Maximal border subrank tensors

Chia-Yu Chang

Texas A&M University

IPAM, February 6

Notations

A, B, C: n-dimensional vector spaces over $\mathbb{C}$

$\{a_i\}, \{b_i\}, \{c_i\}$: bases of A, B, C, respectively

$\{\alpha_i\}, \{\beta_i\}, \{\gamma_i\}$: dual bases

$\{e_i\}$: the standard basis of $\mathbb{C}^s$, $s \in \mathbb{N}$

$\langle s \rangle := \sum_{i=1}^{s} e_i \otimes e_i \otimes e_i \in \mathbb{C}^s \otimes \mathbb{C}^s \otimes \mathbb{C}^s$: the **unit tensor** of size s

A tensor $T \in A \otimes B \otimes C$ can be viewed as a linear map $T_A : A^* \to B \otimes C$.
Similarly, we have T_B and T_C.
We say a tensor $T \in A \otimes B \otimes C$ is **concise** if T_A, T_B, and T_C are injective.

In particular, the unit tensor $\langle s \rangle$ is concise in $\mathbb{C}^s \otimes \mathbb{C}^s \otimes \mathbb{C}^s$.

Rank, Border rank, Subrank, and Border subrank

Definition

Let $T \in A \otimes B \otimes C$.
The **rank** of T, $\mathbf{R}(T)$, is the minimal positive integer r such that

$$T \in (\mathrm{Hom}(\mathbb{C}^r, A) \times \mathrm{Hom}(\mathbb{C}^r, B) \times \mathrm{Hom}(\mathbb{C}^r, C)) \cdot \langle r \rangle.$$

The **border rank** of T, $\underline{\mathbf{R}}(T)$, is the minimal positive integer r such that

$$T \in \overline{(\mathrm{Hom}(\mathbb{C}^r, A) \times \mathrm{Hom}(\mathbb{C}^r, B) \times \mathrm{Hom}(\mathbb{C}^r, C)) \cdot \langle r \rangle}.$$

The **subrank** of T, $\mathbf{Q}(T)$, is the maximal positive integer s such that

$$\langle s \rangle \in (\mathrm{Hom}(A, \mathbb{C}^s) \times \mathrm{Hom}(B, \mathbb{C}^s) \times \mathrm{Hom}(C, \mathbb{C}^s)) \cdot T.$$

The **border subrank** of T, $\underline{\mathbf{Q}}(T)$, is the maximal positive integer s such that

$$\langle s \rangle \in \overline{(\mathrm{Hom}(A, \mathbb{C}^s) \times \mathrm{Hom}(B, \mathbb{C}^s) \times \mathrm{Hom}(C, \mathbb{C}^s)) \cdot T}.$$

Properties of (Border) Rank and (Border) Subrank

$\mathbf{R}(T) := \min\{r : T \in (\mathrm{Hom}(\mathbb{C}^r, A) \times \mathrm{Hom}(\mathbb{C}^r, B) \times \mathrm{Hom}(\mathbb{C}^r, C)) \cdot \langle r \rangle\}$

$\underline{\mathbf{R}}(T) := \min\{r : T \in \overline{(\mathrm{Hom}(\mathbb{C}^r, A) \times \mathrm{Hom}(\mathbb{C}^r, B) \times \mathrm{Hom}(\mathbb{C}^r, C)) \cdot \langle r \rangle}\}$

$\mathbf{Q}(T) := \max\{s : \langle s \rangle \in (\mathrm{Hom}(A, \mathbb{C}^s) \times \mathrm{Hom}(B, \mathbb{C}^s) \times \mathrm{Hom}(C, \mathbb{C}^s)) \cdot T\}$

$\underline{\mathbf{Q}}(T) := \max\{s : \langle s \rangle \in \overline{(\mathrm{Hom}(A, \mathbb{C}^s) \times \mathrm{Hom}(B, \mathbb{C}^s) \times \mathrm{Hom}(C, \mathbb{C}^s)) \cdot T}\}$

For any $T \in A \otimes B \otimes C$, we have

- $\mathbf{Q}(T) \leq \underline{\mathbf{Q}}(T) \leq n$, where $n = \dim(A) = \dim(B) = \dim(C)$

  T is of maximal (border) subrank if "$= n$"
- $\mathbf{Q}(T) \leq \underline{\mathbf{Q}}(T) \leq \underline{\mathbf{R}}(T) \leq \mathbf{R}(T)$
- $\mathbf{Q}(T) \leq \underline{\mathbf{Q}}(T) \leq n \leq \underline{\mathbf{R}}(T) \leq \mathbf{R}(T)$ if T is concise

Motivation from Complexity Theory

- The *exponent of matrix multiplication* is defined as

$$\omega := \inf\{h \in \mathbb{R} : \mathbf{R}(M_{\langle n,n,n\rangle}) = O(n^h)\},$$

where $M_{\langle n,n,n\rangle}$ is the $n \times n \times n$ matrix multiplication tensor.

- [Str69]: $2 \leq \omega \leq \log_2 7 < 2.81 < 3$
- A well-known method to find upper bounds on ω is the *laser method* [Str87]: study an intermediate tensor T which is
  1. of small border rank (low cost)
  2. close to being a matrix multiplication tensor (high value)
- The intermediate tensors of large (asymptotic) subrank are good to get bounds for ω.

Motivation

For a generic tensor T, $\mathbf{R}(T) = \underline{\mathbf{R}}(T) =$ maximum border rank $\sim n^2/3$.
How about $\mathbf{Q}(T)$ and $\underline{\mathbf{Q}}(T)$?
Unknown!

Theorem (Derksen, Makam, Zuiddam, 2022)

The generic subrank of tensors in $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ has bounds

$$3(\lfloor\sqrt{n/3 + 1/4} - 1/2\rfloor) \leq \mathbf{Q}(n) \leq \lfloor\sqrt{3n-2}\rfloor.$$

In particular, the generic subrank is not maximal.

Proposition

The border subrank of generic tensors in $\mathbb{C}^n \otimes \mathbb{C}^n \otimes \mathbb{C}^n$ is at most $n-1$ for $n \geq 3$.

Main result today: A lower bound of the dimension of the set of maximal border subrank tensors

Maximal subrank tensors

View $\langle n \rangle = \sum_{i=1}^{n} a_i \otimes b_i \otimes c_i \in A \otimes B \otimes C$, since A, B, C: n-dimensional.

Note $\mathbf{Q}(\langle n \rangle) = \underline{\mathbf{Q}}(\langle n \rangle) = n$.

Proposition

The orbit of the unit tensor $(\mathrm{GL}(A) \times \mathrm{GL}(B) \times \mathrm{GL}(C)) \cdot \langle n \rangle$ *consists of all maximal subrank tensors.*

Proof.

If $Q(T) = n$, then there exist $X \in \mathrm{End}(A)$, $Y \in \mathrm{End}(B)$, and $Z \in \mathrm{End}(C)$ such that

$$\langle n \rangle = (X \otimes Y \otimes Z) \cdot T \in \mathrm{im}(X) \otimes \mathrm{im}(Y) \otimes \mathrm{im}(Z).$$

Since $\langle n \rangle$ is concise, we get that X, Y, Z are invertible. □

Maximal border subrank tensors

$\underline{\mathbf{Q}}(T) = n$ if and only if

$$\langle n \rangle \in \overline{(\mathrm{End}(A) \times \mathrm{End}(B) \times \mathrm{End}(C)) \cdot T} = \overline{(\mathrm{GL}(A) \times \mathrm{GL}(B) \times \mathrm{GL}(C)) \cdot T}$$

Define $\underline{\mathbf{Q}}_{\max} := \overline{\{T \in A \otimes B \otimes C : \underline{\mathbf{Q}}(T) = n\}}$.

Then $\overline{(\mathrm{GL}(A) \times \mathrm{GL}(B) \times \mathrm{GL}(C)) \cdot \langle n \rangle} \subset \underline{\mathbf{Q}}_{\max}$.

Write $G = \mathrm{GL}(A) \times \mathrm{GL}(B) \times \mathrm{GL}(C)$.

Hence, $\dim(\underline{\mathbf{Q}}_{\max}) \geq \dim(G) - \dim(G_{\langle n \rangle}) = 3n^2 - 2n$,

where $G_{\langle n \rangle} := \{g \in G : g \cdot \langle n \rangle = \langle n \rangle\}$ is the symmetry group of $\langle n \rangle$.

Main Theorem

$$\dim(\underline{\mathbf{Q}}_{\max}) \geq \frac{2n^3 + 3n^2 - 2n}{3} \sim \frac{2}{3} n^3.$$

The nullcone by the symmetry group of the unit tensor

Define the nullcone $\mathcal{N}_{G_{\langle n \rangle}} := \{w \in A \otimes B \otimes C : 0 \in \overline{G_{\langle n \rangle} \cdot w}\}$, and let

$$\mathrm{Cone}(\langle n \rangle, \mathcal{N}_{G_{\langle n \rangle}}) := \{v + w : v \in \mathbb{C} \cdot \langle n \rangle \text{ and } w \in \mathcal{N}_{G_{\langle n \rangle}}\} \subset \underline{\mathbf{Q}}_{\max}.$$

Then we have $\overline{G \cdot \mathrm{Cone}(\langle n \rangle, \mathcal{N}_{G_{\langle n \rangle}})} \subset \underline{\mathbf{Q}}_{\max}$.

Proposition

The symmetry group of the unit tensor is $G_{\langle n \rangle} = \mathfrak{S}_n \ltimes \mathbf{T}$, *where*

$$\mathbf{T} := \{(\lambda, \mu, \nu) \in G : \lambda, \mu, \nu \text{: diagonal, } \lambda\mu\nu = Id_n\}$$

is a maximal torus and $\mathfrak{S}_n$ *is the symmetric group on n elements.*

Let $\mathcal{N}_{\mathbf{T}} := \{w \in A \otimes B \otimes C : 0 \in \overline{\mathbf{T} \cdot w}\} \subset \mathcal{N}_{G_{\langle n \rangle}}$.

By Hilbert-Mumford criterion, $\mathcal{N}_{G_{\langle n \rangle}} = G_{\langle n \rangle} \cdot \mathcal{N}_{\mathbf{T}}$.

Since $\mathbf{T}$ is normal in $G_{\langle n \rangle}$, we have that $\mathcal{N}_{G_{\langle n \rangle}} = \mathcal{N}_{\mathbf{T}}$.

The nullcone defined by the torus

Let $x_{ijk} = \alpha_i \otimes \beta_j \otimes \gamma_k$. The coordinate ring of $A \otimes B \otimes C$ is $\mathbb{C}[x_{ijk}]$.

$$\begin{aligned}
\mathcal{N}_{G_{\langle n\rangle}} &= \mathcal{N}_{\mathbf{T}} \\
&= \mathrm{Zeros}(\{f \in \mathbb{C}[x_{ijk}] : f\text{: homogeneous}, \deg(f) > 0, g \cdot f = f \quad \forall g \in \mathbf{T}\}) \\
&= \mathrm{Zeros}(\{f \in \mathbb{C}[x_{ijk}] : f\text{: monomials}, \deg(f) > 0, g \cdot f = f \quad \forall g \in \mathbf{T}\})
\end{aligned}$$

since $\mathbf{T}$ is a torus and monomials span the weight vectors.

Thus $\mathcal{N}_{G_{\langle n\rangle}}$ is a union of linear spaces.

In particular,

$$\begin{aligned}
\mathcal{N}_{G_{\langle n\rangle}} &\subset \mathrm{Zeros}(\{x_{iii}, x_{iij}x_{jji}, x_{iji}x_{jij}, x_{ijj}x_{jii}, x_{ijk}x_{jki}x_{kij} : \text{distinct } 1 \leq i, j, k \leq n\rangle\}) \\
&= \bigcup \text{ of linear spaces with } \dim = n^3 - \left(n + 3\binom{n}{2} + 2\binom{n}{3}\right)
\end{aligned}$$

$W := \langle a_i \otimes b_j \otimes c_k : \text{ at least one of } j, k \text{ less than } i\rangle$ is one of the linear spaces.

Dimension of the nullcone

Claim: $W \subset \mathcal{N}_{G_{\langle n \rangle}} := \{w \in A \otimes B \otimes C : 0 \in \overline{G_{\langle n \rangle} \cdot w}\}$
Proof of claim: Let $c(t) \in G_{\langle n \rangle}$ be defined as

$$c(t) := \left(\begin{pmatrix} t^{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & t^{\lambda_n} \end{pmatrix}, \begin{pmatrix} t^{\mu_1} & & 0 \\ & \ddots & \\ 0 & & t^{\mu_n} \end{pmatrix}, \begin{pmatrix} t^{\nu_1} & & 0 \\ & \ddots & \\ 0 & & t^{\nu_n} \end{pmatrix} \right)$$

where $\lambda_k = 2^n - 2^{n-k+1}$ and $\mu_k = \nu_k = 2^{n-k} - 2^{n-1}$ for $k = 1, \ldots, n$.
Then

$$\begin{aligned} c(t) \cdot (a_i \otimes b_j \otimes c_k) &= t^{\lambda_i + \mu_j + \nu_k}(a_i \otimes b_j \otimes c_k) \\ &= t^{2^{n-j} + 2^{n-k} - 2^{n-i+1}}(a_i \otimes b_j \otimes c_k) \end{aligned}$$

which tends to zero as $t \to 0$ when $a_i \otimes b_j \otimes c_k \in W$.

Dimension of the cone

Thus we can focus on W rather than on the whole nullcone $\mathcal{N}_{G_{\langle n \rangle}}$.
Consider the cone over W with vertex $\langle n \rangle$

$$\mathrm{Cone}(\langle n \rangle, W) := \{v + w : v \in \mathbb{C} \cdot \langle n \rangle \text{ and } w \in W\} \subset \mathrm{Cone}(\langle n \rangle, \mathcal{N}_{G_{\langle n \rangle}}) \subset \underline{\mathbf{Q}}_{\max},$$

which has dimension

$$\dim(W) + 1 = \frac{4n^3 - 3n^2 - n}{6} + 1.$$

The orbit closure of the cone $\overline{G \cdot \mathrm{Cone}(\langle n \rangle, W)}$ is also a subset of $\underline{\mathbf{Q}}_{\max}$.

Dimension of the orbit closure of the cone

Let $v = \langle n \rangle + w \in \mathrm{Cone}(\langle n \rangle, W)$ be a general point, where $w \in W$, and let $\mathrm{Tran}_G(v, \mathrm{Cone}(\langle n \rangle, W)) := \{g \in G : g \cdot v \in \mathrm{Cone}(\langle n \rangle, W)\}$.
The orbit closure of the cone $\overline{G \cdot \mathrm{Cone}(\langle n \rangle, W)}$ has dimension

$$\dim(G) + \dim(\mathrm{Cone}(\langle n \rangle, W)) - \dim(\mathrm{Tran}_G(v, \mathrm{Cone}(\langle n \rangle, W))).$$

We can compute $\dim(\mathrm{Tran}_G(v, \mathrm{Cone}(\langle n \rangle, W)))$ by considering its tangent space at the identity element of G

$$\{(x, y, z) \in \mathfrak{g} : (x, y, z).(\langle n \rangle + w) \in \mathrm{Cone}(\langle n \rangle, W)\},$$

where $\mathfrak{g} = \mathrm{End}(A) \oplus \mathrm{End}(B) \oplus \mathrm{End}(C)$.

Lower bound

The dimension of the tangent space of $\mathrm{Tran}_G(v, \mathrm{Cone}(\langle n \rangle, W))$ is

$$\frac{3n^2 + n + 2}{2}.$$

Thus the dimension of $\overline{G \cdot \mathrm{Cone}(\langle n \rangle, W)}$ is

$$\begin{aligned} &\dim(G) + \dim(\mathrm{Cone}(\langle n \rangle, W)) - \dim(\mathrm{Tran}_G(v, \mathrm{Cone}(\langle n \rangle, W))) \\ &= 3n^2 + \frac{4n^3 - 3n^2 - n}{6} + 1 - \frac{3n^2 + n + 2}{2} \\ &= \frac{2n^3 + 3n^2 - 2n}{3}. \end{aligned}$$

This gives a lower bound of the dimension of $\underline{\mathbf{Q}}_{\max}$.

Future Questions

(1) The nullcone $\mathcal{N}_{G_{\langle n \rangle}}$ contains the union of W and its permutations. We have examples showing that the nullcone is larger, but not necessarily larger dimensional. What are all components of the nullcone?

(2) A tensor $T \in A \otimes B \otimes C$ is of maximal border subrank if the unit tensor $\langle n \rangle$ lies in the orbit closure $\overline{G \cdot T}$. Can we apply a one parameter subgroup of G on the tensor T to approach the unit tensor?

(3) Tensors in the orbit closure of the cone $\mathrm{Cone}(\langle n \rangle, \mathcal{N}_{G_{\langle n \rangle}})$ are of maximal border subrank. Do all maximal border subrank tensors lie in the orbit closure of the cone? Do we have this if (2) is true?

Future Questions

(4) [Baiggi, C., Draisma, Rupniewski]: An upper bound on dimension of $\underline{\mathbf{Q}}_{\mathsf{max}}$ is

$$n^3 - \lfloor n/3 \rfloor^3 + 6n^2.$$

Can we reduce the gap between the upper and the lower bound?

(5) For a generic tensor T, there are nontrivial upper and lower bounds on its subrank [DMZ22]. Can we find nontrivial upper or lower bounds on border subrank of a generic tensor?

(6) We have methods for finding upper bounds of subrank and border subrank that rely on other notion of rank, for example, geometric rank, slice rank, and G-stable rank. Can we develop a strategy to find the subrank and border subrank of a tensor or upper bounds on them that does not rely on other notions of rank?

Thank you!

References I

Harm Derksen, Visu Makam, and Jeroen Zuiddam.

Subrank and optimal reduction of scalar multiplications to generic tensors.

In *37th Computational Complexity Conference*, volume 234 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 9, 23. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022.

Volker Strassen.

Gaussian elimination is not optimal.

Numer. Math., 13:354–356, 1969.

Volker Strassen.

Relative bilinear complexity and matrix multiplication.

J. Reine Angew. Math., 375/376:406–443, 1987.