Introduction to the Geometry of Tensors Part 2:

The complexity of matrix multiplication

J.M. Landsberg

Texas A&M University and Univ. Toulouse (chaire d'excellence)

Supported by NSF grants AF-1814254 and AF-2203618

Outline

1. Strassen's spectacular failure
2. Geometric formulation of the problem
3. Bini's sleepless nights
4. lower bounds, barriers, and a path to overcome them
5. upper bounds, barriers, and 2 paths to overcome them

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 1979, 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,

2011-21 Stouthers,Vasilevska-Williams,LeGall, Alman and Vasilevska-Williams exponent improved by .004.

The matrix multiplication tensor

Set $N = n^2$.

Matrix multiplication is a bilinear map

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

In other words

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

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

Strassen's algorithm as a rank expression

Rank one tensors correspond to bilinear maps that can be computed using one scalar multiplication.

The rank of a tensor T is essentially the number of scalar multiplications needed to compute the corresponding bilinear map.

standard presentation is $M_{\langle n\rangle} = \sum_{i,j,k=1}^{n} x^i_j \otimes y^j_k \otimes z^k_i$

Strassen's presentation is

$$\begin{aligned}
M_{\langle 2\rangle} = & x^1_1 \otimes y^1_1 \otimes z^1_1 \\
& + (-x^1_2 + x^2_1 - x^2_2) \otimes (-y^1_2 + y^2_1 - y^2_2) \otimes (-z^1_2 + z^2_1 - z^2_2) \\
& + (x^1_2 + x^2_2) \otimes (y^1_2 + y^2_2) \otimes (z^1_2 + z^2_2) \\
& + (-x^2_1 + x^2_2) \otimes (-y^2_1 + y^2_2) \otimes (-z^2_1 + z^2_2) \\
& + \mathbb{Z}_3 \cdot [x^1_2 \otimes y^2_1 \otimes (z^1_1 - z^1_2 + z^2_1 - z^2_2)]
\end{aligned}$$

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

$\omega = 2$

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.

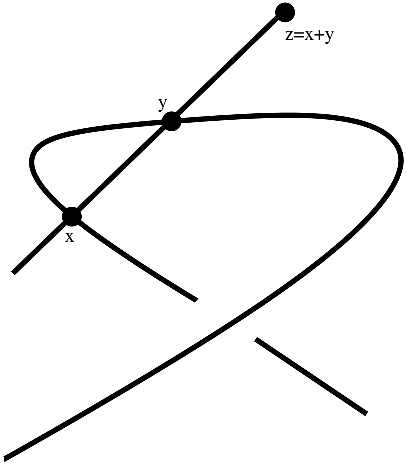
They used numerical methods.

Their code appeared to have a problem.

Recall our picture

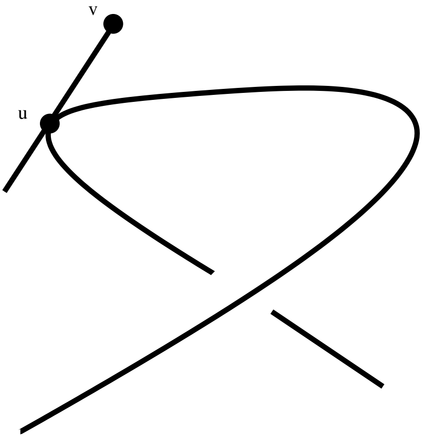
$\{$ tensors of rank two$\}$ $=$

$\{$ points on a secant line to set of tensors of rank one$\}$



Tensors of rank 5: points on a secant 5-plane

Recall our second picture



Theorem (Bini 1980) $\underline{\mathbf{R}}(M_{\langle n\rangle}) = O(n^{\omega})$, so border rank is also a legitimate complexity measure.

How to *disprove* astounding conjecture?

Let $\sigma_r \subset \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N = \mathbb{C}^{N^3}$: 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 I thought this would be easy (review)

Consider rank at most r matrices:
$\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B)) = \{[T] \mid \underline{\mathbf{R}}(T) \leq r\}$

Invariant under changes of bases $\Rightarrow$ its ideal
$I_{\sigma_r(Seg(\mathbb{P}A \times \mathbb{P}B))} \subset Sym(A^* {\otimes} B^*)$ invariant under changes of bases

Special case: rank one - saw matrix has rank one iff size two minors zero. Degree two polynomials.

Recall homogeneous degree two polynomials on matrices:

$S^2(A^* {\otimes} B^*) = S^2 A^* {\otimes} S^2 B^* \oplus \Lambda^2 A^* {\otimes} \Lambda^2 B^*$

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 less than 12

2013 Hauenstein-Ikenmeyer-L: No polynomials in ideal of σ_6 of degree less than 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) $\underline{\mathbf{R}}(M_{\langle 2\rangle}) = 7$.

Polynomials via a retreat to linear algebra

$T \in A \otimes B \otimes C = \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$ may be recovered from the linear space $T(C^*) \subset A \otimes B$.

tensors up to changes of bases $\sim$ linear subspaces of spaces of matrices up to changes of bases.

Idea (Strassen 1983, E. Toeplitz 1877): instead of tensor, work with N-dimensional space of $N \times N$ matrices.

$\rightsquigarrow$ (Strassen 1983) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \frac{3}{2}\mathbf{n}^2$

Variant: (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 for $M_{\langle 2 \rangle}$.

Retreat to linear algebra, cont'd

Perspective: Strassen mapped space of tensors to space of matrices, found equations by taking minors.

Classical trick in algebraic geometry to find equations via minors.

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

These equations were found using *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).

Punch line: Found equations by exploiting symmetry of σ_r

Bad News: Barriers

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

For the experts: Determinantal methods detect zero dimensional schemes (want zero dimensional smoothable schemes).

Spans of zero dimensional (local) schemes of length $6m$ on Segre fill ambient space. (Bernardi-Ranestad+Buczynski)

In particular, cannot use to show $\underline{\mathbf{R}}(T) > 6m$.

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:

$T \in A \otimes B \otimes C$, the *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: $(g_1, g_2, g_3) \in GL_{\mathbf{n}}^{\times 3}$

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

$\rightsquigarrow$ breakthrough: new lower bounds by exploiting G_T (L-Michalek, **Buczyńska-Buczyński**, Conner-Harper-L, Conner-Huang-L)

Recent breakthroughs: new lower bounds by exploiting G_T

first breakthrough $\rightsquigarrow$ (L-Michalek 2017)
$\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq 2\mathbf{n}^2 - \log_2 \mathbf{n} - 1$

Open: Hay in a haystack: A random tensor in $\mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$ has border rank $\sim \frac{m^2}{3}$. Find an *explicit* sequence of tensors of border rank $m^{1+\epsilon}$.

previously: $2m - \sqrt{m}$ (L-Ottaviani, 2013)

first breakthrough $\rightsquigarrow$ (L-Michalek, 2020) $(2.02)m$

BB-breakthrough $\rightsquigarrow$ (Conner-Harper-L 2020) short proof of border rank of $M_{\langle 2\rangle}$ and determined border rank of $M_{\langle 2,2,3\rangle}$ and $M_{\langle 2,2,4\rangle}$

Work in "Haiman-Sturmfels multi-graded Hilbert scheme" allows unsaturated ideals.

Next step: use *deformation theory* to break lower bound barriers. (Implemented with Jelisiejew and Pal in $m = 5$ case last lect.)

Upper bounds

Ideal method: Upper bound $\underline{\mathbf{R}}(M_{\langle n\rangle})$ directly.

So far little success.

Idea: work indirectly, utilize combinatorics and **probability** (following Shannon/Erdös).

work of Schönhage, then Strassen $\rightsquigarrow$ "Strassen's laser method":

Recall definitions which (ahistorically) appeared in quantum information theory:

- $T \in A\otimes B\otimes C$, $T' \in A'\otimes B'\otimes C'$, *Kronecker product* $T \boxtimes T' \in (A\otimes A')\otimes(B\otimes B')\otimes(C\otimes C')$, and Kronecker powers $T^{\boxtimes k} \in (A^{\otimes k})\otimes(B^{\otimes k})\otimes(C^{\otimes k})$
- Say T *degenerates* to T' if $T' \in \overline{GL(A) \times GL(B) \times GL(C) \cdot T}$. In this case $\underline{\mathbf{R}}(T') \leq \underline{\mathbf{R}}(T)$.

Upper bounds

Start with tensor T where $\underline{\mathbf{R}}(T)$ is minimal or close to minimal and with "nice" combinatorial structure.

Show for large k, there **exists** a degeneration of $T^{\boxtimes k}$ to $M_{\langle n \rangle}$ for some large n.

(in fact certain restricted "random" degenerations)

Since have upper bound on $\underline{\mathbf{R}}(T^{\boxtimes k})$, get upper bound on $\underline{\mathbf{R}}(M_{\langle n \rangle})$.

Responsible for all upper bounds since 1987.

Barriers to Upper bounds

Ambanius-Filimus-LeGall 2014: can never prove $\omega < 2.3$ with favorite tensor "big CW"

Much follow-up work $\rightsquigarrow$ Christandl-Vrana-Zuiddam: geometric explanation, recall

$\underset{\sim}{\mathbf{R}}(T) := \lim_{N\to\infty}(\underline{\mathbf{R}}(T^{\boxtimes N}))^{\frac{1}{N}}$, $\underset{\sim}{\mathbf{Q}}(T) := \lim_{N\to\infty}(\underline{\mathbf{Q}}(T^{\boxtimes N}))^{\frac{1}{N}}$

if $\underset{\sim}{\mathbf{R}}(T)/\underset{\sim}{\mathbf{Q}}(T) > 1$ then cannot prove $\omega = 2$ using T in laser method also quantitative limits

$T \in \mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$, $\underset{\sim}{\mathbf{R}}(T)$ not known unless $\underline{\mathbf{R}}(T) = m$ (recall $m \leq \underset{\sim}{\mathbf{R}}(T) \leq \underline{\mathbf{R}}(T)$)

$\underset{\sim}{\mathbf{Q}}(T)$ known for special tensors (method from celestial mechanics - Strassen), additional methods via auxiliary geometric quantites (see lect. 1)

$\rightsquigarrow$ two strategies

Strategy 1:

Start with T with $\underset{\sim}{\mathbf{Q}}(T) = m$, prove upper bound on $\underset{\sim}{\mathbf{R}}(T)$ via $\underline{\mathbf{R}}(T)$ or perhaps $\underline{\mathbf{R}}(T^{\boxtimes 2})$

Ex. (Coppersmith-Winograd1988) $cw_2 \in \mathbb{C}^3 \otimes \mathbb{C}^3 \otimes \mathbb{C}^3$ has $\underset{\sim}{\mathbf{Q}}(cw_2) = 3$. known if $\underset{\sim}{\mathbf{R}}(cw_2)$ is 3, then $\omega = 2$ also known

$\underline{\mathbf{R}}(cw_2) = 4$

$\rightsquigarrow$ **1988 Determine $\underline{\mathbf{R}}(cw_2^{\boxtimes 2})$**

(2021, Conner-Huang-L, using BB-breakthrough+)
$\underline{\mathbf{R}}(cw_2^{\boxtimes 2}) = 16 = 4^2$ (Bad news for laser method)

Ex. (Conner-Gesmundo-Ventura 2019) $skewcw_2 \in \mathbb{C}^3 \otimes \mathbb{C}^3 \otimes \mathbb{C}^3$ has $\underset{\sim}{\mathbf{Q}}(skewcw_2) = 3$. and if $\underset{\sim}{\mathbf{R}}(skewcw_2)$ is 3, then $\omega = 2$ also showed

$\underline{\mathbf{R}}(skewcw_2) = 5 > 4$

(2020, Conner-Harper-L, using breakthrough)
$\underline{\mathbf{R}}(skewcw_2^{\boxtimes 2}) = 17 << 5^2$ (Hopeful news for laser method)

Strategy 2:

Start with T with $\underline{\mathbf{R}}(T) = m$.

Explains motivation for lect. 1: **Open**: Classify concise tensors of minimal border rank. (state of art: March 2022 Jelisiejew-L-Pal $m \leq 5$)

Work in progress (with Alman, Conner, and Vasilevska-Williams): are any new tensors better than CW?

goal prove ω less than 2.3, as useful minimal border rank tensors have barriers i.e., $\underset{\sim}{\mathbf{Q}}(T) < 1$.

Thank you for your attention

For more on **tensors**, their geometry and applications, resp. **geometry and complexity**, resp. **asymptotic geometry, moment maps, (quantum) information theory...** :