebra and $R \subset A$ a maximal right ideal, then any computation of $Mult_A$ separates $(R, A, 0)$.*

Proof. Since ϕ separates $(A, 0, 0)$ it separates $(R, 0, 0)$. Let $B_1 \subset B$ be maximal such that ϕ separates $(R, B_1, 0)$. If $B_1 \neq B$, then there exists a non-zero $b \in B$ such that $Ab \subseteq \langle RB \rangle = R$, a contradiction, as a left ideal cannot be contained in a right ideal. $\square$

As a corollary we obtain a very easy proof that $\mathbf{R}(M_{m,m,m}) \geq 2m^2 - m$.

Definition 14.2.4. A more natural and general definition of separation (which, to avoid confusion, we call Separation) is as follows: Given $T \in V_1^* \otimes \cdots \otimes V_n^*$, ϕ a computation of T and $U_j \subseteq V_j$, we will say ϕ *Separates* $(U_1, \ldots, U_n)$ if we have a decomposition $\phi = \phi_1 + \cdots + \phi_n + \psi$ with each

$$\phi_j : U_j \to V_1^* \otimes \cdots \otimes V_{j-1}^* \otimes V_{j+1}^* \otimes \cdots \otimes V_n^*$$

injective and $\text{length}(\phi) = \sum_i \text{length}(\phi_i) + \text{length}(\psi)$.

If ϕ Separates (A_1, B_1, C_1), then the length of ϕ is at least $\dim A_1 + \dim B_1 + \dim C_1$, so the conclusion of the corresponding Separation lemma is a little stronger than that of the separation lemma (but the hypotheses are stronger as well). Note that the hypotheses are also basis independent, unlike the separation lemma.

We leave the statement and proof of the analogous Extension lemma to the reader.

14.3. **Acknowledgments.** Many colleagues generously helped the author in the preparation of this article. Special thanks are due to E. Allman, M. Bläser, P. Bürgisser, L. Garcia, D. Gross, J. Morton, G. Ottaviani, C. Robles, and the anonymous referee for numerous suggestions to improve this article. In particular, the new proof of Bläser's theorem arose out of discussions with P. Bürgisser.

ABOUT THE AUTHOR

J. M. Landsberg is a professor of mathematics at Texas A&M University. His research applies exterior differential systems and representation theory to problems in differential and algebraic geometry.

REFERENCES

1. H. Abo, G. Ottaviani, and P. Peterson, *Induction for secant varieties of Segre varieties*, preprint, math.AG/0607191.
2. A. Alder and V. Strassen, *On the algorithmic complexity of associative algebras*, Theoret. Comput. Sci. **15** (1981), no. 2, 201–211. MR623595 (82g:68038)
3. Valery B. Alekseyev, *On the complexity of some algorithms of matrix multiplication*, J. Algorithms **6** (1985), no. 1, 71–85. MR780851 (86g:68070)
4. J. Alexander and A. Hirschowitz, *Polynomial interpolation in several variables*, J. Algebraic Geom. **4** (1995), no. 2, 201–222. MR1311347 (96f:14065)
5. Elizabeth S. Allman and John A. Rhodes, *Phylogenetic ideals and varieties for the general Markov model*, Advances in Applied Mathematics (to appear).
6. ———, *Phylogenetic invariants for the general Markov model of sequence mutation*, Math. Biosci. **186** (2003), no. 2, 113–144. MR2024609 (2004j:92048)
7. Wolf Barth, *Submanifolds of low codimension in projective space*, Proceedings of the International Congress of Mathematicians (Vancouver, B.C., 1974), Vol. 1, Canad. Math. Congress, Montreal, Que., 1975, pp. 409–413. MR0422294 (54:10285)
8. Ingemar Bengtsson and Karol Życzkowski, *Geometry of quantum states*, Cambridge University Press, Cambridge, 2006, An Introduction to Quantum Entanglement. MR2230995
9. Dario Bini, Milvio Capovani, Francesco Romani, and Grazia Lotti, $O(n^{2.7799})$ *complexity for* $n \times n$ *approximate matrix multiplication*, Inform. Process. Lett. **8** (1979), no. 5, 234–235. MR534068 (80h:68024)
10. Markus Bläser, *A* $\frac{5}{2}n^2$-*lower bound for the rank of* $n \times n$-*matrix multiplication over arbitrary fields*, 40th Annual Symposium on Foundations of Computer Science (New York, 1999), IEEE Computer Soc., Los Alamitos, CA, 1999, pp. 45–50. MR1916183
11. ———, *On the complexity of the multiplication of matrices of small formats*, J. Complexity **19** (2003), no. 1, 43–60. MR1951322 (2003k:68040)
12. M. Brambilla and G. Ottaviani, *On the Alexander-Hirschowitz theorem*, preprint math.AG/0701409.

13. Roger W. Brockett and David Dobkin, *On the optimal evaluation of a set of bilinear forms*, Linear Algebra and Appl. **19** (1978), no. 3, 207–235. MR0495183 (58:13915)

14. 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. MR1440179 (99c:68002)

15. M. V. Catalisano, A. V. Geramita, and A. Gimigliano, *On the rank of tensors, via secant varieties and fat points*, Zero-Dimensional Schemes and Applications (Naples, 2000), Queen's Papers in Pure and Appl. Math., vol. 123, Queen's Univ., Kingston, ON, 2002, pp. 133–147. MR1898833

16. ———, *Ranks of tensors, secant varieties of Segre varieties and fat points*, Linear Algebra Appl. **355** (2002), 263–285. MR1930149 (2003g:14070)

17. ———, *Higher secant varieties of Segre-Veronese varieties*, Projective Varieties with Unexpected Properties, Walter de Gruyter GmbH & Co. KG, Berlin, 2005, pp. 81–107. MR2202248

18. ———, *Secant varieties of Grassmann varieties*, Proc. Amer. Math. Soc. **133** (2005), no. 3, 633–642 (electronic). MR2113908 (2006d:14053)

19. M.V. Catalisano, A.V. Geramita, and A. Gimigliano, *On the ideals of secant varieties to certain rational varieties*, preprint math.AG/0609054.

20. J.A. Cavender and J. Felsenstein, *Invariants of phylogenies in a simple case with discrete states*, J. Classification **4** (1987), 57–71.

21. Ciro Ciliberto, *Geometric aspects of polynomial interpolation in more variables and of Waring's problem*, European Congress of Mathematics, Vol. I (Barcelona, 2000), Progr. Math., vol. 201, Birkhäuser, Basel, 2001, pp. 289–316. MR1905326 (2003i:14058)

22. H. Cohn, R. Kleinberg, B. Szegedy, and C. Umans, *Group-theoretic algorithms for matrix multiplication*, Proceedings of the 46th Annual Symposium on Foundations of Computer Science (2005), 379–388.

23. H. Cohn and C. Umans, *A group theoretic approach to fast matrix multiplication*, Proceedings of the 44th Annual Symposium on Foundations of Computer Science (2003), no. 2, 438–449.

24. Don Coppersmith and Shmuel Winograd, *Matrix multiplication via arithmetic progressions*, J. Symbolic Comput. **9** (1990), no. 3, 251–280. MR1056627 (91i:68058)

25. J. Eisert and D. Gross, *Multi-particle entanglement*, Lectures on Quantum Information (D. Bruss and G. Leuchs, eds.), Wiley-VCH, Weinheim, 2006, pp. 237–252.

26. Jens Eisert and Hans J. Briegel, *Schmidt measure as a tool for quantifying multiparticle entanglement*, Phys. Rev. A **64** (2001), no. 022306, 1–4.

27. William Fulton and Joe Harris, *Representation theory*, A first course, Graduate Texts in Mathematics, vol. 129, Readings in Mathematics, Springer-Verlag, New York, 1991. MR1153249 (93a:20069)

28. Luis David Garcia, Michael Stillman, and Bernd Sturmfels, *Algebraic geometry of Bayesian networks*, J. Symbolic Comput. **39** (2005), no. 3-4, 331–355. MR2168286 (2006g:68242)

29. Joe Harris, *Algebraic geometry*, A first course, Graduate Texts in Mathematics, vol. 133, Springer-Verlag, New York, 1995, corrected reprint of the 1992 original. MR1182558 (93j:14001)

30. Robin Hartshorne, *Varieties of small codimension in projective space*, Bull. Amer. Math. Soc. **80** (1974), 1017–1032. MR0384816 (52:5688)

31. J. E. Hopcroft and L. R. Kerr, *On minimizing the number of multiplications necessary for matrix multiplication*, SIAM J. Appl. Math. **20** (1971), 30–36. MR0274293 (43:58)

32. Serkan Hoşten and Suela Ruffa, *Introductory notes to algebraic statistics*, Rend. Istit. Mat. Univ. Trieste **37** (2005), no. 1-2, 39–70 (2006). MR2227048

33. Thomas A. Ivey and J. M. Landsberg, *Cartan for beginners: differential geometry via moving frames and exterior differential systems*, Graduate Studies in Mathematics, vol. 61, American Mathematical Society, Providence, RI, 2003. MR2003610 (2004g:53002)

34. Julian D. Laderman, *A noncommutative algorithm for multiplying 3 × 3 matrices using 23 multiplications*, Bull. Amer. Math. Soc. **82** (1976), no. 1, 126–128. MR0395320 (52:16117)

35. James A. Lake, *A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony*, Mol. Biol. Evol. **4** (1987), no. 2, 167–191.

36. J. M. Landsberg, *The border rank of the multiplication of 2 × 2 matrices is seven*, J. Amer. Math. Soc. **19** (2006), no. 2, 447–459 (electronic). MR2188132 (2006j:68034)

37. J. M. Landsberg and L. Manivel, *On the ideals of secant varieties of Segre varieties*, Found. Comput. Math. **4** (2004), no. 4, 397–422. MR2097214 (2005m:14101)

38. J. M. Landsberg and Laurent Manivel, *On the projective geometry of rational homogeneous varieties*, Comment. Math. Helv. **78** (2003), no. 1, 65–100. MR1966752 (2004a:14050)

39. J.M. Landsberg and L. Manivel, *Generalizations of Strassen's equations for Secant varieties of Segre varieties*, Communications in Algebra **36** (2008), 1–18.

40. J.M. Landsberg and J. Morton, *Computational complexity and geometry*, book in preparation.

41. J.M. Landsberg and J. Weyman, *On the ideals and singularities of secant varieties of Segre varieties*, preprint, math.AG/0601452, Bull. London Math. Soc. **39** (2007), no. 4, 685–697. MR2346950

42. R. Lazarsfeld and A. Van de Ven, *Topics in the geometry of projective space*, DMV Seminar, vol. 4, Birkhäuser Verlag, Basel, 1984, recent work of F. L. Zak, with an addendum by Zak. MR808175 (87e:14045)

43. Thomas Lickteig, *A note on border rank*, Inform. Process. Lett. **18** (1984), no. 3, 173–178. MR760371 (86c:68040)

44. ———, *Typical tensorial rank*, Linear Algebra Appl. **69** (1985), 95–120. MR798367 (87f:15017)

45. G. Ottaviani, *Symplectic bundles on the plane, secant varieties and Lüroth quartics revisited*, preprint, math.AG/0702151.

46. Giorgio Ottaviani and Elena Rubei, *Quivers and the cohomology of homogeneous vector bundles*, Duke Math. J. **132** (2006), no. 3, 459–508. MR2219264

47. Lior Pachter and Bernd Sturmfels (eds.), *Algebraic statistics for computational biology*, Cambridge University Press, New York, 2005. MR2205865 (2006i:92002)

48. A. Schönhage, *Partial and total matrix multiplication*, SIAM J. Comput. **10** (1981), no. 3, 434–455. MR623057 (82h:68070)

49. Jean-Pierre Serre, *Linear representations of finite groups*, Springer-Verlag, New York, 1977, translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42. MR0450380 (56:8675)

50. F. Severi, *Sintorno ai punti doppi impropri di una superficie generale dello spazio a quattro dimensioni, e a suio punti tripli apparenti*, Rend. Circ. Mat. Palermo **15** (1901), no. 2, 33–51.

51. J. Sidman and S. Sullivant, *Secant varieties and prolongations*, preprint, arXiv:math/0611696.

52. V. Strassen, *Rank and optimal computation of generic tensors*, Linear Algebra Appl. **52/53** (1983), 645–685. MR709378 (85b:15039)

53. ———, *Relative bilinear complexity and matrix multiplication*, J. Reine Angew. Math. **375/376** (1987), 406–443. MR882307 (88h:11026)

54. Volker Strassen, *Gaussian elimination is not optimal*, Numer. Math. **13** (1969), 354–356. MR0248973 (40:2223)

55. R. C. Vaughan and T. D. Wooley, *Waring's problem: a survey*, Number Theory for the Millennium, III (Urbana, IL, 2000), A K Peters, Natick, MA, 2002, pp. 301–340. MR1956283 (2003j:11116)

56. Abraham Waksman, *On Winograd's algorithm for inner products*, IEEE Trans. Computers **C-19** (1970), no. 4, 360–361. MR0455534 (56:13772)

57. Jerzy Weyman, *Cohomology of vector bundles and syzygies*, Cambridge Tracts in Mathematics, vol. 149, Cambridge University Press, Cambridge, 2003. MR1988690 (2004d:13020)

58. S. Winograd, *On multiplication of 2×2 matrices*, Linear Algebra and Appl. **4** (1971), 381–388. MR0297115 (45:6173)

59. F. L. Zak, *Projections of algebraic varieties*, Mat. Sb. (N.S.) **116(158)** (1981), no. 4, 593–602, 608. MR665860 (84i:14012)

60. ———, *Tangents and secants of algebraic varieties*, Translations of Mathematical Monographs, vol. 127, American Mathematical Society, Providence, RI, 1993, translated from the Russian manuscript by the author. MR1234494 (94i:14053)

Department of Mathematics, Texas A&M University, College Station, Texas 77843-3368

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