On rank and subrank of the matrix multiplication tensors

Jeroen Zuiddam

University of Amsterdam
Matrix multiplication tensors

- characterize computational complexity of multiplying matrices
Matrix multiplication tensors

- Characterize computational complexity of multiplying matrices

- For this we need to understand tensor rank and subrank

helps
Matrix multiplication tensors

- characterize computational complexity of multiplying matrices

- for this we need to understand tensor rank and subrank

- these tensors are highly structured

↑ closed under powering
Definition

\[ \mathbf{M}_n = \sum e_{ij} \otimes e_{jk} \otimes e_{ki} \in \mathbb{F}^{n^2 \times n^2 \times n^2} \]
Definition

\[ \text{MM}_n := \sum e_{ij} \otimes e_{jk} \otimes e_{ki} \in \mathbb{F}^{n^2 \times n^2 \times n^2} \]

Complexity

\[ \begin{array}{ccc}
\begin{array}{c}
{n} \\
A \cdot B
\end{array} & \leftrightarrow & R(\text{MM}_n) \\
\text{number of scalar +/- needed?} & \uparrow & \text{tensor rank}
\end{array} \]
Definition

\[ \text{Definition} \]

\[ \text{MM}_n := \sum_{eij} \otimes e_{jk} \otimes e_{ki} \in \mathbb{F}^{n^2 \times n^2 \times n^2} \]

Complexity

\[ \text{Complexity} \]

\[ \text{number of scalar +/- needed?} \]

\[ \text{R(\text{MM}_n)} \]

↓

tensor rank

Structured

\[ \text{Structured} \]

\[ \text{MM}_a \otimes \text{MM}_b = \text{MM}_{ab} \]
**Definition**

\[ MM_n = \sum e_{ij} \otimes e_{jk} \otimes e_{ki} \in \mathbb{R}^{n^2 \times n^2 \times n^2} \]

**Complexity**

Number of scalar +/- needed?

\[ \begin{bmatrix} A \\
B \end{bmatrix} \quad \mapsto \quad R(MM_n) \]

Tensor rank

**Structured**

\[ MM_a \otimes MM_b = MM_{ab} \]

**Exponent**

\[
\lim_{m \to \infty} R \left( (MM_2)^{\otimes m} \right)^{1/m} = 2^w
\]
Tensor rank $R$
Create tensor from small diagonal tensor

Subrank $Q$
Create large diagonal tensor from tensor
Tensor rank $R$

Create tensor from small diagonal tensor

\[ T = U \otimes V \otimes W \cdot \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i \]

Create large diagonal tensor from tensor

\[ (T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i) \]
Tensor rank \( R \)

Create tensor from small diagonal tensor

\[
T = U \otimes V \otimes W \cdot \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i
\]

\((T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i)\)

\[\uparrow \text{find} \]

Subrank \( Q \)

Create large diagonal tensor from tensor

\[
\sum_{i=1}^{s} e_i \otimes e_i \otimes e_i = U \otimes V \otimes W \cdot T
\]

\[\uparrow \text{find} \]
Tensor rank $R$

Create tensor from small diagonal tensor

\[ T = U \otimes V \otimes W \cdot \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i \]

\[ T = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i \]

Subrank $Q$

Create large diagonal tensor from tensor

\[ \sum_{i=1}^{s} e_i \otimes e_i \otimes e_i = U \otimes V \otimes W \cdot T \]

Applications:

- Matrix multiplication
- Circuit complexity [Raz]

Applications:

- Matrix multiplication (also!)
- Additive combinatorics
Subramle application: Matrix multiplication barriers
Subraman application: Matrix multiplication barriers

History of bounds on exponent w

- best upper bound: 3
- best lower bound: 2

Sixties | Eighties | Now
Subrank application: Matrix multiplication barriers

History of bounds on exponent $\omega$

- best upper bound
- best lower bound
- lower bound under extra assumptions, "barrier"

Sixties Eighties Now
Ordering

\[ S \leq T \iff S = U \oplus V \oplus W \cdot T \text{ for some } U, V, W \]
Ordering

\[ S \leq T \text{ iff } S = U \odot V \odot W \cdot T \text{ for some } U, V, W \]

Rank upper bound phrased in ordering

\[ MM_m \leq \sum_{i=1}^{r} e_i \odot e_i \odot e_i \quad \Rightarrow \quad w \leq \log_m r \]
Ordering

\[ S \leq T \text{ iff } S = U \otimes V \otimes W \cdot T \text{ for some } U, V, W \]

Rank upper bound phrased in ordering

\[
MM_m \leq \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i \quad \Rightarrow \quad \omega \leq \log_m r
\]

In practice, we go via specific "intermediate" tensor \( T \)

\[
MM_m \leq T^{\otimes n} \leq I_r
\]
Ordering

\[ S \leq T \iff S = U \otimes V \otimes W \cdot T \] for some \( U, V, W \)

Rank upper bound phrased in ordering

\[ \text{MM}_n \leq \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i \quad \Rightarrow \quad w \leq \log_m r \]

In practice, we go via specific "intermediate" tensor \( T \)

\[ \text{MM}_n \leq T \otimes n \leq I_r \]

Theorem (Strassen) \( \uparrow \) essentially maximal subrank \( m^{2-o(1)} \)
Ordering

\[ S \leq T \iff S = U \otimes V \otimes W \cdot T \text{ for some } U, V, W \]

Rank upper bound phrased in ordering

\[ \text{MM}_m \leq \sum_{i=1}^{r} e_i \otimes e_i \otimes e_i \Rightarrow \omega \leq \log_m r \]

In practice, we go via specific "intermediate" tensor \( T \)

\[ \text{MM}_m \leq T^\otimes n \leq I_r \]

Theorem (Strassen) \( \uparrow \) essentially maximal subrank \( m^{2-o(1)} \)

Barrier Theorem (CVZ) small subrank of \( T^\otimes n \) implies \( \text{barrier} \)
Two directions in the rest of the talk:

1. How to upper bound subrank? for barriers and other applications

2. How to circumvent these barriers and improve upper bound on w?
Methods to upper bound subrank:

- Tensor rank: decompose into sum of simple tensors: $u \otimes v \otimes w$
Methods to upper bound subrank:

- **Tensor rank**: Decompose into sum of simple tensors: $u \otimes v \otimes w$
  - Change notion of "simple"
- **Slice rank [Tao]**: Decompose into:
  $$\sum_{j} u \otimes v_{j} \otimes w_{j}, \sum_{j} u_{j} \otimes v_{j} \otimes w_{j}, \sum_{j} u_{j} \otimes v_{j} \otimes w_{j}$$
Methods to upper bound subrank:

- Tensor rank: decompose into sum of simple tensors: $u \otimes v \otimes w$
  - Change notion of "simple"

- Slice rank [Tao]: decompose into:
  \[ \sum_{j} u \otimes v_j \otimes w_j, \sum_{j} u_j \otimes v \otimes w_j, \sum_{j} u_j \otimes v_j \otimes w \]

- Analytic rank [Gowers & Wolf] (over prime char.) counting, bias
Methods to upper bound subrank:

- Tensor rank: decompose into sum of simple tensors: \( u \otimes v \otimes w \)
  
  Change notion of "simple"

- Slice rank [Tao]: decompose into:
  \[
  \sum_j u \otimes v_j \otimes w_j, \ \sum_j u_j \otimes v \otimes w_j, \ \sum_j u_j \otimes v_j \otimes w
  \]

- Analytic rank [Gowers & Wolf] (over prime char.) counting, bias
  
  Extend to char. 0

- Geometric Rank [Kopparty-Moshkovitz-Z] dimension of variety
Methods to upper bound subrank:

- **Tensor rank**: decompose into sum of simple tensors: $u \otimes v \otimes w$

  Change notion of "simple"

- **Slice rank** [Tao]: decompose into:
  
  $$\sum_j u \otimes v_j \otimes w_j, \sum_j u_j \otimes v \otimes w_j, \sum_j u_j \otimes v_j \otimes w$$

- **Analytic rank** [Gowers & Wolf] (over prime char.) counting, bias

  Extend to char. 0

- **Geometric Rank** [Kopparty-Moshkovitz-Z] dimension of variety

- **G-stable rank** [Derksen] one-parameter subgroups, invariant theory
Geometric rank $\text{GR}(T)$

$$\text{codim} \left\{ (x, y) \in F^n \times F^n : \forall z \ T(x, y, z) = 0 \right\}$$
Geometric rank \( \text{GR}(T) \)

\[
\text{codim} \left\{ (x,y) \in \mathbb{F}^n \times \mathbb{F}^n : \forall z \ T(x,y,z) = 0 \right\}
\]

**Theorem** [Kopparty-Moshkovitz-Z]

\[ Q(T) \leq \text{GR}(T) \leq \text{SR}(T) \]
Geometric rank $\text{GR}(T)$

$$\text{codim} \{(x, y) \in \mathbb{F}^n \times \mathbb{F}^n : \forall z \ T(x, y, z) = 0\}$$

\underline{Theorem} [Kopparty-Moshkovitz-Z]

- $Q(T) \leq \text{GR}(T) \leq \text{SR}(T)$
  \[\text{[Strassen 1988]}\]
- $\text{GR}(MM^m_m) = \lceil \frac{3}{4}m^2 \rceil \quad \Rightarrow \quad \text{Solves problem: } Q(MM^m_m) = \lceil \frac{3}{4}m^2 \rceil$. 
**Geometric rank** $GR(T)$

$$\text{codim } \{ (x, y) \in \mathbb{F}^n \times \mathbb{F}^n : \forall z \ T(x, y, z) = 0 \}$$

**Theorem [Kopparty-Moshkovitz-Z]**

- $Q(T) \leq GR(T) \leq SR(T)$

  $\leftarrow [\text{Strassen 1988}]$

- $GR(MM_m) = \left\lceil \frac{3}{4} m^2 \right\rceil$ $\Rightarrow$ Solves problem: $Q(MM_m) = \left\lceil \frac{3}{4} m^2 \right\rceil$.  

- invariant under permuting $x, y$ and $z$

- "extends" analytic rank to char. 0
Geometric rank \( \text{GR}(T) \)

\[
\text{codim \{ (x, y) \in \mathbb{F}^n : \forall z \ T(x, y, z) = 0 \}}
\]

**Theorem** [Kopparty-Moshkovitz-Z]

- \( Q(T) \leq \text{GR}(T) \leq \text{SR}(T) \)
  - \( \text{GR}(\text{MM}_m) = \left\lceil \frac{3}{4} m^2 \right\rceil \Rightarrow \text{Solves problem: } Q(\text{MM}_m) = \left\lceil \frac{3}{4} m^2 \right\rceil \).
- Invariant under permuting \( x, y \) and \( z \)
- "Extends" analytic rank to char. 0

**Theorem** [Moshkovitz-Cohen] \( \text{GR} \) equals \( \text{SR} \) up to constant!

**Theorem** [Derksen-Makam-Z] Huge gap between \( Q \) and \( \text{GR} \).
2. How to improve upper bounds on \( w \)? (circumvent barriers?)

Traditional

- **Rank bounds**

\[
MM_m \leq I_r \quad \Rightarrow \quad w \leq \log_m r
\]
2. How to improve upper bounds on $w$? (circumvent barriers?)

**Traditional**

- Rank bounds
  \[ MM_m \leq I_r \Rightarrow w \leq \log_m r \]

- Schönhage's tau theorem
  \[ \bigoplus_{i} MM_{m_i} \leq I_r \Rightarrow \sum_{i} m_i^w \leq r \]
• Schönhage's tau theorem

\[ \oplus \mu \mu m_i \leq \text{Ir} \quad \Rightarrow \quad \sum_i m_i^w \leq r \]
• Schönhage's tau theorem

\[ \bigoplus_{i} m_i \leq r \Rightarrow \sum_{i} m_i^{w} \leq r \]

Non-traditional

Strassen's theory of asymptotic spectra

Positivstellensatz

Tensor inequalities \[\Leftrightarrow\] Real geometry

\[ \chi \in [1, \infty) \]
- Schönhage's tau theorem
  \[ \bigoplus_i M_i \leq I_r \implies \Sigma_i m_i^w \leq r \]

- Non-traditional

- Strassen's theory of asymptotic spectra

- Strassen's generalized tau theorem (from non-trivial connectedness)

- Positivstellensatz
  Tensor inequalities \( \iff \) Real geometry
  \[ \mapsto \chi \in [1, \infty) \]
- Schönhage's tau theorem

\[ \oplus \mathbf{MM}_{m_i} \leq \mathbf{r} \quad \Rightarrow \quad \sum_i m_i^w \leq r \]

Non-traditional

Strassen's theory of asymptotic spectra

- Strassen's generalized tau theorem (from non-trivial connectedness)

\[ \oplus \mathbf{MM}_{m_i} \leq \oplus \mathbf{MM}_{m_j} \quad \Rightarrow \quad \sum_i m_i^w \leq \sum_j m_j^w \]

Positivstellensatz

Tensor inequalities \( \longrightarrow \) Real geometry

\( \Rightarrow \) \( \chi \in \Gamma, \infty \)

Upcoming paper with Wigderson: survey, exposition, extensions