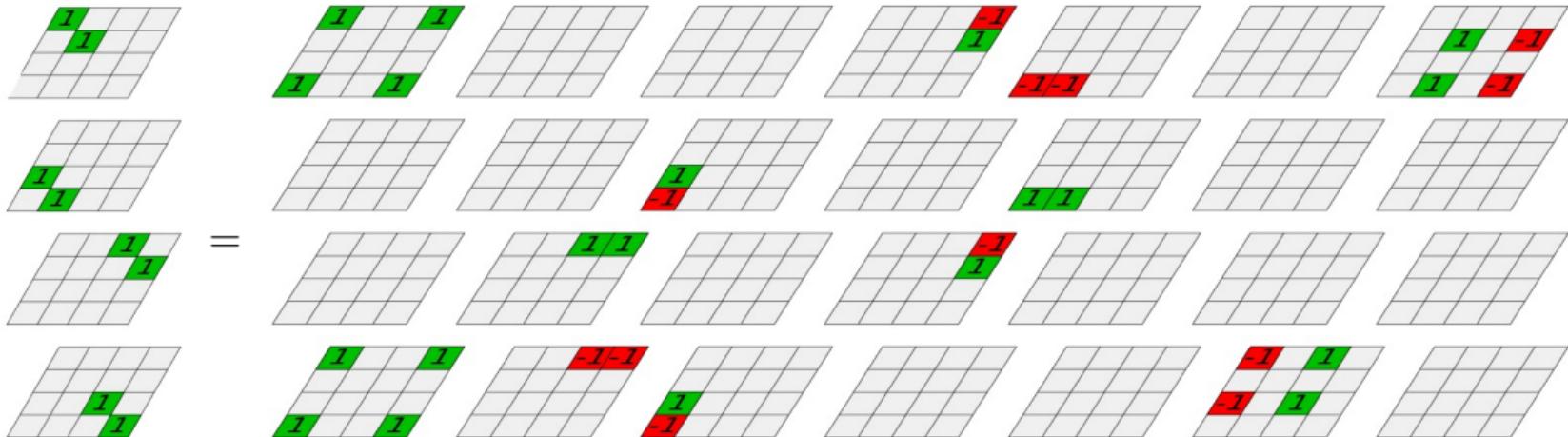


What is... Fast Matrix Multiplication

Dion Gijswijt



Matrix multiplication

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \quad c_{ik} = \sum_j a_{ij} b_{jk}$$

Matrix multiplication

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \quad c_{ik} = \sum_j a_{ij} b_{jk}$$

Multiplying $n \times n$ matrices takes $O(n^3)$ operations (n^3 multiplications and $n^3 - n^2$ additions).

Matrix multiplication

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \quad c_{ik} = \sum_j a_{ij} b_{jk}$$

Multiplying $n \times n$ matrices takes $O(n^3)$ operations (n^3 multiplications and $n^3 - n^2$ additions).

Can we do better?

Exponent of matrix multiplication

$$\omega = \inf\{\alpha : \text{We can multiply } n \times n \text{ matrices in time } O(n^\alpha)\}$$

We know that $2 \leq \omega \leq 3$.

Strassen's bilinear algorithm (1969)



$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Strassen's bilinear algorithm (1969)



$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$M_1 := (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$M_2 := (A_{21} + A_{22}) \cdot B_{11}$$

$$M_3 := (A_{11} + A_{12}) \cdot B_{22}$$

$$M_4 := A_{11} \cdot (B_{12} - B_{22})$$

$$M_5 := A_{22} \cdot (B_{21} - B_{11})$$

$$M_6 := (A_{21} - A_{11}) \cdot (B_{11} + B_{12})$$

$$M_7 := (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

$$C_{11} = M_1 - M_3 + M_5 + M_7$$

$$C_{12} = M_3 + M_4$$

$$C_{21} = M_2 + M_5$$

$$C_{22} = M_1 - M_2 + M_4 + M_6$$

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$M_1 := (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$M_2 := (A_{21} + A_{22}) \cdot B_{11}$$

$$M_3 := (A_{11} + A_{12}) \cdot B_{22} \qquad \qquad C_{11} = M_1 + M_4 - M_5 + M_7$$

$$M_4 := A_{11} \cdot (B_{12} - B_{22})$$

$$M_5 := A_{22} \cdot (B_{21} - B_{11})$$

$$M_6 := (A_{21} - A_{11}) \cdot (B_{11} + B_{12})$$

$$M_7 := (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

$$C_{12} = M_3 + M_5$$

$$C_{21} = M_2 + M_4$$

$$C_{22} = M_1 - M_2 + M_3 + M_6$$

Recursive application to $2^n \times 2^n$ matrices

Number of operations $f(n)$ satisfies $f(n+1) \leq 7f(n) + 18 \cdot 4^n$. Hence, $f(n) = O(7^n)$.

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \cdot \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$M_1 := (A_{11} + A_{22}) \cdot (B_{11} + B_{22})$$

$$M_2 := (A_{21} + A_{22}) \cdot B_{11}$$

$$M_3 := (A_{11} + A_{12}) \cdot B_{22}$$

$$C_{11} = M_1 + M_4 - M_5 + M_7$$

$$M_4 := A_{11} \cdot (B_{12} - B_{22})$$

$$C_{12} = M_3 + M_5$$

$$M_5 := A_{22} \cdot (B_{21} - B_{11})$$

$$C_{21} = M_2 + M_4$$

$$M_6 := (A_{21} - A_{11}) \cdot (B_{11} + B_{12})$$

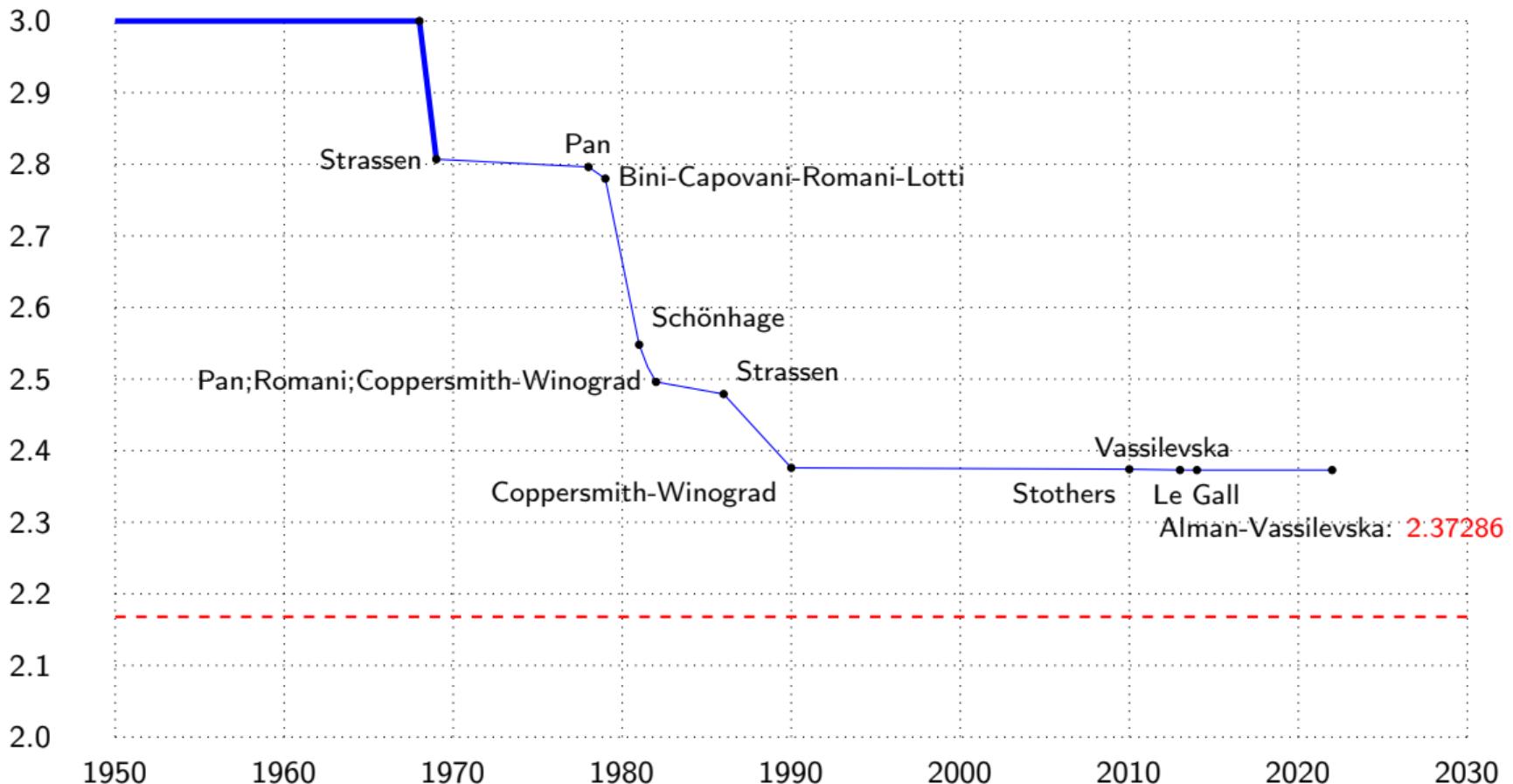
$$C_{22} = M_1 - M_2 + M_3 + M_6$$

$$M_7 := (A_{12} - A_{22}) \cdot (B_{21} + B_{22})$$

Recursive application to $2^n \times 2^n$ matrices

Number of operations $f(n)$ satisfies $f(n+1) \leq 7f(n) + 18 \cdot 4^n$. Hence, $f(n) = O(7^n)$.

⇒ We can multiply $N \times N$ matrices in time $O(N^{\log_2 7}) = O(N^{2.81})$.



Applications

- Algorithms in linear algebra. E.g. matrix inversion, determinant have complexity $O(n^\omega)$.

Applications

- Algorithms in linear algebra. E.g. matrix inversion, determinant have complexity $O(n^\omega)$.
- Linear programming.

Applications

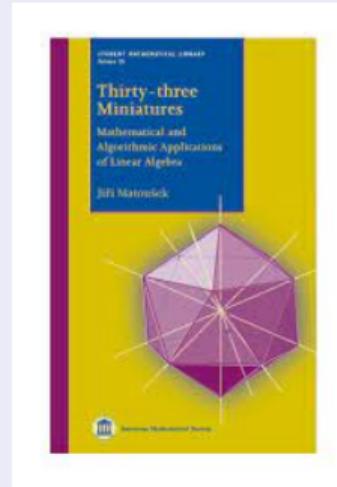
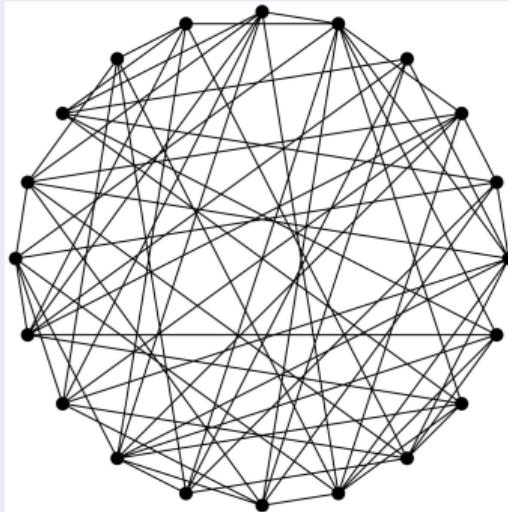
- Algorithms in linear algebra. E.g. matrix inversion, determinant have complexity $O(n^\omega)$.
- Linear programming.
- Algorithms on graphs. E.g. all-pair shortest path, graph matching.

Applications

- Algorithms in linear algebra. E.g. matrix inversion, determinant have complexity $O(n^\omega)$.
- Linear programming.
- Algorithms on graphs. E.g. all-pair shortest path, graph matching.

Puzzle: How to find a triangle in a graph?

[Matoušek, 33 miniatures]

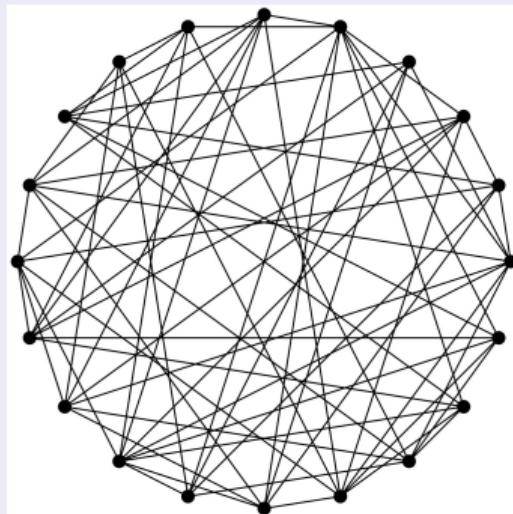


Applications

- Algorithms in linear algebra. E.g. matrix inversion, determinant have complexity $O(n^\omega)$.
- Linear programming.
- Algorithms on graphs. E.g. all-pair shortest path, graph matching.

Puzzle: How to find a triangle in a graph?

[Matoušek, 33 miniatures]



Let A be the $n \times n$ adjacency matrix.

Find i, j s.t. $A_{ij} \neq 0$ and $(A^2)_{ij} \neq 0$.

Takes only $O(n^\omega)$ time.

Tensor / Trilinear form

$$T = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} x_i y_j z_k$$

- The index sets I, J, K are finite.
- The c_{ijk} are coefficients in field \mathbb{F} .
- The x_i, y_j, z_k are variables.

Tensor / Trilinear form

$$T = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} x_i y_j z_k$$

- The index sets I, J, K are finite.
- The c_{ijk} are coefficients in field \mathbb{F} .
- The x_i, y_j, z_k are variables.

Tensor rank

We say that $T \neq 0$ has rank 1 if

$$T = \left(\sum_i \alpha_i x_i \right) \cdot \left(\sum_j \beta_j y_j \right) \cdot \left(\sum_k \gamma_k z_k \right).$$

That is: $c_{ijk} = \alpha_i \beta_j \gamma_k$ for all i, j, k .

In general: $R(T)$ is minimum r in a decomposition $T = T_1 + \cdots + T_r$ of rank 1 tensors.

Direct sum

Given

$$T = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} x_i y_j z_k, \quad T' = \sum_{i' \in I'} \sum_{j' \in J'} \sum_{k' \in K'} c_{i'j'k'} x_{i'} y_{j'} z_{k'}$$

Assume $I \cap I' = J \cap J' = K \cap K' = \emptyset$.

The **direct sum** is $T + T'$ viewed with variable sets $I \cup I', J \cup J', K \cup K'$.

Direct sum

Given

$$T = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} x_i y_j z_k, \quad T' = \sum_{i' \in I'} \sum_{j' \in J'} \sum_{k' \in K'} c_{i'j'k'} x_{i'} y_{j'} z_{k'}$$

Assume $I \cap I' = J \cap J' = K \cap K' = \emptyset$.

The **direct sum** is $T + T'$ viewed with variable sets $I \cup I', J \cup J', K \cup K'$.

Compare to matrices: $A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$

Direct sum

Given

$$T = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} x_i y_j z_k, \quad T' = \sum_{i' \in I'} \sum_{j' \in J'} \sum_{k' \in K'} c_{i'j'k'} x_{i'} y_{j'} z_{k'}$$

Assume $I \cap I' = J \cap J' = K \cap K' = \emptyset$.

The **direct sum** is $T + T'$ viewed with variable sets $I \cup I', J \cup J', K \cup K'$.

Compare to matrices: $A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$

- $R(T \oplus T') \leq R(T) + R(T')$ (sub additive)

Kronecker product

Given

$$T = \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} c_{ijk} x_i y_j z_k, \quad T' = \sum_{i' \in I'} \sum_{j' \in J'} \sum_{k' \in K'} c_{i'j'k'} x_{i'} y_{j'} z_{k'}$$

the Kronecker product is

$$T \otimes T' = \sum_{(i,i') \in I \times I'} \sum_{(j,j') \in J \times J'} \sum_{(k,k') \in K \times K'} c_{ijk} c_{i'j'k'} x_{ii'} y_{jj'} z_{kk'}$$

- $R(T \otimes T') \leq R(T)R(T')$ (sub multiplicative)

MM tensors

Define

$$\langle \ell, m, n \rangle = \sum_{i=1}^{\ell} \sum_{j=1}^m \sum_{k=1}^n x_{ij} y_{jk} z_{ki}.$$

This tensor encodes multiplication of $\ell \times m$ and $m \times n$ matrices.

MM tensors

Define

$$\langle \ell, m, n \rangle = \sum_{i=1}^{\ell} \sum_{j=1}^m \sum_{k=1}^n x_{ij} y_{jk} z_{ki}.$$

This tensor encodes multiplication of $\ell \times m$ and $m \times n$ matrices.

- Suppose $AB = C$.
- Substitute $x_{ij} = A_{ij}$ and $y_{jk} = B_{jk}$.
- Then C_{ik} is the coefficient of z_{ki} .

MM tensors

Define

$$\langle \ell, m, n \rangle = \sum_{i=1}^{\ell} \sum_{j=1}^m \sum_{k=1}^n x_{ij} y_{jk} z_{ki}.$$

This tensor encodes multiplication of $\ell \times m$ and $m \times n$ matrices.

- Suppose $AB = C$.
- Substitute $x_{ij} = A_{ij}$ and $y_{jk} = B_{jk}$.
- Then C_{ik} is the coefficient of z_{ki} .

Lemma

- $\langle \ell, m, n \rangle \otimes \langle \ell', m', n' \rangle \cong \langle \ell\ell', mm', nn' \rangle$
- $R(\langle \ell, m, n \rangle) = R(\langle n, \ell, m \rangle)$

So $R(\langle \ell mn, \ell mn, \ell mn \rangle) \leq R(\langle \ell, m, n \rangle)^3$

$$\begin{aligned}
\langle 2, 2, 2 \rangle &= x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + \\
&\quad x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22} + x_{22}y_{21}z_{12} + x_{22}y_{22}z_{22} \\
&= \\
&\quad (x_{11} + x_{22}) \cdot (y_{11} + y_{22}) \cdot (z_{11} + z_{22}) \\
&\quad \quad x_{11} \cdot (y_{12} - y_{22}) \cdot (z_{21} + z_{22}) \\
&\quad \quad x_{22} \cdot (y_{21} - y_{11}) \cdot (z_{12} + z_{11}) \\
&\quad (x_{21} + x_{22}) \cdot y_{11} \cdot (z_{12} - z_{22}) \\
&\quad (x_{12} + x_{11}) \cdot y_{22} \cdot (z_{21} - z_{11}) \\
&\quad (x_{12} - x_{22}) \cdot (y_{21} + y_{22}) \cdot z_{11} \\
&\quad (x_{21} - x_{11}) \cdot (y_{12} + y_{11}) \cdot z_{22}
\end{aligned}$$

So $R(\langle 2, 2, 2 \rangle) \leq 7$.

Theorem

- Suppose that $R(\langle n, n, n \rangle) \leq r$. Then $n^\omega \leq r$.
- Suppose that $R(\langle \ell, m, n \rangle) \leq r$. Then $(\ell mn)^\omega \leq r^3$.

Theorem

- Suppose that $R(\langle n, n, n \rangle) \leq r$. Then $n^\omega \leq r$.
- Suppose that $R(\langle \ell, m, n \rangle) \leq r$. Then $(\ell mn)^\omega \leq r^3$.

Examples

- (Strassen) $R(\langle 2, 2, 2 \rangle) \leq 7$. So $\omega \leq \log(7)/\log(2) < 2.81$

Theorem

- Suppose that $R(\langle n, n, n \rangle) \leq r$. Then $n^\omega \leq r$.
- Suppose that $R(\langle \ell, m, n \rangle) \leq r$. Then $(\ell mn)^\omega \leq r^3$.

Examples

- (Strassen) $R(\langle 2, 2, 2 \rangle) \leq 7$. So $\omega \leq \log(7)/\log(2) < 2.81$
- (Pan) $R(\langle 70, 70, 70 \rangle) \leq 143640$. So $\omega \leq \log(143640)/\log(70) < 2.796$.

Theorem

- Suppose that $R(\langle n, n, n \rangle) \leq r$. Then $n^\omega \leq r$.
- Suppose that $R(\langle \ell, m, n \rangle) \leq r$. Then $(\ell mn)^\omega \leq r^3$.

Examples

- (Strassen) $R(\langle 2, 2, 2 \rangle) \leq 7$. So $\omega \leq \log(7)/\log(2) < 2.81$
- (Pan) $R(\langle 70, 70, 70 \rangle) \leq 143640$. So $\omega \leq \log(143640)/\log(70) < 2.796$.

We know that $R(\langle 3, 3, 3 \rangle) \in \{19, 20, 21, 22, 23\}$.

Rank 19 would give $\omega < 2.681$.

Theorem

- Suppose that $R(\langle n, n, n \rangle) \leq r$. Then $n^\omega \leq r$.
- Suppose that $R(\langle \ell, m, n \rangle) \leq r$. Then $(\ell mn)^\omega \leq r^3$.

Examples

- (Strassen) $R(\langle 2, 2, 2 \rangle) \leq 7$. So $\omega \leq \log(7)/\log(2) < 2.81$
- (Pan) $R(\langle 70, 70, 70 \rangle) \leq 143640$. So $\omega \leq \log(143640)/\log(70) < 2.796$.

We know that $R(\langle 3, 3, 3 \rangle) \in \{19, 20, 21, 22, 23\}$.

Rank 19 would give $\omega < 2.681$.

Testing if $R(\langle 3, 3, 3 \rangle) \leq 19$ can in principle be done by solving a non-convex opt. problem in $19 \cdot 3 \cdot 9 = 513$ variables.

Finding rank decomposition

Suppose $T = \sum_{i,j,k=1}^n c_{ijk}x_iy_jz_k$. To test if

$$T = \sum_{\ell=1}^r \left(\sum_i \alpha_i^{(\ell)} x_i \right) \cdot \left(\sum_j \beta_j^{(\ell)} y_j \right) \cdot \left(\sum_k \gamma_k^{(\ell)} z_k \right)$$

for some numbers $\alpha_i^{(\ell)}, \beta_j^{(\ell)}, \gamma_k^{(\ell)}$, solve:

$$\min_{\alpha, \beta, \gamma} \left\| \sum_{ijk} [c_{ijk} - \sum_{\ell=1}^r \alpha_i^{(\ell)} \beta_j^{(\ell)} \gamma_k^{(\ell)}] \right\|.$$

For instance, using Alternate Least Squares.

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

This is $\langle 2, 2, 2 \rangle$ where we set $x_{22} = 0$.

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

Bini et al. found:

$$T = \frac{1}{\lambda} \left[\begin{array}{l} (x_{12} + \lambda x_{11}) \cdot (y_{12} + \lambda y_{22}) \cdot z_{21} \\ + (x_{21} + \lambda x_{11}) \cdot y_{11} \cdot (z_{11} + \lambda z_{12}) \\ - x_{12} \cdot y_{12} \cdot (z_{11} + z_{21} + \lambda z_{22}) \\ - x_{21} \cdot (y_{11} + y_{12} + \lambda y_{21}) \cdot z_{11} \\ + (x_{12} + x_{21}) \cdot (y_{12} + \lambda y_{21}) \cdot (z_{11} + \lambda z_{22}) \end{array} \right] + O(\lambda)$$

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

Bini et al. found:

$$T = \frac{1}{\lambda} \left[\begin{array}{l} (x_{12} + \lambda x_{11}) \cdot (y_{12} + \lambda y_{22}) \cdot z_{21} \\ + (x_{21} + \lambda x_{11}) \cdot y_{11} \cdot (z_{11} + \lambda z_{12}) \\ - x_{12} \cdot y_{12} \cdot (z_{11} + z_{21} + \lambda z_{22}) \\ - x_{21} \cdot (y_{11} + y_{12} + \lambda y_{21}) \cdot z_{11} \\ + (x_{12} + x_{21}) \cdot (y_{12} + \lambda y_{21}) \cdot (z_{11} + \lambda z_{22}) \end{array} \right] + O(\lambda)$$

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

Bini et al. found:

$$T = \frac{1}{\lambda} \begin{bmatrix} (x_{12} + \lambda x_{11}) \cdot (y_{12} + \lambda y_{22}) \cdot z_{21} \\ + (x_{21} + \lambda x_{11}) \cdot y_{11} \cdot (z_{11} + \lambda z_{12}) \\ - x_{12} \cdot y_{12} \cdot (z_{11} + z_{21} + \lambda z_{22}) \\ - x_{21} \cdot (y_{11} + y_{12} + \lambda y_{21}) \cdot z_{11} \\ + (x_{12} + x_{21}) \cdot (y_{12} + \lambda y_{21}) \cdot (z_{11} + \lambda z_{22}) \end{bmatrix} + O(\lambda)$$

Border rank

Border rank $\underline{R}(T)$ is at most 5.

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

$$T' = x_{23}y_{31}z_{12} + x_{23}y_{32}z_{22} + x_{22}y_{21}z_{12} + x_{22}y_{22}z_{22} + x_{13}y_{31}z_{11} + x_{13}y_{32}z_{21}$$

T' is obtained from T by:

for each $x_{ij}y_{jk}z_{ki}$ change $i \leftrightarrow 3 - i$ and $j \leftrightarrow 4 - j$.

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

$$T' = x_{23}y_{31}z_{12} + x_{23}y_{32}z_{22} + x_{22}y_{21}z_{12} + x_{22}y_{22}z_{22} + x_{13}y_{31}z_{11} + x_{13}y_{32}z_{21}$$

T' is obtained from T by:

for each $x_{ij}y_{jk}z_{ki}$ change $i \leftrightarrow 3 - i$ and $j \leftrightarrow 4 - j$.

- $\langle 2, 3, 2 \rangle = T + T'$
- $\underline{R}(T') = \underline{R}(T) \leq 5$.

Bini and border rank

$$T = x_{11}y_{11}z_{11} + x_{11}y_{12}z_{21} + x_{12}y_{21}z_{11} + x_{12}y_{22}z_{21} + x_{21}y_{11}z_{12} + x_{21}y_{12}z_{22}$$

$$T' = x_{23}y_{31}z_{12} + x_{23}y_{32}z_{22} + x_{22}y_{21}z_{12} + x_{22}y_{22}z_{22} + x_{13}y_{31}z_{11} + x_{13}y_{32}z_{21}$$

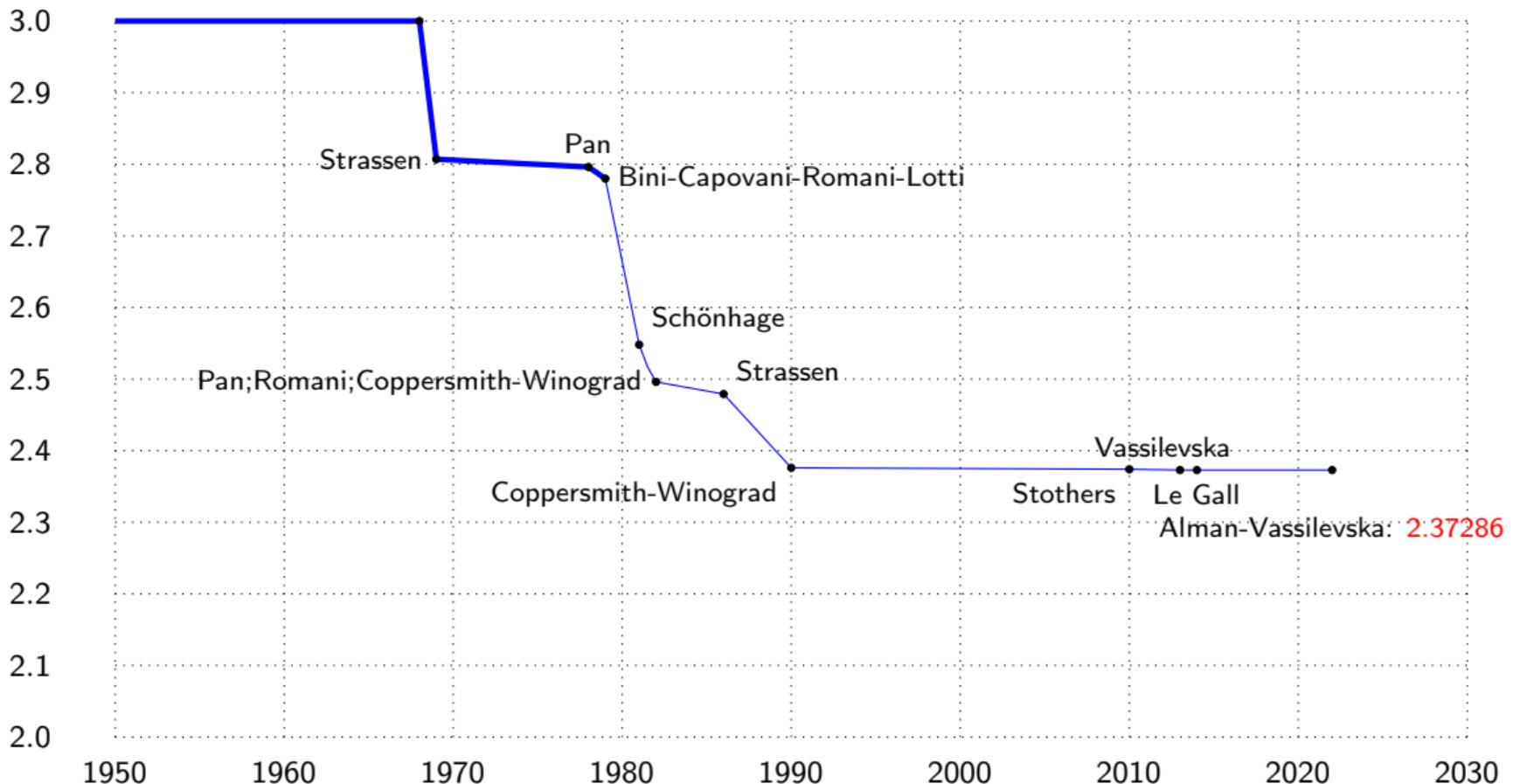
T' is obtained from T by:

for each $x_{ij}y_{jk}z_{ki}$ change $i \leftrightarrow 3 - i$ and $j \leftrightarrow 4 - j$.

- $\langle 2, 3, 2 \rangle = T + T'$
- $\underline{R}(T') = \underline{R}(T) \leq 5$.

Conclusion

Since $\underline{R}(\langle 2, 3, 2 \rangle) \leq 10$ we have $\omega \leq 3 \log(10) / \log(2 \cdot 3 \cdot 2) < 2.78$.



Schönhage's direct sum theorem



Observation 1

Border rank can be strictly subadditive !

For matrices: $\text{rank} \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = \text{rank } A + \text{rank } B$

Border rank is not additive

- $\underline{R}(\langle 3, 1, 3 \rangle) = 9, \quad \underline{R}(\langle 1, 4, 1 \rangle) = 4$
- $\underline{R}(\langle 3, 1, 3 \rangle \oplus \langle 1, 4, 1 \rangle) \leq 10$

Shitov [2019] found example with $R(S \oplus T) < R(S) + R(T)$, disproving Strassen's conjecture.

Schönhage's direct sum theorem



Observation 2

Direct sums of equal MM-tensors simulate larger MM-tensors.

Schönhage's direct sum theorem



Observation 2

Direct sums of equal MM-tensors simulate larger MM-tensors.

Suppose $T = 7 \odot \langle n, n, n \rangle$ (direct sum of 7 copies of $\langle n, n, n \rangle$).

Then T can ‘simulate’ $\langle 2, 2, 2 \rangle \otimes \langle n, n, n \rangle = \langle 2n, 2n, 2n \rangle$.

Schönhage's direct sum theorem



Observation 2

Direct sums of equal MM-tensors simulate larger MM-tensors.

Suppose $T = 7 \odot \langle n, n, n \rangle$ (direct sum of 7 copies of $\langle n, n, n \rangle$).

Then T can ‘simulate’ $\langle 2, 2, 2 \rangle \otimes \langle n, n, n \rangle = \langle 2n, 2n, 2n \rangle$.

In general, if $t > k^\omega$, then $t \odot \langle n, n, n \rangle$ can ‘simulate’ $\langle kn, kn, kn \rangle$.

Schönhage's direct sum theorem



Theorem

Let T be a tensor such that

- $\underline{R}(T) \leq r$
- T ‘contains’ direct sum $\langle \ell_1, m_1, n_1 \rangle \oplus \cdots \oplus \langle \ell_t, m_t, n_t \rangle$

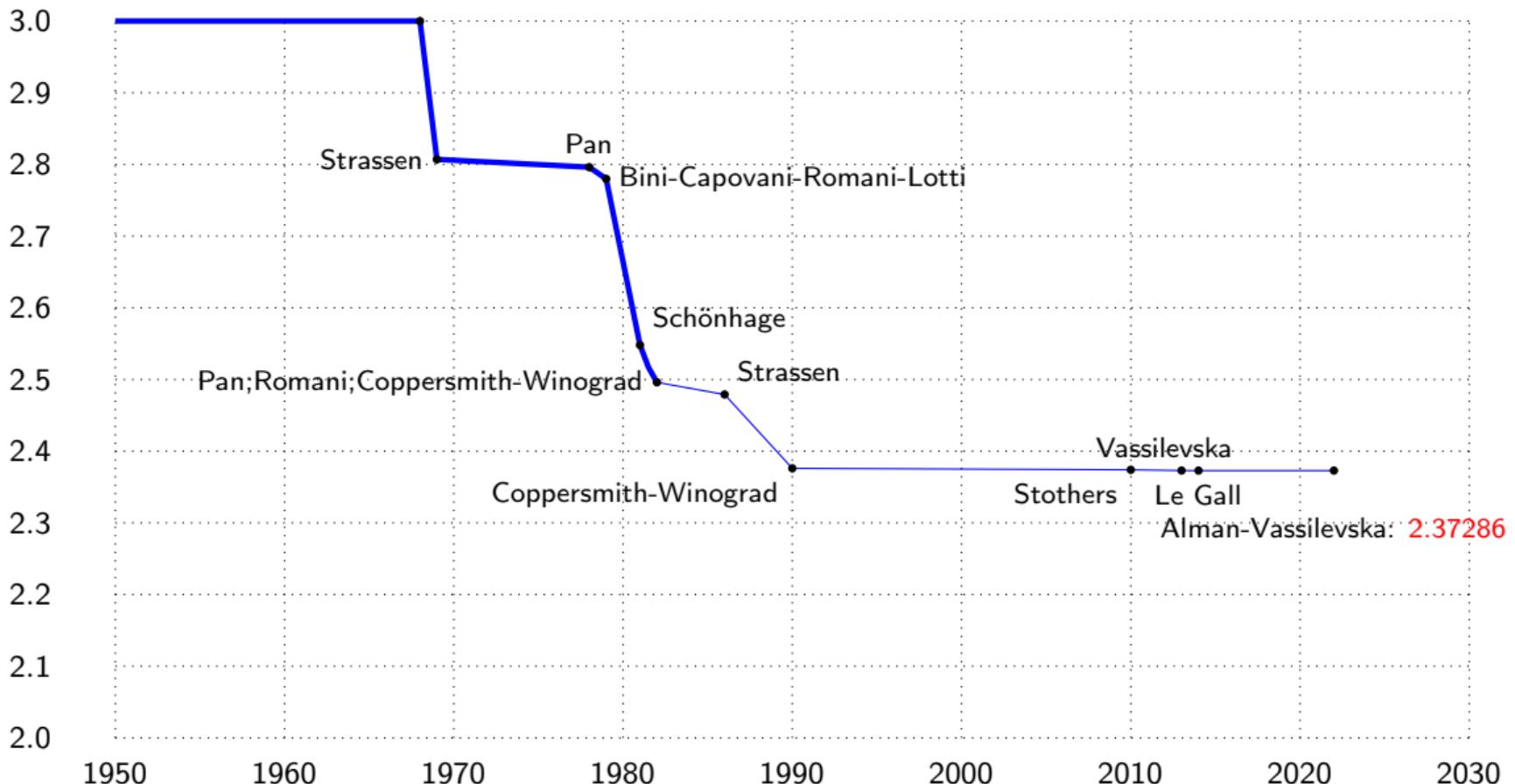
Then $(\ell_1 m_1 n_1)^{\omega/3} + \cdots + (\ell_t m_t n_t)^{\omega/3} \leq r$.

Proof.

High powers of T contain many copies of the same matrix multiplication. □

$$\underline{R}(\langle 3, 1, 3 \rangle \oplus \langle 1, 4, 1 \rangle) \leq 10$$

So $9^{\omega/3} + 4^{\omega/3} \leq 10$. It follows that $\omega \leq 2.594$.



Coppersmith-Winograd tensor

Fix positive integer q . Coppersmith-Winograd tensor:

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q x_0y_iz_i + \sum_{i=1}^q x_iy_0z_i + \sum_{i=1}^q x_iy_iz_0$$

Coppersmith-Winograd tensor

Fix positive integer q . Coppersmith-Winograd tensor:

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q x_0y_i z_i + \sum_{i=1}^q x_i y_0 z_i + \sum_{i=1}^q x_i y_i z_0$$

- CW_q is sum of 6 MM-tensors. For example:

$$x_0y_0z_{q+1} \cong \langle 1, 1, 1 \rangle \quad (\text{multiply } 1 \times 1 \text{ and } 1 \times 1 \text{ 'matrices'})$$

$$\sum_{i=1}^q x_i y_0 z_i \cong \langle q, 1, q \rangle \quad (\text{multiply } q \times 1 \text{ and } 1 \times q \text{ matrices})$$

- But not a **direct sum**.

Coppersmith-Winograd tensor

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_i z_i + x_i y_0 z_i + x_i y_i z_0)$$

Partition variables

$$X_0 = \{x_0\}, X_1 = \{x_1, \dots, x_q\}, X_2 = \{x_{q+1}\}$$

$$Y_0 = \{y_0\}, Y_1 = \{y_1, \dots, y_q\}, Y_2 = \{y_{q+1}\}$$

$$Z_0 = \{z_0\}, Z_1 = \{z_1, \dots, z_q\}, Z_2 = \{z_{q+1}\}$$



Denote by T_{ijk} restriction of CW_q to X_i, Y_j, Z_k .

Then T_{ijk} is a MM-tensor if $i + j + k = 2$ and zero otherwise.

Laser method

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_iz_i + x_iy_0z_i + x_iy_iz_0)$$

Can show: $R(CW_q) \leq q + 2$.

Laser method

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_i z_i + x_i y_0 z_i + x_i y_i z_0)$$

Can show: $\underline{R}(CW_q) \leq q + 2$.

Laser method

Basic tensor: CW_q . Consider $T = (CW_q)^{\otimes n}$ for large n .

Laser method

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_i z_i + x_i y_0 z_i + x_i y_i z_0)$$

Can show: $\underline{R}(CW_q) \leq q + 2$.

Laser method

Basic tensor: CW_q . Consider $T = (CW_q)^{\otimes n}$ for large n .

- $\underline{R}(T) \leq (q + 2)^n$ is ‘small’.

Laser method

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_i z_i + x_i y_0 z_i + x_i y_i z_0)$$

Can show: $\underline{R}(CW_q) \leq q + 2$.

Laser method

Basic tensor: CW_q . Consider $T = (CW_q)^{\otimes n}$ for large n .

- $\underline{R}(T) \leq (q + 2)^n$ is ‘small’.
- T consists of $3^n \times 3^n \times 3^n$ blocks. Each block is a MM-tensor or zero.

Laser method

$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_i z_i + x_i y_0 z_i + x_i y_i z_0)$$

Can show: $\underline{R}(CW_q) \leq q + 2$.

Laser method

Basic tensor: CW_q . Consider $T = (CW_q)^{\otimes n}$ for large n .

- $\underline{R}(T) \leq (q + 2)^n$ is ‘small’.
- T consists of $3^n \times 3^n \times 3^n$ **blocks**. Each block is a MM-tensor or zero.
- Can find a large direct sum of MM-tensors inside
(Uses the ‘course structure’ and [Salem-Spencer sets](#)).

Laser method

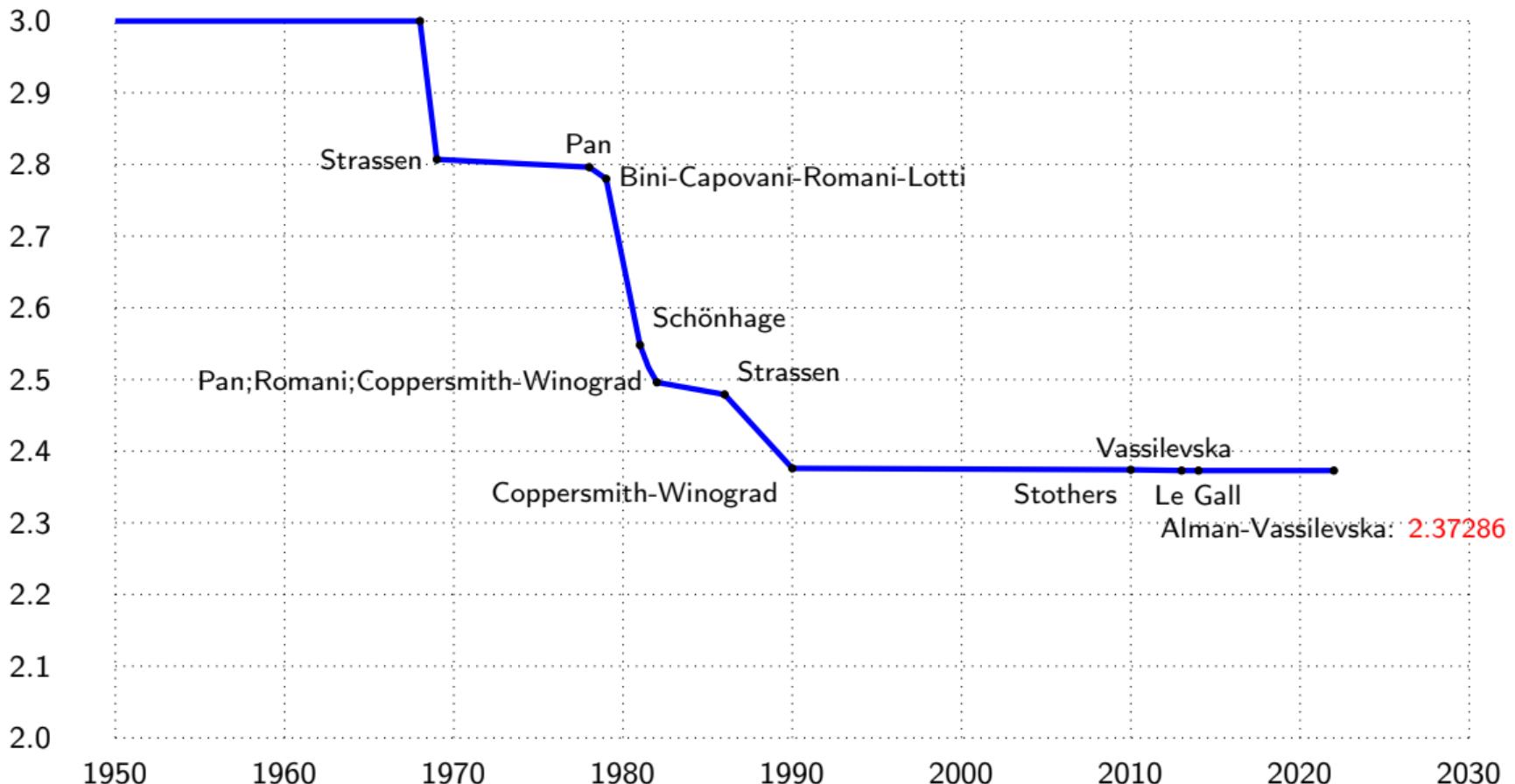
$$CW_q = x_0y_0z_{q+1} + x_0y_{q+1}z_0 + x_{q+1}y_0z_0 + \sum_{i=1}^q (x_0y_iz_i + x_iy_0z_i + x_iy_iz_0)$$

Can show: $\underline{R}(CW_q) \leq q + 2$.

Laser method

Basic tensor: CW_q . Consider $T = (CW_q)^{\otimes n}$ for large n .

- $\underline{R}(T) \leq (q + 2)^n$ is ‘small’.
- T consists of $3^n \times 3^n \times 3^n$ blocks. Each block is a MM-tensor or zero.
- Can find a large direct sum of MM-tensors inside
(Uses the ‘course structure’ and Salem-Spencer sets).
- By Schönhage’s theorem we get bound on ω .

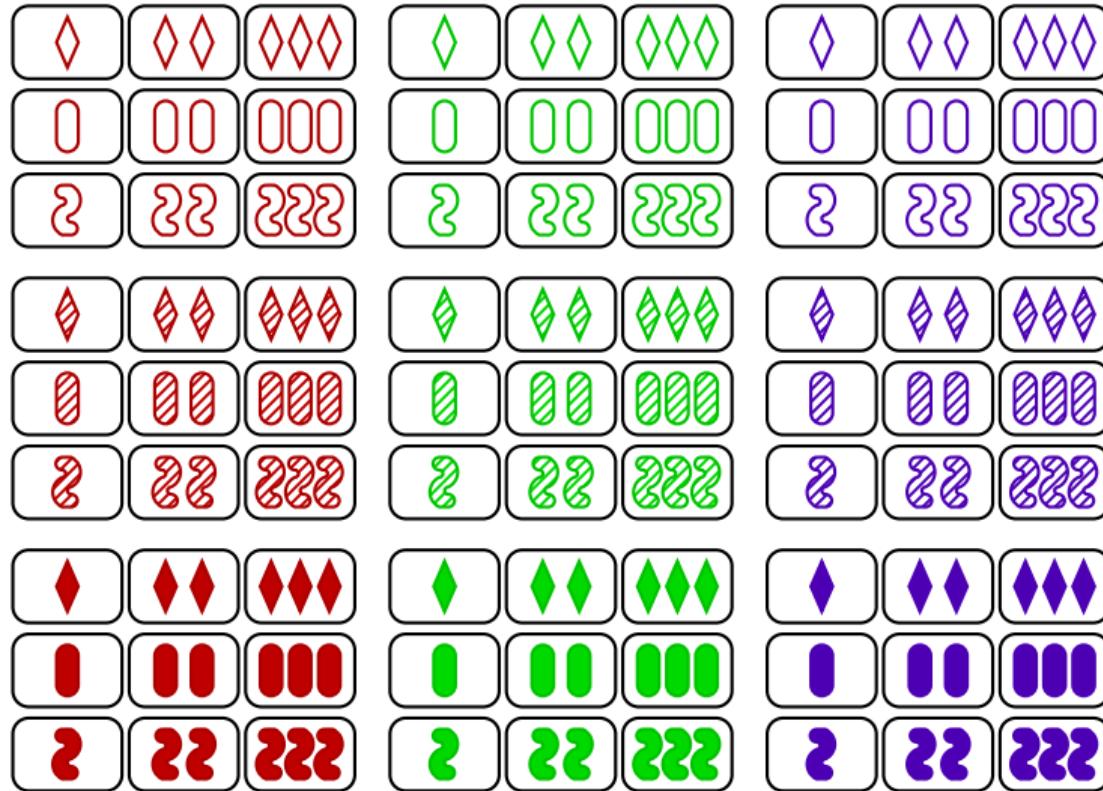


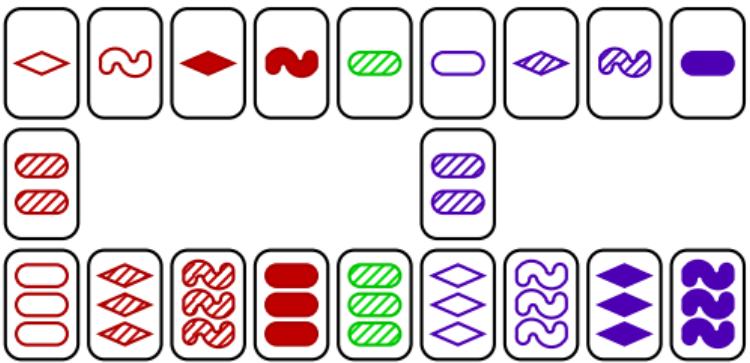
Could it be that $\omega = 2$?

Several ideas and conjectures for reaching $\omega = 2$:

- CW equivoluminous subsets conjecture
(would disprove Erdős-Szemerédi sunflower conjecture and give large **cap sets**).
- Existence of large **multicoloured sum-free sets**
(Context of Cohn-Umans group-algebra approach)
- Better analysis of (powers of) CW_q . Perhaps different basic tensor?

Cap set problem & Slice rank





Pellegrino cap (1971)

	R	G	B
1	• . .	. • .	. . •
2	• •
3	• •

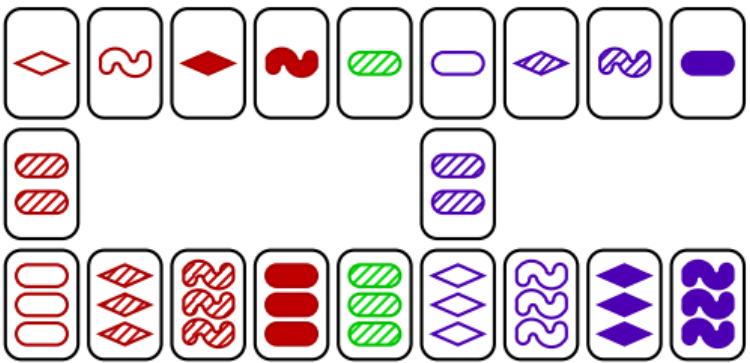
Cap set problem

Cap set:

subset $A \subseteq \mathbb{F}_3^n$ that has no three points on a line.

Cap set Problem:

is there a $c < 3$ such that every cap set has size $O(c^n)$.



Pellegrino cap (1971)

	R	G	B
1	• . .	. • .	• . .
2	•	• . .
3	•	• . .

Cap set problem

Cap set:

subset $A \subseteq \mathbb{F}_3^n$ that has no three points on a line.

Cap set Problem:

is there a $c < 3$ such that every cap set has size $O(c^n)$.

Theorem (Ellenberg-G. 2016)

Cap sets in \mathbb{F}_3^n have size $O(2.756^n)$.

Could it be that $\omega = 2$?

Several ideas and conjectures for reaching $\omega = 2$:

- CW equivoluminous subsets conjecture
(would disprove Erdős-Szemerédi sunflower conjecture and give large cap sets).
- Existence of large multicoloured sum-free sets
(Context of Cohn-Umans group-algebra approach)
- Better analysis of (powers of) CW_q . Perhaps different basic tensor?

Could it be that $\omega = 2$?

Several ideas and conjectures for reaching $\omega = 2$:

- CW equivoluminous subsets conjecture FALSE
(would disprove Erdős-Szemerédi sunflower conjecture and give large cap sets).
- Existence of large multicoloured sum-free sets
(Context of Cohn-Umans group-algebra approach)
- Better analysis of (powers of) CW_q . Perhaps different basic tensor?

Could it be that $\omega = 2$?

Several ideas and conjectures for reaching $\omega = 2$:

- CW equivoluminous subsets conjecture FALSE
(would disprove Erdős-Szemerédi sunflower conjecture and give large cap sets).
- Existence of large multicoloured sum-free sets FALSE (Blasiak et al)
(Context of Cohn-Umans group-algebra approach)
- Better analysis of (powers of) CW_q . Perhaps different basic tensor?

Slice rank

The proof of Cap set Theorem was reformulated by Tao in terms of [slice rank](#).

Slice rank

The proof of Cap set Theorem was reformulated by Tao in terms of slice rank.

Tensor T has slice rank 1 if of the form:

$$\left(\sum_i \alpha_i x_i \right) \cdot \left(\sum_{j,k} \beta_{jk} y_j z_k \right) \quad \text{or} \quad \left(\sum_j \alpha_j y_j \right) \cdot \left(\sum_{i,k} \beta_{ik} x_i z_k \right) \quad \text{or} \quad \left(\sum_k \alpha_k z_k \right) \cdot \left(\sum_{i,j} \beta_{ij} x_i y_j \right)$$

Slice rank

The proof of Cap set Theorem was reformulated by Tao in terms of slice rank.

Tensor T has slice rank 1 if of the form:

$$\left(\sum_i \alpha_i x_i \right) \cdot \left(\sum_{j,k} \beta_{jk} y_j z_k \right) \quad \text{or} \quad \left(\sum_j \alpha_j y_j \right) \cdot \left(\sum_{i,k} \beta_{ik} x_i z_k \right) \quad \text{or} \quad \left(\sum_k \alpha_k z_k \right) \cdot \left(\sum_{i,j} \beta_{ij} x_i y_j \right)$$

Slice rank satisfies:

- $SR(\langle \ell, m, n \rangle) = \min(\ell m, \ell n, mn)$.

Slice rank

The proof of Cap set Theorem was reformulated by Tao in terms of slice rank.

Tensor T has slice rank 1 if of the form:

$$\left(\sum_i \alpha_i x_i \right) \cdot \left(\sum_{j,k} \beta_{jk} y_j z_k \right) \quad \text{or} \quad \left(\sum_j \alpha_j y_j \right) \cdot \left(\sum_{i,k} \beta_{ik} x_i z_k \right) \quad \text{or} \quad \left(\sum_k \alpha_k z_k \right) \cdot \left(\sum_{i,j} \beta_{ij} x_i y_j \right)$$

Slice rank satisfies:

- $SR(\langle \ell, m, n \rangle) = \min(\ell m, \ell n, mn)$.
- Can upper bound $SR(CW_q^{\otimes n})$ by a counting argument.

Slice rank

The proof of Cap set Theorem was reformulated by Tao in terms of slice rank.

Tensor T has slice rank 1 if of the form:

$$\left(\sum_i \alpha_i x_i \right) \cdot \left(\sum_{j,k} \beta_{jk} y_j z_k \right) \quad \text{or} \quad \left(\sum_j \alpha_j y_j \right) \cdot \left(\sum_{i,k} \beta_{ik} x_i z_k \right) \quad \text{or} \quad \left(\sum_k \alpha_k z_k \right) \cdot \left(\sum_{i,j} \beta_{ij} x_i y_j \right)$$

Slice rank satisfies:

- $SR(\langle \ell, m, n \rangle) = \min(\ell m, \ell n, mn)$.
- Can upper bound $SR(CW_q^{\otimes n})$ by a counting argument.
- Can lower bound slice rank of direct sums of MM-tensors.

Slice rank

The proof of Cap set Theorem was reformulated by Tao in terms of slice rank.

Tensor T has slice rank 1 if of the form:

$$\left(\sum_i \alpha_i x_i\right) \cdot \left(\sum_{j,k} \beta_{jk} y_j z_k\right) \quad \text{or} \quad \left(\sum_j \alpha_j y_j\right) \cdot \left(\sum_{i,k} \beta_{ik} x_i z_k\right) \quad \text{or} \quad \left(\sum_k \alpha_k z_k\right) \cdot \left(\sum_{i,j} \beta_{ij} x_i y_j\right)$$

Slice rank satisfies:

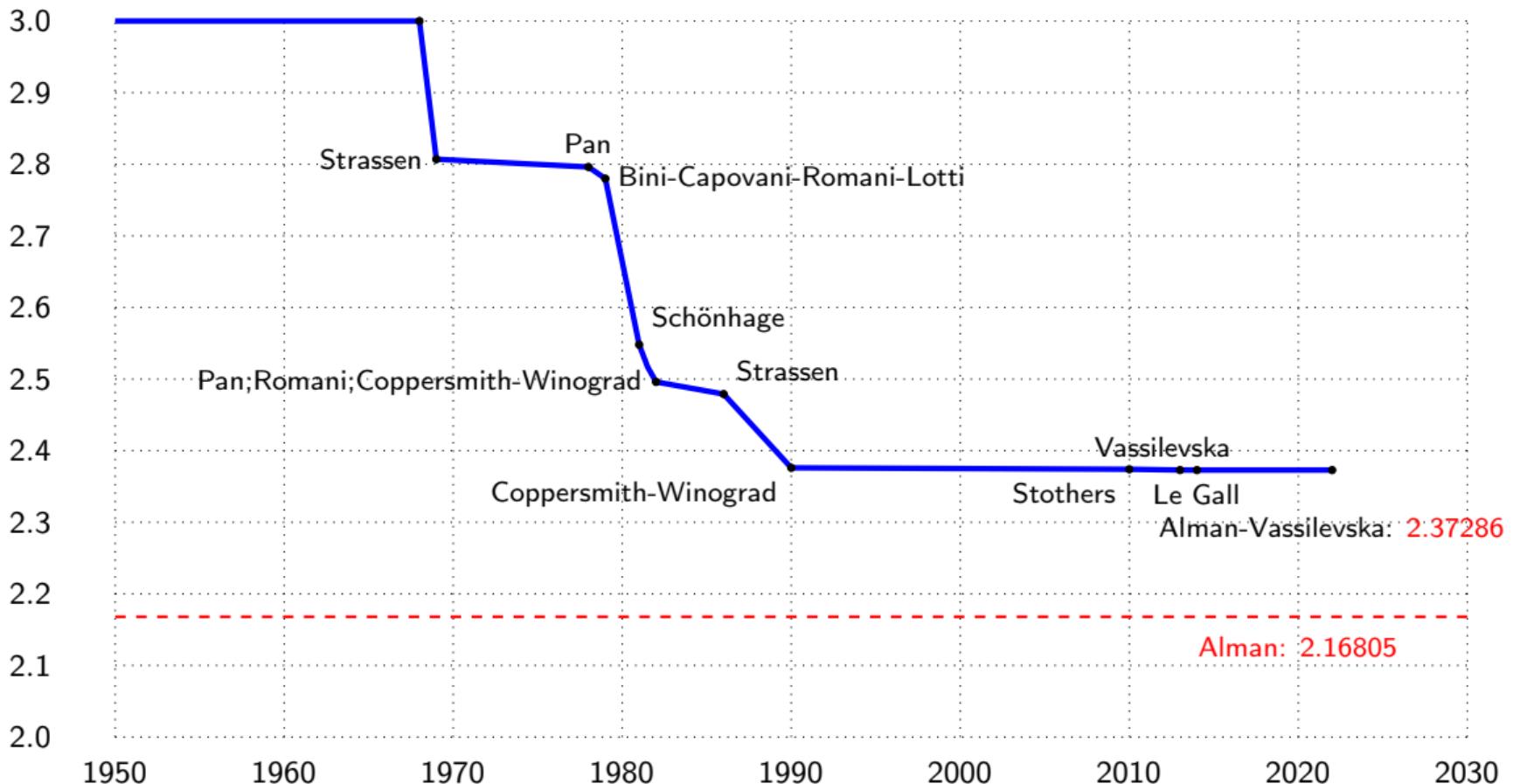
- $SR(\langle \ell, m, n \rangle) = \min(\ell m, \ell n, mn)$.
- Can upper bound $SR(CW_q^{\otimes n})$ by a counting argument.
- Can lower bound slice rank of direct sums of MM-tensors.

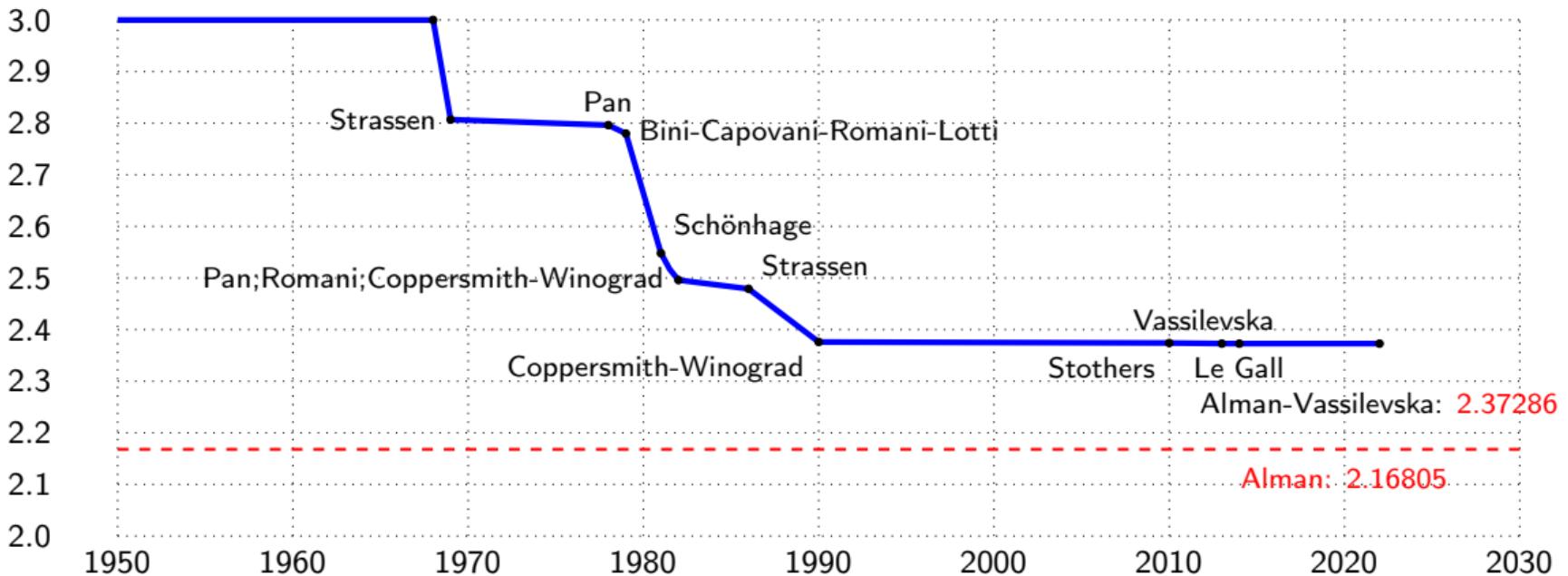
Barriers for methods

In recent years: various barriers to fast matrix multiplication.

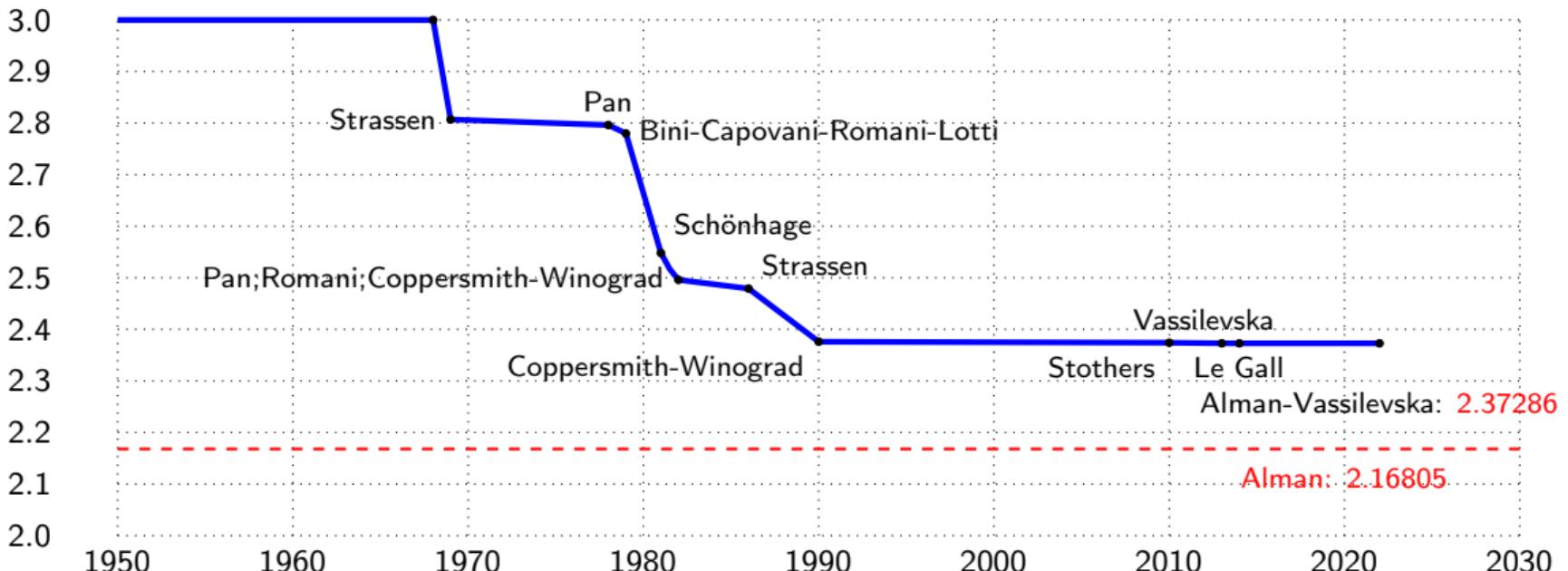
Using slice rank Alman and Alman-Vassilevska showed:

Current methods applied to CW_q cannot go beyond $\omega \leq 2.16805$.





Open problem: Is $\omega = 2$?



Open problem: Is $\omega = 2$?

Thanks!