Tensor decomposition and tensor rank from the point of view of Classical Algebraic Geometry *RTG Workshop* Tensors and their Geometry in High Dimensions (September 26-29, 2012) *UC Berkeley*

Giorgio Ottaviani

Università di Firenze

CIME will organize a summer course in Levico Terme (Italy) June 10th - 15th Combinatorial Algebraic Geometry

Lecturers

- Aldo Conca
- Sandra Di Rocco
- Ian Draisma
- Bernd Sturmfels
- Filippo Viviani



- Wednesday Rank and symmetric rank. Tensor decomposition. Classical apolarity and Sylvester algorithm. Secant varieties. Clebsch quartics. Sum of squares, sum of k-th powers.
- **Thursday** Cases where classical apolarity fails. Vector bundles and non abelian apolarity. Equations for secant varieties, infinitesimal criterion for smoothness. Scorza map and Lüroth quartics. Identifiability.
- **Friday** Actions of *SL*(2). The complexity of Matrix Multiplication Algorithm.

Any tensor $t \in \mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_d}$ defines by contraction a function f_t over the product $S = S_{m_1-1} \times \ldots \times S_{m_d-1}$ of the corresponding d spheres. $f_t : S \to \mathbb{R}$

Theorem (Lim, Qi)

The critical points of f_t corresponds to tensors $(x_1,\ldots,x_d)\in S$ such that

$$t(x_1,\ldots,\hat{x}_i,\ldots,x_d)=\lambda_i x_i$$

Theorem (Friedland-O)

The number of singular d-ples of a general tensor t over \mathbb{C} is the coefficient of $\prod_i t_i^{m_i-1}$ in the polynomial

$$\prod_i rac{\hat{t_i}^{m_i}-t_i^{m_i}}{\hat{t_i}-t_i}$$

where $\hat{t}_i = \sum_{j \neq i} t_j$

The format $k_0 \times k_1 \times \ldots \times k_p$ with $k_0 = \max_j k_j$ is called boundary format if

$$k_0 - 1 = \sum_{i=1}^{p} (k_i - 1)$$

This is the format where is possible to define the diagonal and it is the analog of square matrix.

The basic example is given by the multiplication tensor $S^{k_1-1}\mathbb{C}^2 \otimes \ldots \otimes S^{k_p-1}\mathbb{C}^2 \to S^{\sum_i (k_i-1)}\mathbb{C}^2$ which sits in $\otimes_{i=0}^p (S^{k_i-1}\mathbb{C}^2)$ In the boundary format case it is well defined a unique "diagonal" given by elements $a_{i_0...i_p}$ satisfying $i_0 = \sum_{j=1}^p i_j$



Definition

A p + 1-dimensional tensor of boundary format $A \in V_0 \otimes \ldots \otimes V_p$ is called triangulable if there exist bases in V_j such that $a_{i_0,\ldots,i_p} = 0$ for $i_0 > \sum_{t=1}^p i_t$

Definition

A p + 1-dimensional tensor of boundary format $A \in V_0 \otimes \ldots \otimes V_p$ is called diagonalizable if there exist bases in V_j such that $a_{i_0,\ldots,i_p} = 0$ for $i_0 \neq \sum_{t=1}^p i_t$

The "identity" matrices

Definition

A p + 1-dimensional tensor of boundary format $A \in V_0 \otimes \ldots \otimes V_p$ is an identity if one of the following equivalent conditions holds i) there exist bases in V_j such that

$$\mathbf{a}_{i_0,\dots,i_p} = \left\{ egin{array}{ccc} 0 & ext{for} & i_0
eq \sum_{t=1}^p i_t \\ 1 & ext{for} & i_0 = \sum_{t=1}^p i_t \end{array}
ight.$$

ii) there exist a vector space U of dimension 2 and isomorphisms $V_j \simeq S^{k_j}U$ such that A belongs to the unique one dimensional SL(U)-invariant subspace of $S^{k_0}U \otimes S^{k_1}U \otimes \ldots \otimes S^{k_p}U$

The equivalence between i) and ii) follows easily from the following remark: the matrix A satisfies the condition ii) if and only if it corresponds to the natural multiplication map $S^{k_1}U \otimes \ldots \otimes S^{k_p}U \rightarrow S^{k_0}U$ (after a suitable isomorphism $U \simeq U^{\vee}$ has been fixed).

Intrinsic characterizations

The definitions of triangulable, diagonalizable and identity apply to elements of $\mathbb{P}(V_0 \otimes \ldots \otimes V_p)$ as well. In particular all identity matrices fill a distinguished orbit in $\mathbb{P}(V_0 \otimes \ldots \otimes V_p)$. We denote by *Stab* $(A) \subset SL(V_0) \times \ldots \times SL(V_p)$ the stabilizer subgroup of A and by Stab $(A)^0$ its connected component containing the identity. The main results are the following.

Theorem

([AO]) Let $A \in \mathbb{P}(V_0 \otimes \ldots \otimes V_p)$ of boundary format such that Det $A \neq 0$. Then

A is triangulable \iff A is not stable for $SL(V_0) \times \ldots \times SL(V_p)$

Theorem

([AO]) Let $A \in \mathbb{P}(V_0 \otimes \ldots \otimes V_p)$ be of boundary format such that Det $A \neq 0$. Then

A is diagonalizable \iff Stab(A) contains a subgroup $\simeq \mathbb{C}^*$ Giorgio Ottaviani

Tensor decomposition and tensor rank

The proof of the above two theorems relies on the Hilbert-Mumford criterion. The proof of the following theorem needs more geometry.

Theorem

([AO] for p = 2, [D] for $p \ge 3$) Let $A \in \mathbb{P}(V_0 \otimes V_1 \otimes \ldots \otimes V_p)$ of boundary format such that Det $A \ne 0$. Then there exists a 2-dimensional vector space U such that SL(U) acts over $V_i \simeq S^{k_i}U$ and according to this action on $V_0 \otimes \ldots \otimes V_p$ we have Stab $(A)^0 \subset SL(U)$. Moreover the following cases are possible

 $Stab (A)^{0} \simeq \begin{cases} 0 & (trivial \ subgroup) \\ \mathbb{C} \\ \mathbb{C}^{*} \\ SL(2) & (this \ case \ occurs \ if \ and \ only \ if \ A \ is \ an \ identity) \end{cases}$

When A is an identity then Stab $(A) \simeq SL(2)$.

Weierstrass canonical form, case $2 \times k \times (k+1)$

The case $2 \times k \times (k+1)$ has boundary format and it was solved by Weierstrass.

Theorem (Weierstrass)

All nondegenerate matrices of type $2 \times k \times (k+1)$ are GL(2) × GL(k) × GL(k+1) equivalent to the identity matrix having the two slices



The proof shows first that there is a dense orbit.

Let (x_0, x_1) be homogeneous coordinates on \mathbb{P}^1 . The identity matrix appearing in Weierstrass canonical form corresponds to the morphism of vector bundles given by

$$I_k(x_0, x_1) := \begin{pmatrix} x_0 & x_1 & \\ & \ddots & \ddots & \\ & & x_0 & x_1 \end{pmatrix}$$

The format $2 \times k \times (k+1)$ is a building block for all the other formats $2 \times b \times c$. The canonical form illustrated by the following Theorem is called the Kronecker canonical form (there is an extension in the degenerate case that we do not pursue here).

Theorem (Kronecker, 1890)

Let $2 \leq b < c$. There exist unique $n, m, q \in \mathbb{N}$ satisfying

$$\begin{cases} b = nq + m(q+1) \\ c = n(q+1) + m(q+2) \end{cases}$$

such that the general tensor $t \in \mathbb{C}^2 \otimes \mathbb{C}^b \otimes \mathbb{C}^c$ decomposes as n blocks $2 \times q \times (q+1)$ and m blocks $2 \times (q+1) \times (q+2)$ in Weierstrass form.

The Fibonacci blocks

Kac has generalized this statement to the format $2 \le w \le s \le t$ satisfying the inequality $t^2 - wst + s^2 \ge 1$. The result is interesting because it gives again a canonical form. Given w, define by the recurrence relation $a_0 = 0$, $a_1 = 1$, $a_j = wa_{j-1} - a_{j-2}$ For w = 2 get 0, 1, 2, ... and Kronecker's result. For w = 3 get 0, 1, 3, 8, 21, 55, ... (odd Fibonacci numbers)



Figure: A decomposition in two Fibonacci blocks

Theorem (Kac, 1980)

Let $2 \le w \le s \le t$ satisfying the inequality $t^2 - wst + s^2 \ge 1$. Then there exist unique $n, m, j \in \mathbb{N}$ satisfying

$$\left\{ egin{array}{ll} s=& \mathit{na}_{j} &+ \mathit{ma}_{j+1} \ t=& \mathit{na}_{j+1} &+ \mathit{ma}_{j+2} \end{array}
ight.$$

such that the general tensor $t \in \mathbb{C}^w \otimes \mathbb{C}^s \otimes \mathbb{C}^t$ decomposes as n blocks $w \times a_j \times a_{j+1}$ and m blocks $w \times a_{j+1} \times a_{j+2}$ which are denoted "Fibonacci blocks". They can be described by representation theory (see [Brambilla]).

The original proof of Kac uses representations of quivers. In [Brambilla] there is an independent proof in the language of vector bundles.

Relevance of matrix multiplication algorithm

Many numerical algorithms use matrix multiplication. The complexity of matrix multiplication algorithm is crucial in many numerical routines.

 $M_{m,n} =$ space of $m \times n$ matrices

Matrix multiplication is a bilinear operation

$$egin{array}{lll} M_{m,n} imes M_{n,l} o & M_{m,l} \ (A,B) & \mapsto A \cdot B \end{array}$$

where $A \cdot B = C$ is defined by $c_{ij} = \sum_k a_{ik} b_{kj}$. This usual way to multiply a $m \times n$ matrix with a $n \times l$ matrix requires *mnl* multiplications and ml(n-1) additions, so asympotically 2mnl elementary operations. The usual way to multiply two 2×2 matrices requires eight multiplication and four additions.

Matrix multiplication can be seen as a tensor

 $t_{m,n,l} \in M_{m,n} \otimes M_{n,l} \otimes M_{m,l}$ $t_{m,n,l}(A \otimes B \otimes C) = \sum_{i,j,k} a_{ik}b_{kj}c_{ji} = tr(ABC)$ and the number of multiplications needed coincides with the rank of $t_{m,n,l}$ with respect to the Segre variety $\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C$ of decomposable tensors.

Allowing approximations, the border rank of t is a good measure of the complexity of the algorithm of matrix multiplication.

Strassen showed explicitly

$$\begin{split} M_{2,2,2} = & a_{11} \otimes b_{11} \otimes c_{11} + a_{12} \otimes b_{21} \otimes c_{11} + a_{21} \otimes b_{11} \otimes c_{21} + a_{22} \otimes b_{21} \otimes c_{21} \\ & + a_{11} \otimes b_{12} \otimes c_{12} + a_{12} \otimes b_{22} \otimes c_{12} + a_{21} \otimes b_{12} \otimes c_{22} + a_{22} \otimes b_{22} \otimes c_{22} \end{split}$$

$$= (a_{11} + a_{22}) \otimes (b_{11} + b_{22}) \otimes (c_{11} + c_{22}) + (a_{21} + a_{22}) \otimes b_{11} \otimes (c_{21} - c_{22}) + a_{11} \otimes (b_{12} - b_{22}) \otimes (c_{12} + c_{22}) + a_{22} \otimes (-b_{11} + b_{21}) \otimes (c_{21} + c_{11}) + (a_{11} + a_{12}) \otimes b_{22} \otimes (-c_{11} + c_{12}) + (-a_{11} + a_{21}) \otimes (b_{11} + b_{12}) \otimes c_{22} + (a_{12} - a_{22}) \otimes (b_{21} + b_{22}) \otimes c_{11}.$$

$$(1)$$

Dividing a matrix of size $2^k \times 2^k$ into 4 blocks of size $2^{k-1} \times 2^{k-1}$ one shows inductively that are needed 7^k multiplications and $9 \cdot 2^k + 18 \cdot 7^{k-1}$ additions, so in general $\leq C7^k$ elementary operations.

The number 7 of multiplications needed turns out to be the crucial measure.

The exponent of matrix multiplication ω is defined to be $\underline{\lim}_n \log_n$ of the arithmetic cost to multiply $n \times n$ matrices, or equivalently, $\underline{\lim}_n \log_n$ of the minimal number of multiplications needed. A consequence of Strassen bound is that $\omega \leq \log_2 7 = 2.81 \dots$ The border rank in case 3×3 is still unknown. Our results are as follows:

Theorem (O-Landsberg)

Let $n \leq m$.

$$br(T_{m,n,l}) \geq \frac{nl(n+m-1)}{m}$$

Corollary

$$br(T_{n,n,l}) \ge 2nl - l$$
$$br(T_n) \ge 2n^2 - n.$$

Thus for 3×3 matrices, the state of the art is $15 \leq br(M_{(3,3,3)}) \leq 21$, the upper bound is due to Schönhage .

Bläser bound

Bläser proved the following lower bounds for the *rank* of matrix multiplication are $\mathbf{R}(M_{m,n,l} \ge lm + mn + l - m + n - 3)$, $\mathbf{R}(M_{n,n,l}) \ge 2ln - l + 2n - 2$, and $\mathbf{R}(M_n) \ge \frac{5}{2}n^2 - 3n$. Recent improvements due to Landsberg.

We define, for every p, a linear map

$$(M_{\langle m,n,l\rangle})^{\wedge p}_A \colon \mathbb{C}^{nl\binom{mn}{p}} \to \mathbb{C}^{ml\binom{mn}{p+1}}$$

and we prove that $\mathbf{R}(M_{m,n,l}) \geq {\binom{mn-1}{p}} \operatorname{rank} \left[(M_{\langle m,n,l \rangle})_A^{\wedge p} \right]$. We then compute the rank of the linear map $(M_{\langle m,n,l \rangle})_A^{\wedge p}$.

Let A, B, C be complex vector spaces of dimensions a, b, c, with $b \le c$, and with dual vector spaces A^*, B^*, C^* .

The most naïve equations for $\sigma_r(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ are the so-called *flattenings*. Given $T \in A \otimes B \otimes C$, consider $T_B : B^* \longrightarrow A \otimes C$ as a linear map. Then $\operatorname{br}(T) \geq \operatorname{rank}(T_B)$ and similarly for cyclic permutations of A, B, C. We consider

$$T_A^{\wedge p}: B^* \otimes \wedge^p A \longrightarrow \wedge^{p+1} A \otimes C.$$

The setting, II

To avoid redundancies, assume $b \le c$ and $p \le \lfloor \frac{a}{2} \rfloor - 1$. Then, if $T = a \otimes b \otimes c$ is of rank one,

$$\operatorname{rk}((a \otimes b \otimes c)_A^{\wedge p}) = {a-1 \choose p}.$$

To see this, expand $a = \alpha_1$ to a basis $\alpha_1 \dots \alpha_a$ of A with dual basis $\alpha^1 \dots \alpha^a$ of A^* . Then $T_A^{\wedge p} = [\alpha^{i_1} \wedge \dots \alpha^{i_p} \otimes b] \otimes [\alpha_1 \wedge \alpha_{i_1} \wedge \dots \alpha_{i_p} \otimes c]$, so the image is isomorphic to $\wedge^p(A/\alpha_1) \otimes c$. When T is generic, we expect $T_A^{\wedge p}$ to be injective, thus potentially

When T is generic, we expect $T_A^{\mu\nu}$ to be injective, thus potentially obtaining modules of equations up to

$$r = \frac{b\binom{a}{p}}{\binom{a-1}{p}} = \frac{ba}{a-p}.$$

Since this is an increasing function of p, one gets the most equations taking p equal to its maximal value, $p = \lceil \frac{a}{2} \rceil - 1$.

Corollary

Set $a \leq b \leq c$. Then the maps $T_A^{\wedge p}$ give nontrivial equations for $\sigma_r(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ for $r \leq 2a - \sqrt{a}$.

Determining the precise module structure of the equations (i.e., which irreducible submodules actually contribute nontrivial equations) appears to be difficult.

3×3 case

Our computation started in the 3×3 case.

In this case the multiplication tensor sits in $\mathbb{C}^9 \otimes \mathbb{C}^9 \simeq \mathbb{C}^{729}$ **Question** What's the length of the tensor decomposition of this tensor that can be achieved by numerical techniques ? We think at

$$t_{ij,kl,mn} = \begin{cases} 1 \text{ if } j = k, l = m, n = i \\ 0 \text{ otherwise} \end{cases}$$

We have a contraction map

$$\mathbb{C}^{1134} \simeq \mathbb{C}^9 \otimes \wedge^4 \mathbb{C}^9 \to \mathbb{C}^9 \otimes \wedge^4 \mathbb{C}^9 \simeq \mathbb{C}^{1134}$$

and the maximum rank expected is 1134 The answer was 918.

Indeed, Let M, N, L be vector spaces of dimensions m, n, l. Write $A = M \otimes N^*$, $B = N \otimes L^*$, $C = L \otimes M^*$, so a = mn, b = nl, c = ml. The matrix multiplication operator $M_{\leq m,n,l>}$ is $M_{\leq m,n,l>} = Id_M \otimes Id_N \otimes Id_L \in A \otimes B \otimes C$. We compute the kernel of the map

Matrix multiplication tensor is quite special. One cannot expect a unique honest decomposition. Indeed it is invariant by a big isotropy group, because

$$tr(ABC) = tr\left((G^{-1}AH)(H^{-1}BK)(K^{-1}CG)\right)$$

We will apply the inheritance principle to the case of an (n + m - 1)-plane $A' \subset A = \mathbb{C}^{nm}$. Assume $b \leq c$, so $n \leq m$. The essential idea for the proof is to choose a subspace $A' \subset M \otimes N^*$ on which the restriction of $M_{< m,n,l>n}$ becomes injective. Take a vector space W of dimension 2, and fix isomorphisms $N \simeq S^{n-1}W$. $M \simeq S^{m-1}W^*$. Let A' be the direct summand $S^{m+n-2}W^* \subset S^{n-1}W^* \otimes S^{n-1}W^* = M \otimes N^*$. Recall that $S^{\alpha}W$ may be interpreted as the space of homogenous polynomials of degree α in two variables. If $f \in S^{\alpha}W$ and $g \in S^{\beta}W^*$ then we can perform the contraction $g \cdot f \in S^{\alpha-\beta}W$. In the case $f = I^{\alpha}$ is the power of a linear form I, then the contraction $g \cdot I^{\alpha}$ equals $I^{\alpha-\beta}$ multiplied by the value of g at the point *I*, so that (for $\beta \leq \alpha$) $g \cdot I^{\alpha} = 0$ if and only if *I* is a root of *g*. Consider the natural skew-symmetrization map

$$A'\otimes\wedge^{n-1}(A')\longrightarrow\wedge^n(A')$$

Recall that representation theory distinguishes a complement A'' to A, so the projection $M \otimes N^* \longrightarrow A'$ is well defined. Compose with the projection

$$M \otimes N^* \otimes \wedge^{n-1}(A') \longrightarrow A' \otimes \wedge^{n-1}(A')$$

to obtain

$$M \otimes N^* \otimes \wedge^{n-1}(A') \longrightarrow \wedge^n (A').$$

Now the equation is equivalent to a map

$$\psi'_{p}: N^{*} \otimes \wedge^{n-1}(A') \longrightarrow M^{*} \otimes \wedge^{n}(A').$$

We claim it is injective. (Note that when n = m the source and target space are dual to each other.)

The proof, II

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

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

We show this dual map is surjective. Let $l^{n-1} \otimes (l_1^{m+n-2} \wedge \cdots \wedge l_{n-1}^{m+n-2}) \in S^{n-1}W \otimes \wedge^{n-1}S^{m+n-2}W$ with $l_i \in W$. Such elements span the target so it will be sufficient to show any such element is in the image. Assume first that l is distinct from the l_i . Since $n \leq m$, there is a polynomial $g \in S^{m-1}W^*$ which vanishes on l_1, \ldots, l_{n-1} and is nonzero on l. Then, up to a nonzero scalar, $g \otimes (l_1^{m+n-2} \wedge \cdots \wedge l_{n-1}^{m+n-2} \wedge l^{m+n-2})$ maps to our element.

Since the image is closed (being a linear space), the condition that l is distinct from the l_i may be removed by taking limits. Finally, $\psi'_p \otimes Id_L$ is the map induced from the restricted matrix multiplication operator and we may repeat the general arguments. To complete the proof, observe that an element of rank one in $A' \otimes B \otimes C$ induces a map of rank $\binom{n+m-2}{n-1}$. So the rank of the multiplication operator must be at least

$$\frac{\dim L\otimes N^*\otimes\wedge^{n-1}(A')}{\binom{n+m-2}{n-1}}=nI\frac{\binom{n+m-1}{n-1}}{\binom{n+m-2}{n-1}}=\frac{nI(n+m-1)}{m}$$

which proves our result.

Note that if we have two different X-decompositions of a tensor $f = \sum_{i=1}^{r} x_i = \sum_{j=1}^{r} y_j$ then the tangent space at f of $\sigma_r(X)$ is tangent to X at all points x_i and y_j .

Definition (Chiantini-Ciliberto)

X is called not k-weakly defective if the general tangent hyperplane tangent to X at k general points $x_1, \ldots, x_k \in X$ is tangent only at these points. The locus where it is tangent is called the contact locus.

Theorem

not k-weakly defective \implies k-identifiable

Moreover, the theorem allows computer experiments.

Theorem (Sylvester[1851], Chiantini-Ciliberto, Mella, Ballico, [2002-2005])

The general $f \in S^d \mathbb{C}^{n+1}$, $d \neq 3$, of rank s smaller than the generic one, has a unique Waring decomposition, with the only exceptions

- rank $s = \binom{n+2}{2} 1$ in $S^4 \mathbb{C}^{n+1}$, $2 \le n \le 4$, when there are infinitely many decompositions
- rank 7 in S³C⁵, when there are infinitely many decompositions
- \bullet rank 9 in $S^6 \mathbb{C}^3,$ where there are exactly two decompositions
- rank 8 in $S^4 \mathbb{C}^4$, where there are exactly two decompositions

The cases listed in red are called the *defective cases*. The cases listed in blue are called the *weakly defective cases*.

Weakly defective examples

Assume for simplicity k = 3. Only known examples where the general $f \in V_1 \otimes V_2 \otimes V_3$ (dim $V_i = n_i + 1$) of subgeneric rank s has a NOT UNIQUE decomposition, besides the defective ones, are

- unbalanced case, rank $s = n_1 n_2 + 1$, $n_3 \ge n_1 n_2 + 1$
- rank 6 $(n_1, n_2, n_3) = (3, 3, 3)$ where there are two decompositions
- rank 8 $(n_1, n_2, n_3) = (2, 5, 5)$, sporadic case [CO], maybe six decompositions

Theorem

- The unbalanced case is understood [Chiantini-O. [2011]].
- There is a unique decomposition for general tensor of rank s in $\mathbb{C}^{n+1} \otimes \mathbb{C}^{n+1} \otimes \mathbb{C}^{n+1}$ if $s \leq \frac{3n+1}{2}$ [Kruskal[1977] if $s \leq \frac{(n+2)^2}{16}$ [Chiantini-O. [2011]]
- The exceptions to uniqueness listed in the previous slide are the only ones in the cases
 (i) n_i ≤ 6
 (ii) s ≤ 6 [Chiantini-O. [2011]]

Proof uses a generalization of the inductive technique in [AOP] plus the weak defectivity.

It is given by the subvariety of X where the general hyperplane tangent at k general points is tangent. In the case $\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3$ and k = 6, the contact locus is an elliptic normal curve. In the case $\mathbb{P}^2 \times \mathbb{P}^5 \times \mathbb{P}^5$ and k = 8, the contact locus is $\mathbb{P}^2 \times \mathbb{P}^1 \times \mathbb{P}^1$ embedded with $\mathcal{O}(3, 1, 1)$. Thanks !!