On the geometry of matrix multiplication

J.M. Landsberg

Texas A&M University

Spring 2018 AMS Sectional Meeting

A practical problem: efficient linear algebra

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.

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.

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!*

Tensor formulation of conjecture

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

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

i.e., an element of

$$\mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$$

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

Tensor formulation of conjecture

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.

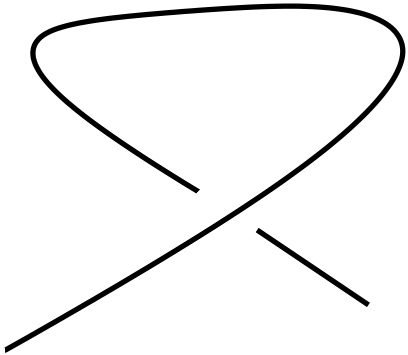
Theorem (Strassen): $\mathbf{R}(M_{\langle n \rangle}) = O(n^{\omega})$ where $M_{\langle n \rangle}$ can be computed using $O(n^{\omega})$ arithmetic operations. ω is called the *exponent* of matrix multiplication.

Astounding Conjecture

$\omega = 2$

Progress: Strassen 1969,Bini et. al. 1979, Schönhage 1981,Strassen 1987,Coppersmith-Winograd 1989 $\rightsquigarrow \omega \leq 2.38$
Stouthers, Williams, LeGall 2011-2014: $\omega \leq 2.3755$.

Geometric formulation of conjecture



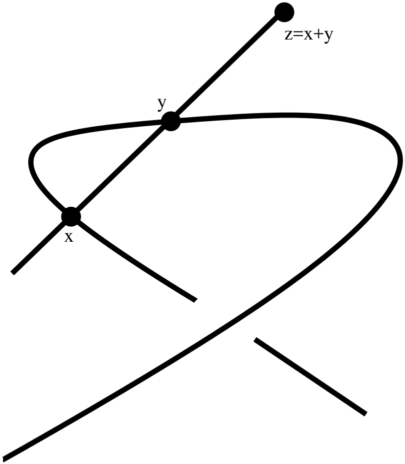
Imagine this curve represents the set of tensors of rank one.

Geometric formulation of conjecture

$\{$ tensors of rank at most two$\}$

$=$

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



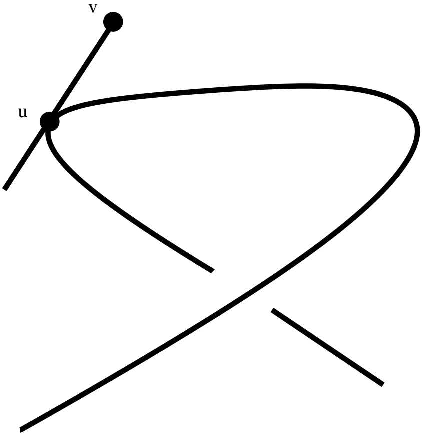
Bini's sleepless nights

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

Wider geometric perspective

Let $X \subset \mathbb{CP}^M$ be a projective variety. Stratify $\mathbb{CP}^M$ by a sequence of nested spaces

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

where

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

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

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

Secant varieties have been studied for a long time.

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

Border rank Lower bounds

- (Classical) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \mathbf{n}^2$.

Idea of proof: $T \in \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N \rightsquigarrow T_{\mathbb{C}^N} : \mathbb{C}^{N*} \to \mathbb{C}^N \otimes \mathbb{C}^N$ and $\underline{\mathbf{R}}(T) \geq \operatorname{rank}(T_{\mathbb{C}^N})$.

Border rank Lower bounds

- (Classical) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq \mathbf{n}^2$.
- (Strassen 1983) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq \frac{3}{2}\mathbf{n}^2$
- (Lickteig 1985) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq \frac{3}{2}\mathbf{n}^2 + \frac{\mathbf{n}}{2} - 1$
- (L 2006, Hauenstein-Ikenmeyer-L 2013) $\underline{\mathbf{R}}(M_{\langle 2\rangle}) = 7$.

Hauenstein-Ikenmeyer-L proof: How to find equations for $\sigma_r(X)$?- representation theory

$Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$ is homogeneous for $G = GL(A) \times GL(B) \times GL(C)$.

For any G-variety $Z \subset \mathbb{P}V_\lambda$, its ideal will be a G-module, so one should not look for individual polynomials, but G-modules of polynomials.

Can do systematically in small cases $\rightsquigarrow$ H-I-L proof

Border rank Lower bounds

- (Classical) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq \mathbf{n}^2$.
- (Strassen 1983) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq \frac{3}{2}\mathbf{n}^2$
- (Lickteig 1985) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq \frac{3}{2}\mathbf{n}^2 + \frac{\mathbf{n}}{2} - 1$
- (L 2006, Hauenstein-Ikenmeyer-L 2013) $\underline{\mathbf{R}}(M_{\langle 2\rangle}) = 7$.
- (L-Ottaviani 2013) $\underline{\mathbf{R}}(M_{\langle n\rangle}) \geq 2\mathbf{n}^2 - \mathbf{n}$

Strassen, L-Ottaviani proofs: Determinantal equations

Idea (L-O): look for G-modules V_μ, V_ν where there exists a G-module inclusion $i : V_\lambda \to V_\mu \otimes V_\nu$. Then, for $p \in V_\lambda$, $x \in X$,

$$\underline{\mathbf{R}}(p) \geq \frac{\operatorname{rank}(i(p))}{\operatorname{rank}(i(x))}.$$

classical proof case: $V_\lambda = A \otimes B \otimes C = \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$, $V_\mu = A$, $V_\nu = B \otimes C$.
L-O proof case: $V_\lambda = A \otimes B \otimes C = \mathbb{C}^N \otimes \mathbb{C}^N \otimes \mathbb{C}^N$,

$V_\mu = \Lambda^p A^* \otimes B$, $V_\nu = \Lambda^{p+1} A \otimes C$.

$p = n << \frac{n^2}{2}$

Border rank Lower bounds

- (Classical) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \mathbf{n}^2$.
- (Strassen 1983) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \frac{3}{2}\mathbf{n}^2$
- (Lickteig 1985) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq \frac{3}{2}\mathbf{n}^2 + \frac{\mathbf{n}}{2} - 1$
- (L 2006, Hauenstein-Ikenmeyer-L 2013) $\underline{\mathbf{R}}(M_{\langle 2 \rangle}) = 7$
- (L-Ottaviani 2013) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq 2\mathbf{n}^2 - \mathbf{n}$
- (L-Michalek 2017) $\underline{\mathbf{R}}(M_{\langle n \rangle}) \geq 2\mathbf{n}^2 - \log_2 \mathbf{n} - 1$

Idea of L-Michalek proof: exploit $G_{M_{\langle n \rangle}}$-orbits

$M_{\langle n \rangle}$ also has symmetry:
As a trilinear map

$$M_{\langle n \rangle}(X, Y, Z) = \mathrm{trace}(XYZ)$$

and for $g \in GL_n$

$$\mathrm{trace}(XYZ) =$$

$$\mathrm{trace}(YZX) = \mathrm{trace}(Z^T Y^T X^T) = \mathrm{trace}((gX)Y(Zg^{-1})) = \mathrm{etc...}$$

$$G_{M_{\langle n \rangle}} = PGL_n^{\times 3} \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$$

Exploit structure of $G_{M_{\langle n \rangle}}$ orbits on the Segre to “degenerate” the L-Ottaviani method.

Game over?

$$\sigma_r(X) = \overline{\cup_R \langle R \rangle}$$

union over $R \subset X$: zero dimensional smoothable subschemes length r.

Consider the *cactus variety*

$$\kappa_r(X) = \overline{\cup_R \langle R \rangle}$$

union over $R \subset X$: zero dimensional subschemes length r.
Work of Bernardi-Ranestad: cactus variety fills when r is small (our case, linear in N)

Work of Buczynski-Galazka: determinantal equations are equations for the cactus variety

$\rightsquigarrow$

Determinantal techniques will never prove $\underline{\mathbf{R}}(M_{\langle n \rangle}) > 6n^2$.

Perhaps try to prove conjecture?

Strassen's algorithm revisited

• [Chiantini-Ikenmeyer-L-Ottaviani 2017, Burichenko 2014]: Strassen's optimal decomposition has $\mathfrak{S}_3 \rtimes (\mathbb{Z}_3 \rtimes \mathbb{Z}_2)$ symmetry, where $\mathfrak{S}_3 \subset PGL_2 \subset PGL_2^{\times 3}$.

$$M_{\langle 2 \rangle} = \mathrm{Id}_2^{\otimes 3} + \mathbb{Z}_3 \rtimes \mathbb{Z}_2 \cdot \left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \right).$$

Work in progress (Ballard-Conner-Ikenmeyer-L-Ryder): look for matrix multiplication decompositions with symmetry.

In particular, cyclic $\mathbb{Z}_3$ symmetry.

Symmetry v. Optimality

The smallest known decomposition of $M_{\langle 3 \rangle}$ is of size 23 (Laderman, 1973).

We found rank 23 decompositions with extra symmetry.

A decomposition with $\mathbb{Z}_4 \times \mathbb{Z}_3$-symmetry

$$
\begin{aligned}
M_{\langle 3 \rangle} = &-\begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}^{\otimes 3} \\
&+ \mathbb{Z}_4 \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}^{\otimes 3} \\
&+ \mathbb{Z}_4 \cdot \begin{pmatrix} 0 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}^{\otimes 3} \\
&+ (\mathbb{Z}_4/\mathbb{Z}_2) \cdot \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}^{\otimes 3} \\
&+ \mathbb{Z}_3 \times \mathbb{Z}_4 \cdot \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 & -1 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

What next?

Other rank decompositions: Johnson-McPharlane (families for $M_{\langle 3 \rangle}$), Pan (rank $143,240$ decomposition of $M_{\langle 70 \rangle}$), Alexeev-Smirnov (rank 40 decomposition of $M_{\langle 3,3,6 \rangle}$ and others).
Work in progress: 4×4 and beyond.

Algebraic geometry

$\mathbb{Z}_3$-symmetry implies $M_{\langle n\rangle} = sM_{\langle n\rangle} \oplus \wedge M_{\langle n\rangle}$ with

$sM_{\langle n\rangle} \in S^3(\mathbb{C}^{\mathbf{n}^2})$, $\wedge M_{\langle n\rangle} \in \Lambda^3(\mathbb{C}^{\mathbf{n}^2})$.

Theorem (Chiantini-Hauenstein-L-Ottaviani-Ikenmeyer, 2017)

$\underline{\mathbf{R}}_S(sM_{\langle n\rangle}) = O(\mathbf{n}^\omega)$ *and* $\mathbf{R}_S(sM_{\langle n\rangle}) = O(\mathbf{n}^\omega)$.

Here $\mathbf{R}_S(P)$ *is the smallest* r *such that* $P = z_1^3 + \cdots + z_r^3$ *and* $\underline{\mathbf{R}}_S(P)$ *is smallest* r *such that* $P \in \sigma_r(v_3(\mathbb{P}^{N-1}))$ *(Waring rank and Waring border rank).*

Cubic polynomial!

linear and quadratic polynomials: know everything!

An amazing rank 18 Waring decomposition of $sM_{\langle 3 \rangle}$

Let $\Gamma = (\mathbb{Z}_3^2 \rtimes SL(2, \mathbb{F}_3)) \rtimes \mathbb{Z}_2$, which has order 432.
The group Γ acts on the following configuration of 9 points in the projective plane:

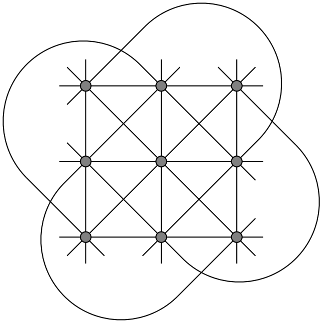


Figure: Hasse configuration. Depiction by D. Eppstein

An amazing Waring rank 18 decomposition of $sM_{\langle 3\rangle}$

Theorem (A. Conner 2017):

$$sM_{\langle 3\rangle} = -\frac{1}{2}\,\mathrm{Id}^{\otimes 3}$$

$$\Gamma \cdot \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}^{\otimes 3}$$

$$+\,\Gamma \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{\frac{2\pi i}{3}} & 0 \\ 0 & 0 & e^{-\frac{2\pi i}{3}} \end{pmatrix}^{\otimes 3}$$

$\Gamma = (\mathbb{Z}_3^2 \rtimes SL(2,\mathbb{F}_3)) \rtimes \mathbb{Z}_2$
The first orbit consists of 9 rank one matrices corresponding to the points in the plane in the picture, and the second orbit consists of 8 rank three matrices.

Thank you for your attention

For more on geometry and complexity: