Matrix multiplication and geometry

J.M. Landsberg

Texas A&M University and Univ. P. Sabatier

Supported by NSF grants AF-1814254 and AF-2203618

Strassen's spectacular failure

Standard algorithm for matrix multiplication, row-column:

$$\begin{pmatrix} * & * & * \\ & & \\ & & \end{pmatrix} \begin{pmatrix} * & & \\ * & & \\ * & & \end{pmatrix} = \begin{pmatrix} * & & \\ & & \\ & & \end{pmatrix}$$

uses $O(n^3)$ arithmetic operations.

Strassen (1968) set out to prove this standard algorithm was indeed the best possible.

At least for 2×2 matrices. At least over $\mathbb{F}_2$.

He failed.

Strassen's algorithm

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

$$
\begin{aligned}
I &= (a_1^1 + a_2^2)(b_1^1 + b_2^2), \\
II &= (a_1^2 + a_2^2)b_1^1, \\
III &= a_1^1(b_2^1 - b_2^2) \\
IV &= a_2^2(-b_1^1 + b_1^2) \\
V &= (a_1^1 + a_2^1)b_2^2 \\
VI &= (-a_1^1 + a_1^2)(b_1^1 + b_2^1), \\
VII &= (a_2^1 - a_2^2)(b_1^2 + b_2^2),
\end{aligned}
$$

If $C = AB$, then

$$
\begin{aligned}
c_1^1 &= I + IV - V + VII, \\
c_1^2 &= II + IV, \\
c_2^1 &= III + V, \\
c_2^2 &= I + III - II + VI.
\end{aligned}
$$

Astounding conjecture

Iterate: $\rightsquigarrow 2^k \times 2^k$ matrices using $7^k \ll 8^k$ multiplications,

and $n \times n$ matrices with $O(n^{2.81})$ arithmetic operations.

Bini 1978, Schönhage 1983, Strassen 1987, Coppersmith-Winograd 1988 $\rightsquigarrow O(n^{2.3755})$ arithmetic operations.

Astounding Conjecture

For all $\epsilon > 0$, $n \times n$ matrices can be multiplied using $O(n^{2+\epsilon})$ arithmetic operations.

$\rightsquigarrow$ *asymptotically, multiplying matrices is nearly as easy as adding them!*

1988-2011 no progress, 2011-14 Stouthers, Vasilevska-Williams, LeGall, 2021 Alman and V-W .004 improvement.

Tensor formulation of conjecture

Set $N = n^2$.
Matrix multiplication is a *bilinear* map

$$M_{\langle n \rangle} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N,$$

Bilinear maps $\mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N$ may also be viewed as trilinear

maps $\mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^{N*} \to \mathbb{C}$.

In other words

$$M_{\langle n \rangle} \in \mathbb{C}^{N*} \otimes \mathbb{C}^{N*} \otimes \mathbb{C}^N.$$

Exercise: As a trilinear map, $M_{\langle n \rangle}(X, Y, Z) = \text{trace}(XYZ)$.

Tensor formulation of conjecture

A tensor $T \in \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N =: A \otimes B \otimes C$ has *rank one* if it is of the form $T = a \otimes b \otimes c$, with $a \in A$, $b \in B$, $c \in C$. Rank one tensors correspond to bilinear maps that can be computed using one scalar multiplication.

The *rank* of a tensor T, $\mathbf{R}(T)$, is the smallest r such that T may be written as a sum of r rank one tensors. The rank is essentially the number of scalar multiplications needed to compute the corresponding bilinear map.

Tensor formulation of conjecture

Theorem (Strassen): $M_{\langle n \rangle}$ can be computed using $O(n^{\tau})$ arithmetic operations $\Leftrightarrow \mathbf{R}(M_{\langle n \rangle}) = O(n^{\tau})$

Let $\omega := \inf_{\tau}\{\mathbf{R}(M_{\langle n \rangle}) = O(n^{\tau})\}$

ω is called the *exponent* of matrix multiplication.

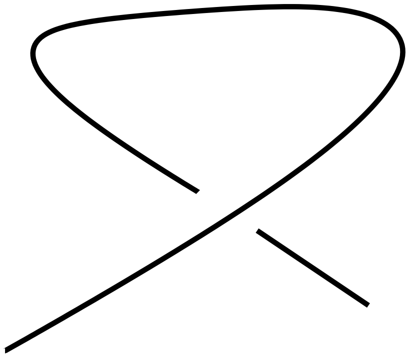
Classical: $\omega \leq 3$.

Corollary of Strassen's algorithm: $\omega \leq \log_2(7) \simeq 2.81$.

Astounding Conjecture

$\omega = 2$

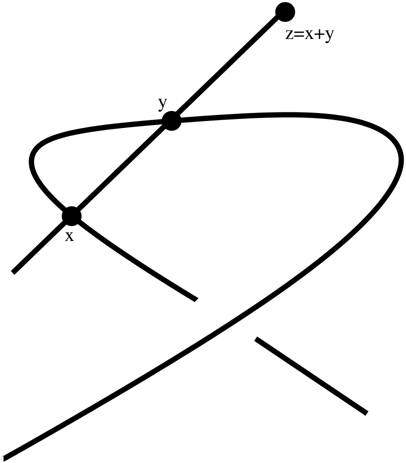
Geometric formulation of conjecture



Imagine this curve represents the set of tensors of rank one sitting in the N^3 dimensional space of tensors.

Geometric formulation of conjecture

$\{$ tensors of rank at most two$\}$ =
$\{$ points on a secant line to set of tensors of rank one$\}$



Conjecture is about a point (matrix multiplication) lying on a secant r-plane to set of tensors of rank one.

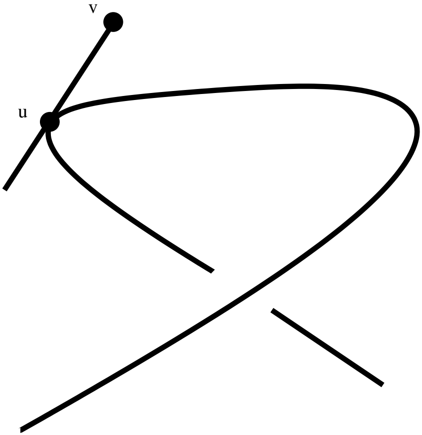
Bini's sleepless nights

Bini-Capovani-Lotti-Romani (1979) investigated if $M_{\langle 2 \rangle}$, with one matrix entry set to zero, could be computed with five multiplications (instead of the naïve 6), i.e., if this reduced matrix multiplication tensor had rank 5.

They used numerical methods.

Their code appeared to have a problem.

The limit of secant lines is a tangent line!



For $T \in \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$, the *border rank of T* $\underline{\mathbf{R}}(T)$ denotes the smallest r such that T is a limit of tensors of rank r.
Theorem (Bini 1980) $\omega = \inf_\tau \{\underline{\mathbf{R}}(M_{\langle n \rangle}) = O(n^\tau)\}$, so border rank is also a legitimate complexity measure.

Wider geometric perspective

Let $X \subset \mathbb{CP}^M$ be a projective variety.

Our case: $M = N^3 - 1$,
$X = Seg(\mathbb{P}^{N-1} \times \mathbb{P}^{N-1} \times \mathbb{P}^{N-1}) \subset \mathbb{P}(\mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N)$.

Stratify $\mathbb{CP}^M$ by a sequence of nested varieties

$$X \subset \sigma_2(X) \subset \sigma_3(X) \subset \cdots \subset \sigma_f(X) = \mathbb{CP}^M$$

where

$$\sigma_r(X) := \overline{\cup_{x_1,\ldots,x_r \in X} \mathrm{span}\{x_1, \ldots, x_r\}}$$

is the variety of secant $\mathbb{P}^{r-1}$'s to X.

Secant varieties have been studied for a long time.

In 1911 Terracini could have predicted Strassen's discovery:
$\sigma_7(Seg(\mathbb{P}^3 \times \mathbb{P}^3 \times \mathbb{P}^3)) = \mathbb{P}^{63}$.

How to *disprove* astounding conjecture?

Let $\sigma_r = \hat{\sigma}_r(Seg(\mathbb{P}^{N-1} \times \mathbb{P}^{N-1} \times \mathbb{P}^{N-1})) \subset \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N = \mathbb{C}^{N^3}$ tensors of border rank at most r.

Find a polynomial P (in N^3 variables) in the ideal of σ_r, i.e., such that $P(T) = 0$ for all $T \in \sigma_r$.

Show that $P(M_{\langle n \rangle}) \neq 0$.

Embarassing (?): had not been known even for $M_{\langle 2 \rangle}$, i.e., for σ_6 when $N = 4$.

Arora and Barak: lower bounds are "**complexity theory's Waterloo** "

Why did I think this would be easy?: Representation Theory

Matrices of rank at most r: zero set of size $r+1$ minors.

Tensors of border rank at most 1: zero set of size 2 minors of flattenings tensors to matrices: $A \otimes B \otimes C = (A \otimes B) \otimes C$.

Tensors of border rank at most 2: zero set of degree 3 polynomials.

Representation theory: systematic way to search for polynomials.

2004 L-Manivel: No polynomials in ideal of σ_6 of degree < 12

2013 Hauenstein-Ikenmeyer-L: No polynomials in ideal of σ_6 of degree < 19. However there are polynomials of degree 19. Caveat: too complicated to evaluate on $M_{\langle 2\rangle}$. Good news: easier polynomial of degree 20 (trivial representation) $\rightsquigarrow$ (L 2006, Hauenstein-Ikenmeyer-L 2013) $\underline{\mathbf{R}}(M_{\langle 2\rangle}) = 7$.

Rank methods

A matrix has rank $\leq r$ iff all size $(r+1)\times(r+1)$ minors are zero.

Classical $A\otimes B\otimes C = (A\otimes B)\otimes C$ border rank $> r$ iff $\exists$ nonzero size $r+1$ minor of flattening.

Only gives weak bounds

To go further, embed $A\otimes B\otimes C$ in larger matrix space:

e.g., (L-Ottaviani 2013) $A\otimes B\otimes C \subset (\Lambda^p A^*\otimes B)\otimes(\Lambda^{p+1}A\otimes C)$ (*)

and take minors.

Results using rank methods

Strassen 1983: $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \frac{3}{2}\mathbf{n}^2$.
Lickteig 1985: $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \frac{3}{2}\mathbf{n}^2 + \frac{\mathbf{n}}{2} - 1$
1985-2012: no further progress other than $M_{\langle 2 \rangle}$

(*) $\rightsquigarrow$ (L-Ottaviani 2013) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq 2\mathbf{n}^2 - \mathbf{n}$

For those familiar with representation theory: (*) found via a $G = GL(A) \times GL(B) \times GL(C)$ module map from $A \otimes B \otimes C$ to a space of matrices (systematic search possible). For those familiar with alg. geom.: (*) natural from vector bundle map perspective (think Koszul)

Punch line: Found equations by exploiting symmetry of σ_r

Bad News: Barriers

Theorem (Bernardi-Ranestad,Buczynski-Galcazka,Efremenko-Garg-Oliviera-Wigderson): Game essentially over for rank (determinantal) methods.

For the experts: Variety of zero dimensional schemes of length r is not irreducible $r > 7$.

Determinantal methods detect zero dimensional schemes (want zero dimensional smoothable schemes).

$$\sigma_r(X) := \overline{\bigcup \{\langle R\rangle \mid \mathrm{length}(R) = r,\ R \subset X,\ R : \mathrm{smoothable}\}}$$

secant variety.
Let

$$\kappa_r(X) := \overline{\bigcup \{\langle R\rangle \mid \mathrm{length}(R) = r,\ R \subset X\}}$$

cactus variety.

Bad News cont'd

Determinantal equations are equations for the cactus variety.

$\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)) = \mathbb{P}(A \otimes B \otimes C)$ when $r \sim \frac{m^2}{3}$

$\kappa_r(Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)) = \mathbb{P}(A \otimes B \otimes C)$ when $r \sim 6m$

Punch line: **Barrier** to progress.

How to go further?

So far, lower bounds via symmetry of σ_r.

The matrix multiplication tensor also has symmetry:

$T \in A \otimes B \otimes C$, define *symmetry group of T*
$G_T := \{g \in GL(A) \times GL(B) \times GL(C) \mid g \cdot T = T\}$

$GL_{\mathbf{n}}^{\times 3} \subset G_{M_{\langle n \rangle}} \subset GL_{\mathbf{n}^2}^{\times 3} = GL(A) \times GL(B) \times GL(C)$:

Proof: $(g_1, g_2, g_3) \in GL_{\mathbf{n}}^{\times 3}$

$$\mathrm{trace}(XYZ) = \mathrm{trace}((g_1 X {g_2}^{-1})(g_2 Y {g_3}^{-1})(g_3 Z {g_1}^{-1}))$$

How to exploit G_T?

Given $T \in A{\otimes}B{\otimes}C$
$\underline{\mathbf{R}}(T) \leq r \Leftrightarrow \exists$ curve $E_t \subset G(r, A{\otimes}B{\otimes}C)$ such that
i) For $t \neq 0$, E_t is spanned by r rank one elements.
ii) $T \in E_0$.

For all $g \in G_T$, gE_t also works.
$\rightsquigarrow$ (L-Michalek 2017) can insist on normalized curves (those with E_0 Borel fixed).

$\rightsquigarrow$ $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq 2\mathbf{n}^2 - \log_2 \mathbf{n} - 1$

More bad news: this method cannot go much further.

New idea: Buczyńska-Buczyński

Use more algebraic geometry: Consider not just curve of r points, but the curve of **ideals** $I_t \in Sym(A^* \oplus B^* \oplus C^*)$ it gives rise to: *border apolarity method*

$T = \lim_{t\to 0} \sum_{j=1}^{r} T_{j,t}$

I_t ideal of $[T_{1,t}] \cup \cdots \cup [T_{r,t}] \subset \mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C$

Can consider limiting ideal but how to take limits?

Answer: Haiman-Sturmfels multigraded Hilbert scheme lives in a product of Grassmannians.

Moreover: Can insist that limiting ideal I_0 is Borel fixed: reduces to small search in each multi-degree.

Instead of single curve $E_t \subset G(r, A{\otimes}B{\otimes}C)$ limiting to Borel fixed point, for each (i,j,k) get curve in $Gr(r, S^iA^*{\otimes}S^jB^*{\otimes}S^kC^*)$, each limiting to Borel fixed point *and* satisfying compatibility conditions.

Upshot: algorithm that either produces all normalized candidate I_0's or proves border rank $> r$.

Border apolarity results: Conner-Harper-L 2019:

$\rightsquigarrow$ very easy computer free algebraic proof $\underline{\mathbf{R}}(M_{\langle 2\rangle}) = 7$

Recall: Strassen $\underline{\mathbf{R}}(M_{\langle 3\rangle}) \geq 14$, L-Ottaviani $\underline{\mathbf{R}}(M_{\langle 3\rangle}) \geq 15$, L-Michalek $\underline{\mathbf{R}}(M_{\langle 3\rangle}) \geq 16$.

Thm. $\underline{\mathbf{R}}(M_{\langle 3\rangle}) \geq 17$

Recall: so far only $\underline{\mathbf{R}}(M_{\langle 2\rangle})$ known among nontrivial matrix multiplication tensors.

Thm. $\underline{\mathbf{R}}(M_{\langle 223\rangle}) = 10$ and Thm. $\underline{\mathbf{R}}(M_{\langle 233\rangle}) = 14$

Thm. For all $\mathbf{n} > 25$, $\underline{\mathbf{R}}(M_{\langle 2\mathbf{n}\mathbf{n}\rangle}) \geq \mathbf{n}^2 + 1.32\mathbf{n} + 1$.
Thm. For all $\mathbf{n} > 14$, $\underline{\mathbf{R}}(M_{\langle 3\mathbf{n}\mathbf{n}\rangle}) \geq \mathbf{n}^2 + 2\mathbf{n}$.

First significant lower bound results for *any* "unbalanced" tensor. Also good for other tensors, e.g. Thm: $\underline{\mathbf{R}}(\det_3) = 17$ important for study of upper bounds on the exponent (another lecture)

Bad News: Still have the barrier

Bad news: off the shelf border apolarity gives determinantal equations— subject to barrier.

Conner-Huang-L 2020: augmentations that enable extensions of the method,

$\rightsquigarrow$ additional results. In particular:

$\underline{\mathbf{R}}(\mathrm{perm}_3) = 16$ (Conner-Huang-L). 2020
important for Strassen's laser method for upper bounds on the exponent of matrix multiplication solved CS problem open since 1988 another lecture – paths to overcome *upper bound* barriers via geometry and representation theory

Path to overcome the lower bound barrier

Effective implementation of deformation theory: allows one to determine if an ideal is a limit of ideals of smooth schemes.

May 2022: with student Arpan Pal and Joachim Jelisiejew, small example of this to determine border rank $\rightsquigarrow$

First effective implementation of deformation theory in the study of tensors.

Path to overcome lower bound barriers!

Thank you for your attention

For more on **tensors**, their geometry and applications, resp. **geometry and complexity**, resp. **recent developments**: