How to find good starting tensors for matrix multiplication

Markus Bläser

Saarland University

Matrix multiplication

$$\begin{pmatrix} z_{1,1} & \dots & z_{1,n} \\ \vdots & \ddots & \vdots \\ z_{n,1} & \dots & z_{n,n} \end{pmatrix} = \begin{pmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,n} \end{pmatrix} \cdot \begin{pmatrix} y_{1,1} & \dots & y_{1,n} \\ \vdots & \ddots & \vdots \\ y_{n,1} & \dots & y_{n,n} \end{pmatrix}$$

$$z_{i,j} = \sum_{k=1}^{n} x_{i,k} y_{k,j}, \qquad 1 \leq i, j \leq n$$

- entries are variables
- allowed operations: addition, multiplication, scalar multiplication

Strassen's algorithm

$$\begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix}.$$

$$\begin{aligned}
p_1 &= (x_{11} + x_{22})(y_{11} + y_{22}), \\
p_2 &= (x_{11} + x_{22})y_{11}, \\
p_3 &= x_{11}(y_{12} - y_{22}), \\
p_4 &= x_{22}(-y_{11} + y_{12}), \\
p_5 &= (x_{11} + x_{12})y_{22}, \\
p_6 &= (-x_{11} + x_{21})(y_{11} + y_{12}), \\
p_7 &= (x_{12} - x_{22})(y_{21} + y_{22}).
\end{aligned}$$

$$\begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} = \begin{pmatrix} p_1 + p_4 - p_5 + p_7 & p_3 + p_5 \\ p_2 + p_4 & p_1 + p_3 - p_2 + p_6 \end{pmatrix}.$$

Strassen's algorithm (2)

- 7 mults, 18 adds

instead of

- 8 mults, 4 adds

Strassen's algorithm (2)

- 7 mults, 18 adds

instead of

- 8 mults, 4 adds

Observation: Strassen's algorithm works over any ring!

Strassen's algorithm (2)

- 7 mults, 18 adds

instead of

- 8 mults, 4 adds

Observation: Strassen's algorithm works over any ring!

$\longrightarrow$ Recurse:

$$\left(\begin{array}{c|c} \quad & \quad \\ \hline \quad & \quad \end{array}\right) \cdot \left(\begin{array}{c|c} \quad & \quad \\ \hline \quad & \quad \end{array}\right) = \left(\begin{array}{c|c} \quad & \quad \\ \hline \quad & \quad \end{array}\right).$$

$$C(n) \leq 7 \cdot C(n/2) + O(n^2), \qquad C(1) = 1$$

Theorem (Strassen)

We can multiply $n \times n$-matrices with $O(n^{\log_2 7}) = O(n^{2.81})$ arithmetic operations

Tensor rank

In general:

- bilinear forms $b_1(X,Y), \ldots b_n(X,Y)$
- in variables $X = \{x_1, \ldots, x_k\}$ and $Y = \{y_1, \ldots, y_m\}$.

Write

$$\sum_{j=1}^{n} b_i z_i = \sum_{h=1}^{k} \sum_{i=1}^{m} \sum_{j=1}^{n} t_{h,i,j} x_h y_i z_j.$$

$$t = (t_{h,i,j}) \in K^k \otimes K^m \otimes K^n$$

is the tensor corresponding to $b_1, \ldots, b_n$.

Tensor rank

Definition

$u \otimes v \otimes w \in U \otimes V \otimes W$ is called a triad "rank-one tensor".

Definition (Rank)

$R(t)$ is the smallest r such that there are rank-one tensors $t_1, \ldots, t_r$ with $t = t_1 + \cdots + t_r$.

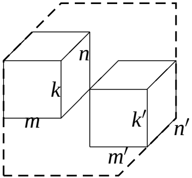
Lemma

Let $t \in U \otimes V \otimes W$ *and* $t' \in U' \otimes V' \otimes W'$.

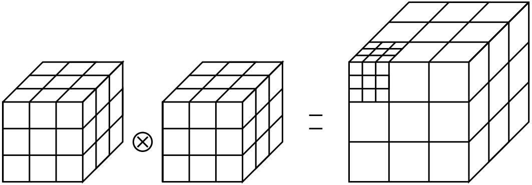
- $R(t \oplus t') \leq R(t) + R(t')$
- $R(t \otimes t') \leq R(t)R(t')$

Sums and products

Direct **sum** $\mathrm{t} \oplus \mathrm{t}' \in (U \oplus U') \otimes (V \oplus V') \otimes (W \oplus W')$:



Tensor **product** $\mathrm{t} \otimes \mathrm{t}' \in (U \otimes U') \otimes (V \otimes V') \otimes (W \otimes W')$:



Matrix multiplication tensor

Example: 2×2-matrix multiplication $\langle 2, 2, 2\rangle$:

$$
\begin{array}{cccccc}
x_{11} & x_{12} & x_{21} & x_{22} & \\
1 & & 1 & & y_{11} \\
 & 1 & & 1 & y_{21} \\
1 & & 1 & & y_{12} \\
 & 1 & & 1 & y_{22} \\
z_{11} & z_{21} & z_{12} & z_{22} &
\end{array}
$$

In general: $t_{(h,h'),(i,i'),(j,j')} = \delta_{h',i}\delta_{i',j}\delta_{j',h}$.

Lemma

- $R(\langle k, m, n\rangle) = R(\langle n, k, m\rangle) = \cdots = R(\langle n, m, k\rangle)$.
- $\langle k, m, n\rangle \otimes \langle k', m', n'\rangle \cong \langle kk', mm', nn'\rangle$.

Strassen's algorithm and tensors

Observation: Tensor product $\cong$ Recursion

Strassen's algorithm:

- $\langle 2, 2, 2\rangle^{\otimes s} = \langle 2^s, 2^s, 2^s\rangle$
- $R(\langle 2, 2, 2\rangle^{\otimes s}) \leq 7^s$

Definition (Exponent of matrix multiplication)

$\omega = \inf\{\tau \mid R(\langle n, n, n\rangle) = O(n^\tau)\}$

Strassen: $\omega \leq \log_2 7 \leq 2.81$

Lemma

If $R(\langle k, m, n\rangle) \leq r$, *then* $\omega \leq 3 \cdot \frac{\log r}{\log kmn}$.

What next?

Maybe we can multiply 2×2-matrices with 6 multiplications?

Theorem (Winograd)

$R(\langle 2, 2, 2\rangle) \geq 7$

Open question (not so open anymore)

Is there a small tensor $\langle n, n, n\rangle$, say, $n \leq 10$, which gives a better bound on the exponent than Strassen?

- Smirnov: $R(\langle 3, 3, 6\rangle) \leq 40 \longrightarrow \omega \leq 2.79$

Border rank (example)

Polynomial multiplication mod X^2:

$$(a_0 + a_1X)(b_0 + b_1X) = \underbrace{a_0b_0}_{f_0} + (\underbrace{a_1b_0 + a_0b_1}_{f_1})X + a_1b_1X^2$$

1	0
0	0

0	1
1	0

Observation

$R(t) = 3$

However, t can be approximated by tensors of rank 2.

$$t(\epsilon) = (1, \epsilon) \otimes (1, \epsilon) \otimes (0, \tfrac{1}{\epsilon}) + (1, 0) \otimes (1, 0) \otimes (1, -\tfrac{1}{\epsilon})$$

1	0
0	0

0	1
1	ϵ

Proof of observation — restrictions

Definition

Let $A : U \to U'$, $B : V \to V'$, $C : W \to W'$ be homomorphism.

- $(A \otimes B \otimes C)(u \otimes v \otimes w) = A(u) \otimes B(v) \otimes C(w)$
- $(A \otimes B \otimes C)t = \sum_{i=1}^{r} A(u_i) \otimes B(v_i) \otimes C(w_i)$ for $t = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$.
- $t' \leq t$ if there are A, B, C such that $t' = (A \otimes B \otimes C)t$. ("restriction").

Lemma

- *If* $t' \leq t$, *then* $R(t') \leq R(t)$
- $R(t) \leq r$ *iff* $t \leq \langle r \rangle$.
  ($\langle r \rangle$ "diagonal" of size r.*)*

Proof of observation

1	0
0	0

0	1
1	0

- Let $t = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$.
- $\text{lin}\{w_1, \ldots, w_r\} = K^2$.
- Asume that $w_r = (1, *)$.
- Let C be the projection along $\text{lin}\{w_r\}$ onto $\text{lin}\{(0,1)\}$.
- $(I \otimes I \otimes C)t = \begin{array}{|c|c|}\hline * & 1 \\ \hline 1 & 0 \\ \hline\end{array}$, which has rank 2.

Border rank

Definition

Let $h \in \mathbb{N}$, $t \in K^{k\times m\times n}$.

1. $R_h(t) = \min\{r \mid \exists u_\rho \in K[\epsilon]^k, v_\rho \in K[\epsilon]^m, w_\rho \in K[\epsilon]^n : \sum_{\rho=1}^{r} u_\rho \otimes v_\rho \otimes w_\rho = \epsilon^h t + O(\epsilon^{h+1})\}$.
2. $\underline{R}(t) = \min_h R_h(t)$. $\underline{R}(t)$ is called the *border rank* of t.

Bini, Capovani, Lotti, Romani: $\underline{R}(\langle 2,2,3\rangle) \le 10$.

Lemma

If $\underline{R}(\langle k,m,n\rangle) \le r$, *then* $\omega \le 3 \cdot \frac{\log r}{\log kmn}$.

Corollary

$\omega \le 2.79$.

Schönhage's τ-theorem

Schönhage: $\underline{R}(\langle k, 1, n\rangle \oplus \langle 1, (k-1)(n-1), 1\rangle) \leq kn + 1$.

Theorem (Schönhage's τ-theorem)

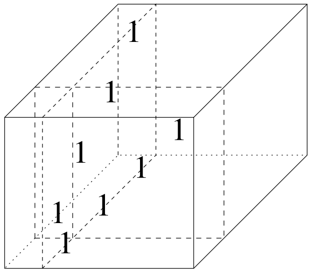
If $\underline{R}(\bigoplus_{i=1}^{p} \langle k_i, m_i, n_i\rangle) \leq r$ *with* $r > p$ *then* $\omega \leq 3\tau$ *where* τ *is defined by*

$$\sum_{i=1}^{p} (k_i \cdot m_i \cdot n_i)^{\tau} = r.$$

Corollary

$\omega \leq 2.55$.

Strassen's tensor



$$\mathrm{Str} = \sum_{i=1}^{q} (\underbrace{e_i \otimes e_0 \otimes e_i}_{\langle q,1,1\rangle} + \underbrace{e_0 \otimes e_i \otimes e_i}_{\langle 1,1,q\rangle})$$

$$= \frac{1}{\epsilon} \sum_{i=1}^{q} (e_0 + \epsilon e_i) \otimes (e_0 + \epsilon e_i) \otimes e_i - \frac{1}{\epsilon} e_0 \otimes e_0 \otimes \sum_{i=1}^{q} e_i + O(\epsilon)$$

Rank versus border rank

Theorem

$R(\mathrm{Str}) = 2q$.

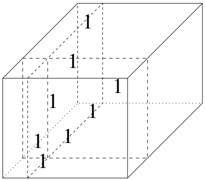
- Let $\mathrm{Str} = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$.
- W.l.o.g. assume that $u_r \notin \mathrm{lin}\{e_0, \ldots, e_{q-1}\}$.
- Let A be the projection along u_r onto $\mathrm{lin}\{e_0, \ldots, e_{q-1}\}$.
- Let B be the projection along e_q onto $\mathrm{lin}\{e_0, \ldots, e_{q-1}\}$.
- $R(A \otimes I \otimes B)\,\mathrm{Str}) \le R(\mathrm{Str}) - 1$.
- $(A \otimes I \otimes B)\,\mathrm{Str}$ is like Str with one inner tensor now being $\langle q-1, 1, 1\rangle$.
- Do this q times and kill q triads.
- We are left with a matrix of rank q.

Gap of almost 2 between rank and border rank.

Laser method

Think of Strassen's tensor having an outer and an inner structure: Cut Str into (combinatorial) cubiods!

- inner tensors: $\langle q, 1, 1\rangle$, $\langle 1, 1, q\rangle$
- outer structure: Put 1 in every cubiod that is nonzero.
  $\longrightarrow \langle 1, 2, 1\rangle$.



$(\mathrm{Str} \otimes \pi\, \mathrm{Str} \otimes \pi^2\, \mathrm{Str})^{\otimes s}$ has

- inner tensors $\langle x, y, z\rangle$ with $xyz = q^{3s}$,
- outer tensor $\langle 2^s, 2^s, 2^s\rangle$.

Degeneration

Definition

1. Let $t = \sum_{\rho=1}^{r} u_\rho \otimes v_\rho \otimes w_\rho \in K^{k \times m \times n}$, $A(\epsilon) \in K[\epsilon]^{k \times k'}$, $B(\epsilon) \in K[\epsilon]^{m \times m'}$, and $C(\epsilon) \in K[\epsilon]^{n \times n'}$. Define

$$(A(\epsilon) \otimes B(\epsilon) \otimes C(\epsilon))t = \sum_{\rho=1}^{r} A(\epsilon)u_\rho \otimes B(\epsilon)v_\rho \otimes C(\epsilon)w_\rho.$$

2. t is a *degeneration* of $t' \in K^{k \times m \times n}$ (" $t \trianglelefteq t'$ "), if there are $A(\epsilon)$, $B(\epsilon)$, $C(\epsilon)$, and q such that

$$\epsilon^q t = (A(\epsilon) \otimes B(\epsilon) \otimes C(\epsilon))t' + O(\epsilon^{q+1}).$$

Remark

$\underline{R}(t) \leq r \Leftrightarrow t \trianglelefteq \langle r \rangle$

Laser method (2)

A degeneration $(A(\epsilon), B(\epsilon), C(\epsilon))$ is called *monomial* if all entries are monomials.

Lemma (Strassen)

$\langle \lceil \frac{3}{4} n^2 \rceil \rangle \trianglelefteq \langle n, n, n \rangle$ *by a monomial degeneration.*

- inner tensors $\langle x, y, z \rangle$ with $xyz = q^{3s}$,
- outer tensor $\langle 2^s, 2^s, 2^s \rangle$.

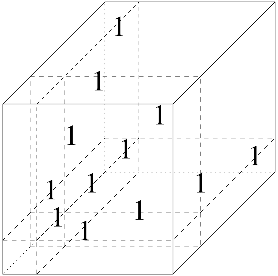
$\longrightarrow 2^{2s}$ independent matrix products with $\langle x, y, z \rangle$ with $xyz = q^{3s}$

Now apply the τ-theorem!

Corollary (Strassen)

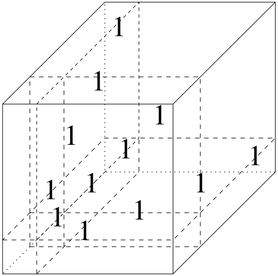
$\omega \leq 2.48$

Coppersmith–Winograd tensor



$$
\begin{aligned}
\epsilon^5\,\mathrm{CW} = \sum_{i=1}^{q} \epsilon \cdot (e_0 + \epsilon^2 e_i) \otimes (e_0 + \epsilon^2 e_i) \otimes (e_0 + \epsilon^2 e_i) \\
- (e_0 + \epsilon^3 \sum_{i=1}^{q} e_i) \otimes (e_0 + \epsilon^3 \sum_{i=1}^{q} e_i) \otimes (e_0 + \epsilon^3 \sum_{i=1}^{q} e_i) \\
+ (1 - q\epsilon) \cdot e_0 \otimes e_0 \otimes e_0 \\
+ O(\epsilon^6)
\end{aligned}
$$

Coppersmith–Winograd tensor



Remark (last time, a doable open question)

$R(CW) = 2q + 1$

Laser method (3)

CW has

- inner structure $\langle q, 1, 1\rangle$, $\langle 1, q, 1\rangle$, $\langle 1, 1, q\rangle$.
- outer structure

1	0
0	0

0	1
1	0

There is a general method how to degenerate large diagonals from arbitrary tensors.
$\longrightarrow$ apply to outer tensor

Corollary (Coppersmith & Winograd)

$\omega \leq 2.41$

Coppersmith & Winograd, Stothers, Vassilevska-Williams, LeGall: $\omega \leq 2.37\ldots$

What did we learn so far?

What did we learn so far?

- How to multiply matrices of astronomic sizes fast!

What did we learn so far?

- How to multiply matrices of astronomic sizes fast!
- If we want to multiply matrices of astronomic sizes even faster, we need tensors
  - with border rank close to $\max\{\dim U, \dim V, \dim W\}$
  - with a "rich" structure

What did we learn so far?

- How to multiply matrices of astronomic sizes fast!
- If we want to multiply matrices of astronomic sizes even faster, we need tensors
  - with border rank close to $\max\{\dim U, \dim V, \dim W\}$
  - with a "rich" structure
- or completely new methods.

Cheap approaches that do not work (not yet?)

	R	$\underline{\mathrm{R}}$
$\langle 2,2,2\rangle$	7	7
$\langle 2,2,3\rangle$	11	[9,10]
$\langle 2,2,4\rangle$	14	[12,14]*
$\langle 2,3,3\rangle$	[14,15]	[10,15]*
$\langle 3,3,3\rangle$	[19,23]	[15,20]

Main tools:

Rank: substition method (Pan), de Groote's twist of it

Border rank: vanishing equations (Strassen, Lickteig, Landsberg & Ottaviani)
in combination with substitution method (Landsberg & Michalek, B & Lysikov)

* Did not find any upper bounds

Characterization problem

Definition

- $S_n(q) = \{t \in K^n \otimes K^n \otimes K^n \mid R(t) \leq q\}$,
- $X_n(q) = \{t \in K^n \otimes K^n \otimes K^n \mid \underline{R}(t) \leq q\}$.

- These definitions are in "complexity-theoretic" terms.
- We need "algebraic" terms.

But: $\{t \mid t \trianglelefteq \langle q \rangle\}$ is not very useful

- We need "*easy to check* algebraic" criteria.

Remark: all tensors considered are tight.

$S_n(n)$

Theorem

$t \in S_n(n)$ *iff* $t \cong \langle n \rangle$

$S_n(n)$

Theorem

$t \in S_n(n)$ *iff* $t \cong \langle n \rangle$

The multiplication in any finite dimensional algebra A can be described by a set of bilinear forms. $\longrightarrow$ tensor t_A

Example:

- $A_\epsilon = K[X]/(X^n - \epsilon) \cong K^n$
- $A_\epsilon \to K[X]/(X^n)$
- $R(A) = 2n - 1$.

Theorem (Alder–Strassen)

$R(A) \geq 2 \dim A -$ *number of maximal twosided ideals.*

$X_n(n)$

Definition

Let $t \in U \otimes V \otimes W$. t is 1_U-generic (1_V, 1_W) if the U-slices (V, W) contain an invertible element.

Proposition (B & Lysikov)

Let t be 1_U- and 1_V-generic. Then there is an algebra A with structural tensor t_A such that $t_A \cong t$.

$X_n(n)$

Theorem (B & Lysikov)

Let A *and* B *be algebras with tensors* t_A *and* t_B*. Then* $t_A \in \overline{GL_n^{\times 3} \cdot t_B}$ *iff* $t_A \in \overline{GL_n \cdot t_B}$.

Theorem (B & Lysikov)

Let t *be* 1_U*- and* 1_V*-generic. Then* $t \in X_n(n)$ *iff there is an algebra* A *such that* $t_A \cong t$ *and* $t_A \in \overline{GL_n \cdot \langle n \rangle}$

Smoothable algebras

Definition

An algebra A of dimension n of the form $K[X_1, \dots, X_m]/I$ for some ideal I is called smoothable if I is a degeneration of some ideal whose zero set consists of n distinct points.

Theorem (B & Lysikov)

Let t be 1_U- and 1_V-generic. Then $t \in X_n(n)$ iff there is a smoothable algebra A such that $t_A \cong t$.

Examples

Cartwright et al.:

- All (commutative) algebras of dimension ≤ 7 are smoothable.
- All algebras generated by two elements are smoothable.
- All algebras with $\dim \operatorname{rad}(A)^2/\operatorname{rad}(A)^3 = 1$
- All algebras defined by a monomial ideal.

- Str^+ has minimal border rank. Its structural tensor is isomorphic to

$$k[X_1, \dots, X_q]/(X_i X_j \mid 1 \leq i, j \leq q)$$

- CW^+ has minimal border rank. Its structural tensor is isomorphic to

$$k[X_1, \dots, X_{q+1}]/(X_i X_j, X_i^2 - X_j^2, X_i^3 | i \neq j)$$

Comon's conjecture

- symmetric tensor = invariant under permutation of dimensions
- symmetric rank = use symmetric rank-one tensors

Conjecture (Comon)

For symmetric tensors, the rank equals the symmetric rank.

Proposition

The border rank Comon conjecture is true for 1-generic tensors of minimal border rank.

$X_n(n+1)$

Theorem

$t \in S_n(n+1) \setminus S_n(n)$ iff t is isomorphic to the multiplication tensors in the algebras

- $K[X]/(X^2) \times K^{n-2}$ *or*
- $T_2 \times K^{n-3}$.

where T_2 is the algebra of upper triangular 2×2-matrices.

Open question (Doable)

What about $X_n(n+1)$ (for 1-generic tensors)?

The asymptotic rank of CW

- We know that $\underline{R}(\mathrm{CW}_q) = q + 2$.
- For fast matrix multiplication, good upper bounds on $\underline{R}(\mathrm{CW}_q^{\otimes N})$ are sufficient.
- In particular, $\underline{R}(\mathrm{CW}_3^{\otimes N})^{1/N} \to 3$ implies $\omega = 2$.

Theorem (B. & Lysikov)

$\underline{R}(\mathrm{CW}_q^{\otimes N}) \geq (q+1)^N + 2^N$.