Tensors of minimal border rank I

J.M. Landsberg
JLP=J. Jelisiejew -L- A. Pal Math. Ann. 2023
GJLM=F. Gesmundo-J-L-T. Mandziuk: work in progress

Owen Professor of Mathematics, Texas A&M University

Supported by NSF grant 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 $A = \begin{pmatrix} a^1_1 & a^1_2 \\ a^2_1 & a^2_2 \end{pmatrix}, \quad B = \begin{pmatrix} b^1_1 & b^1_2 \\ b^2_1 & b^2_2 \end{pmatrix}$. Set

$$\begin{aligned}
I &= (a^1_1 + a^2_2)(b^1_1 + b^2_2),\\
II &= (a^2_1 + a^2_2)b^1_1,\\
III &= a^1_1(b^1_2 - b^2_2)\\
IV &= a^2_2(-b^1_1 + b^2_1)\\
V &= (a^1_1 + a^1_2)b^2_2\\
VI &= (-a^1_1 + a^2_1)(b^1_1 + b^1_2),\\
VII &= (a^1_2 - a^2_2)(b^2_1 + b^2_2),
\end{aligned}$$

$$\begin{aligned}
\text{If } C = AB, \text{ then } c^1_1 &= I + IV - V + VII,\\
c^2_1 &= II + IV,\\
c^1_2 &= III + V,\\
c^2_2 &= I + III - II + VI.
\end{aligned}$$

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 Stothers, Vasilevska-Williams, LeGall, 2021 2023 Alman and V-W .004 improvement.

Tensor formulation of conjecture

Set $N = n^2$.

Matrix multiplication is a *bilinear* map

$$M_{\langle \mathbf{n} \rangle} : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{C}^N,$$
$$(X, Y) \mapsto XY$$

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

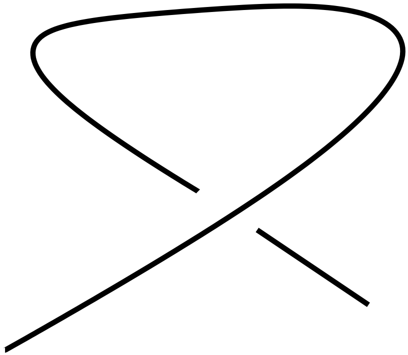
$\mathbb{C}^N \times \mathbb{C}^N \times \mathbb{C}^{N*} \to \mathbb{C}$.

In other words

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

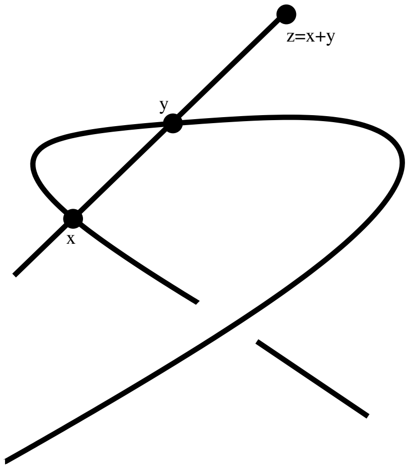
Exercise: As a trilinear map, $M_{\langle \mathbf{n} \rangle}(X, Y, Z) = \mathrm{trace}(XYZ)$.

Geometry Detour: subsets $\rightsquigarrow$ stratifications



Let $X \subset \mathbb{C}^M$ be a subset not contained in a linear subspace.

secant lines



Say X curve $\rightsquigarrow$ 3-dimensional subset.

Can also take secant 2-planes, expect 5-dimensional subset etc....

Geometric definition of rank

Ex. $\mathbb{C}^M = A \otimes B = \mathbb{C}^m \otimes \mathbb{C}^m$, $X = Seg(\mathbb{P}A \times \mathbb{P}B)$: rank one matrices

$\rightsquigarrow$ matrix rank.

Ex. $\mathbb{C}^M$ = space of homogenous polynomials of degree d in **a** variables, $X = v_d(\mathbb{P}A)$: d-th powers $\rightsquigarrow$ Waring rank

Ex. $\mathbb{C}^M = A \otimes B \otimes C = \mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$, $X = Seg(\mathbb{P}A \times \mathbb{P}B \times \mathbb{P}C)$: rank one tensors

$\rightsquigarrow$ tensor rank for $T \in A \otimes B \otimes C$

$\mathbf{R}_X(T)$ is smallest r, such that T lies on a secant r-plane to X.

Tensor formulation of conjecture

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

The tensor rank $\mathbf{R}(T)$ of $T \in A \otimes B \otimes C$ is essentially the number of scalar multiplications needed to compute the corresponding bilinear map.

Tensor formulation of conjecture

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

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

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

$\omega = 2$

Astounding conjecture is about a point lying on a secant r-plane to the set of rank one tensors, more precisely, a sequence of such.

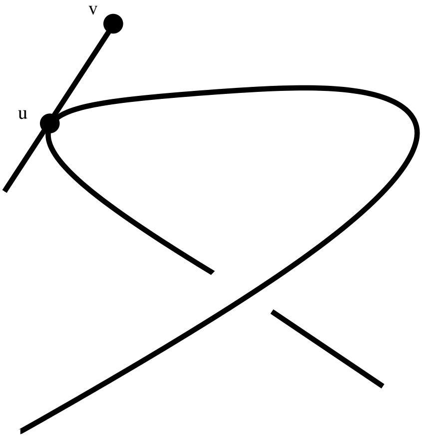
Bini's sleepless nights

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

How to prove upper bounds on ω?

Want to prove upper bounds on $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle})$ when n is large.

To do so need language:

Given $T \in A \otimes B \otimes C$ define its *Kronecker powers*

$T^{\boxtimes N} \in (A^{\otimes N}) \otimes (B^{\otimes N}) \otimes (C^{\otimes N})$,

its tensor powers considered as a 3-way tensor.

Exercise: $M_{\langle \mathbf{n} \rangle}^{\boxtimes N} = M_{\langle n^N \rangle}$.

Exercise: $\underline{\mathbf{R}}(T^{\boxtimes N}) \leq \underline{\mathbf{R}}(T)^N$. More generally, let $M_{\langle \ell, \mathbf{m}, \mathbf{n} \rangle}$ denote the rectangular matrix multiplication tensor. Show $M_{\langle \ell, \mathbf{m}, \mathbf{n} \rangle} \boxtimes M_{\langle \ell', \mathbf{m}', \mathbf{n}' \rangle} = M_{\langle \ell\ell', \mathbf{mm}', \mathbf{nn}' \rangle}$.

Define the *asymptotic rank* $\underset{\sim}{\mathbf{R}}(T) = \lim_{N \to \infty} [\underline{\mathbf{R}}(T^{\boxtimes N})]^{\frac{1}{N}}$

Exercise: $\omega = 2$ iff $\underset{\sim}{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) = n$ for any $n \geq 2$ iff $\underset{\sim}{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) = n \forall n \geq 2$

Isomorphism and degeneration

Let $GL(A)$ group of invertible linear maps $A \to A$ (invertible $\mathbf{a} \times \mathbf{a}$ matrices) Let $G = GL(A) \times GL(B) \times GL(C)$ acts on $A \otimes B \otimes C$. When $\mathbf{a} = \mathbf{b} = \mathbf{c}$ instead take $G = GL(A) \times GL(B) \times GL(C) \ltimes \mathfrak{S}_3$, where $\mathfrak{S}_3$ is the permutation group permuting the three factors.

For all $g \in G$, $\underline{\mathbf{R}}(gT) = \underline{\mathbf{R}}(T)$. "border rank is geometric"

Let $S, T \in A \otimes B \otimes C$. Def. T is *isomorphic* to S if $S \in G \cdot T$ equivalently $T \in G \cdot S$.

Def. T *degenerates* to S if $S \in \overline{G \cdot T}$ and write $T \trianglerighteq S$.

Exercise: If $T \trianglerighteq S$, then $\underline{\mathbf{R}}(T) \geq \underline{\mathbf{R}}(S)$.

$\underline{\mathbf{R}}(T)$ is the smallest r such that $M_{\langle 1 \rangle}^{\oplus r} \trianglerighteq T$.

How to prove upper bounds on ω?

Strassen 1987: find a tensor T where have a good upper bound on $\underline{\mathbf{R}}(T)$. Take $T^{\boxtimes N}$ for N large. Then show $T^{\boxtimes N} \trianglerighteq M_{\langle \mathbf{n} \rangle}$ for some large n. Get $\underline{\mathbf{R}}(M_{\langle \mathbf{n} \rangle}) \leq \underline{\mathbf{R}}(T)^N$.

Amazing: this actually works.

Idea: Can have low cost tensors T that have as subtensors several $M_{\langle 1,1,n \rangle}, M_{\langle 1,n,1 \rangle}, M_{\langle n,1,1 \rangle}$. When take $T^{\boxtimes 3}$ get copies of $M_{\langle n,n,n \rangle}$. Tricky part: making garbage go away in degeneration without losing too much good stuff.

Essentially no progress since 1988. 2014 Ambainus-Filmus-LeGall: explanation why no progress. Geometric explanation Christandl-Vrana-Zuiddam 2019.

Concise tensors

A tensor $T \in A \otimes B \otimes C$ is *concise* if the three maps $T_A : A^* \to B \otimes C$, $T_B : B^* \to A \otimes C$, $T_C : C^* \to A \otimes B$, are all injective.

This just says we cannot realize T in a smaller tensor space

Geometric explanation why progress stalled

Def. *border subrank* of T: $\underline{\mathbf{Q}}(T)$ = largest ρ such that $T \trianglerighteq M_{\langle 1\rangle}^{\oplus \rho}$

Def. *Asymptotic subrank of* T: $\underset{\sim}{\mathbf{Q}}(T) := \lim_{k\to\infty}(\underline{\mathbf{Q}}(T^{\boxtimes k}))^{\frac{1}{k}}$.

Note $\underset{\sim}{\mathbf{R}}(T) \geq \underset{\sim}{\mathbf{Q}}(T)$ and if equality holds and T is concise, both equal m.

Thm. (Christandl-Vrana-Zuiddam building on LeGall and Strassen)
If

$$\frac{\underset{\sim}{\mathbf{R}}(T)}{\underset{\sim}{\mathbf{Q}}(T)} > 1$$

then T cannot be used to prove the exponent is two in the laser method. Moreover, quantitiative version: This ratio measures the potential utility of T for the laser method.

State of the art

No technology exists for showing $T \in \mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$ with $\underline{\mathbf{R}}(T) > m$ has $\underset{\sim}{\mathbf{R}}(T) = m$.

However some technology exists for computing $\underset{\sim}{\mathbf{Q}}(T)$ (comes from probability!)

Best upper bounds on ω via "Big Coppersmith-Winograd tensor"

$$CW_q := \sum_{j=1}^{q} a_0 \otimes b_j \otimes c_j + a_j \otimes b_0 \otimes c_j + a_j \otimes b_j \otimes c_0$$
$$+ a_0 \otimes b_0 \otimes c_{q+1} + a_0 \otimes b_{q+1} \otimes c_0 + a_{q+1} \otimes b_0 \otimes c_0 \in \mathbb{C}^{q+2} \otimes \mathbb{C}^{q+2} \otimes \mathbb{C}^{q+2}$$

with $q = 8$.

But $\underset{\sim}{\mathbf{Q}}(CW_6) \sim 6.44 < 8$, $\underset{\sim}{\mathbf{Q}}(CW_7) \sim 7.087 < 9$, $\underset{\sim}{\mathbf{Q}}(CW_8) \sim 7.087 < 10$ etc.. hence the barrier. Cannot prove $\omega < 2.3$ with it.

State of things

Can sometimes compute $\underset{\sim}{\mathbf{Q}}(T)$. Have explicit tensors with $\underset{\sim}{\mathbf{Q}}(T) = m$. But other than $M_{\langle 1\rangle}^{\oplus m}$, all have $\underline{\mathbf{R}}(T) > m$. In fact

Thm. (GJLM, building on Bläser-Lysikov) If $\underline{\mathbf{R}}(T) = m$ then $\underset{\sim}{\mathbf{Q}}(T) < m$. In particular, minimal border rank tensors cannot be used to prove ω is two.

On the other hand, $\nexists$ techniques for computing $\underset{\sim}{\mathbf{R}}(T)$ when $\underline{\mathbf{R}}(T) > m$.

Two strategies for improving upper bounds on ω:

(1) work with tensors with $\underset{\sim}{\mathbf{Q}}(T) = m$ and try to prove upper bounds on $\underline{\mathbf{R}}(T^{\boxtimes k})$ for k small.

(2) these talks: look for concise tensors with $\underline{\mathbf{R}}(T) = m$, called *tensors of minimal border rank*, and $\underset{\sim}{\mathbf{Q}}(T)$ as large as possible.

History

As early as 1997, hoped to find tensors better than CW_q among tensors of minimal border rank, motivating

Problem 15.2 of *Algebraic Complexity Theory* by Bürgisser, Clausen and Shokrollahi is to *classify tensors of minimal border rank*.

Next lecture: Explain how to classification problem splits into four sub-problems. One of which is equivalent to the following difficult problem in algebraic geometry: which elements of the Hilbert scheme parametrizing zero dimensional schemes of length m lie in the smoothable component?

How to find tensors of minimal border rank?

$$M_{\langle 1\rangle}^{\oplus m}(A^*) = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_m \end{pmatrix}$$

in bases, space of diagonal matrices. Use one such of full rank element (e.g. $\alpha := \alpha_1 + \cdots + \alpha_m$) to obtain an isomorphism $M_{\langle 1\rangle}^{\oplus m}(\alpha) : B^* \to C$.

Then $\mathcal{E} = \mathcal{E}_{M_{\langle 1\rangle}^{\oplus m},\alpha} := M_{\langle 1\rangle}^{\oplus m}(A^*) M_{\langle 1\rangle}^{\oplus m}(\alpha)^{-1} \subset \mathrm{End}(C)$ space of simultaneously diagonal endomorphisms. For tensors isomorphic to $M_{\langle 1\rangle}^{\oplus m}$ get simultaneously diagonalizable endomorphisms.

In particular commuting endomorphisms: this is a closed condition, therefore must hold for degenerations of $M_{\langle 1\rangle}^{\oplus m}$

$\rightsquigarrow$

Strassen (1983): first non-classical equations.

Strassen's equations

Let $T \in A \otimes B \otimes C = \mathbb{C}^m \otimes \mathbb{C}^m \otimes \mathbb{C}^m$ be such that there exists $\alpha \in A^*$ with $T(\alpha)$ full rank. (For such say T is 1_A-generic) Then a necessary condition for T to be of minimal border rank is that $\mathcal{E}_{T,\alpha} := T(A^*)T(\alpha)^{-1} \subset \mathrm{End}(C)$ is abelian.

To make into polynomials, instead of inverse, use cofactor matrix. Explicitly, for all $\alpha, \alpha_1, \alpha_2 \in A^*$,

$$[T(\alpha_1)T(\alpha)^{cof}, T(\alpha_2)T(\alpha)^{cof}] = 0$$

where $[X, Y] = XY - YX$ is commutator.

Get equations of degree $2m = 1 + (m-1) + 1 + (m-1)$.

Key point: semi-continuity of matrix rank

Exercise: can lower to degree $m+1$

Similarly, $\mathcal{E}_{T,\alpha}$ must be closed under composition - further equations

Additional non-classical equations: Koszul flattenings

(L-Ottaviani) 2014: Consider $T : B^* \to A \otimes C$,

and $T \otimes \mathrm{Id}_A : A \otimes B^* \to A \otimes A \otimes C$,

and its projection $T^{\wedge 1} : A \otimes B^* \to \Lambda^2 A \otimes C$.

When $T = M_{\langle 1 \rangle}^{\oplus m}$, $\mathrm{rank}(M_{\langle 1 \rangle}^{\oplus m})^{\wedge 1} = m(m-1)$.

By semi-continuity of matrix rank,

$\underline{\mathbf{R}}(T) \leq m$ implies $\mathrm{rank}\, T^{\wedge 1} \leq m(m-1)$

special case of "Koszul flattenings"

Additional non-classical equations: Symmetry Lie algebras

For $T \in A \otimes B \otimes C$ let

$$\widehat{\mathfrak{g}}_T := \{(X,Y,Z) \in \mathfrak{gl}(A) \times \mathfrak{gl}(B) \times \mathfrak{gl}(C) \mid (X,Y,Z).T = 0\}$$

its (extended) symmetry Lie algebra. Lie algebra of its symmetry group.

Example $\mathfrak{g}_{M_{\langle 1 \rangle}^{\oplus m}} =$

$$\left\{ \left(\begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_m \end{pmatrix}, \begin{pmatrix} \nu_1 & & \\ & \ddots & \\ & & \nu_m \end{pmatrix}, \begin{pmatrix} \tau_1 & & \\ & \ddots & \\ & & \tau_m \end{pmatrix} \right) \mid \mu_j + \nu_j + \tau_j = 0 \right\}$$

In particular $\dim \widehat{\mathfrak{g}}_{M_{\langle 1 \rangle}^{\oplus m}} = 2m$

Thus $\underline{\mathbf{R}}(T) \leq m$ implies $\dim \widehat{\mathfrak{g}}_T \geq 2m$

Additional non-classical equations: centroids/111-algebras

JLP 2023: Define the *centroid* or *111-algebra* of T:

$$\mathcal{A}_{111}^{T} := \{(X, Y, Z) \in \mathfrak{gl}(A) \times \mathfrak{gl}(B) \times \mathfrak{gl}(C) \mid X.T = Y.T = Z.T\}$$

Example $\mathcal{A}_{111}^{M_{\langle 1 \rangle}^{\oplus m}} =$

$$\left\{ \left(\begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_m \end{pmatrix}, \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_m \end{pmatrix}, \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_m \end{pmatrix} \right) \right\}$$

In particular $\dim \mathcal{A}_{111}^{M_{\langle 1 \rangle}^{\oplus m}} = m$

$\rightsquigarrow$ T concise with $\underline{\mathbf{R}}(T) = m$ implies $\dim \mathcal{A}_{111}^{T} \geq m$

Turns out to be particularly useful: next time: gives structure to a class of tensors that previously were not known to have any.

Best tensor for laser method so far: Big Coppersmith-Winograd

$$CW_q := \sum_{j=1}^{q} (a_0 \otimes b_j \otimes c_j + a_j \otimes b_0 \otimes c_j + a_j \otimes b_j \otimes c_0)$$

$$+ a_0 \otimes b_0 \otimes c_{q+1} + a_0 \otimes b_{q+1} \otimes c_0 + a_{q+1} \otimes b_0 \otimes c_0 \in \mathbb{C}^{q+2} \otimes \mathbb{C}^{q+2} \otimes \mathbb{C}^{q+2}$$

Recall: Thm. (Ambainus-Filmus- LeGall 2014) CW_q cannot be used to prove $\omega < 2.3$.

(Christandl-Vrana-Zuiddam 2019) Essentially because $\underset{\sim}{\mathbf{Q}}(CW_q)$ is "small". $\rightsquigarrow$ general geometric test for maximum possible utility of tensor in laser method.

As a space of matrices

$$
CW_6 = \begin{pmatrix}
h & g_1 & g_2 & g_3 & f_1 & f_2 & f_3 & e \\
0 & h & 0 & 0 & 0 & 0 & 0 & f_1 \\
0 & 0 & h & 0 & 0 & 0 & 0 & f_2 \\
0 & 0 & 0 & h & 0 & 0 & 0 & f_3 \\
0 & 0 & 0 & 0 & h & 0 & 0 & g_1 \\
0 & 0 & 0 & 0 & 0 & h & 0 & g_2 \\
0 & 0 & 0 & 0 & 0 & 0 & h & g_3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & h
\end{pmatrix}.
$$

A better (?) tensor

$$T_{better,6} := \begin{pmatrix} h & g_1 & g_2 & g_3 & f_1 & f_2 & f_3 & e \\ 0 & h & 0 & 0 & 0 & g_3 & g_2 & f_1 \\ 0 & 0 & h & 0 & g_3 & 0 & g_1 & f_2 \\ 0 & 0 & 0 & h & g_2 & g_1 & 0 & f_3 \\ 0 & 0 & 0 & 0 & h & 0 & 0 & g_1 \\ 0 & 0 & 0 & 0 & 0 & h & 0 & g_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & h & g_3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & h \end{pmatrix}.$$

We'll see $\underline{\mathbf{R}}(T_{better,6}) = 8$, minimal border rank like CW_6.

Thm (GJLM 2025) $T_{better,6}$ is better than CW_6 for Strassen's laser method in the following sense:

- $T_{better,6} \trianglerighteq CW_6$ (i.e., at least as good)
- $\underset{\sim}{\mathbf{Q}}(T_{better,6}) > \underset{\sim}{\mathbf{Q}}(CW_6)$ (i.e., weaker barrier)

Next time: origin of $T_{better,6}$ and plan to construct similar tensors.

Thank you for your attention

For more on **tensors**, their geometry and applications, resp. **geometry and complexity**, resp. **asymptotic geometry**, resp. **quantum computation and information**: