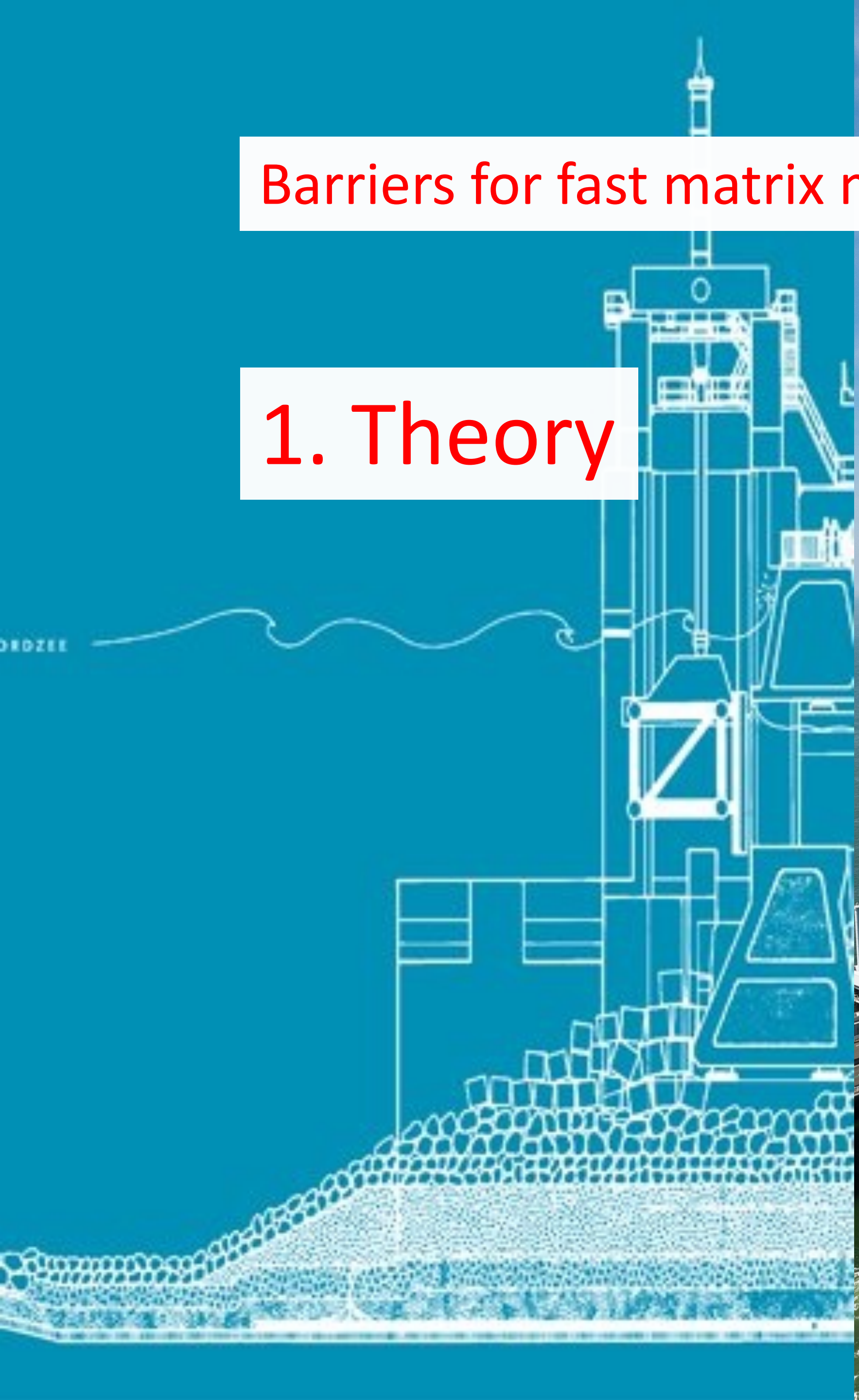
Part two

Jeroen Zuiddam (IAS)

Barriers for fast matrix multiplication

with
Matthias Christandl (Copenhagen)
Péter Vrana (Budapest)

Barriers for fast matrix multiplication



1. Theory

2. Barrier

3. Tools

1.1 Matrix multiplication

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}$$

ω

$O(n^{?})$ multiplications instead of $O(n^3)$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

7 multiplications instead of 8 Strassen 1969

block-wise multiplication

$$\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{pmatrix}$$

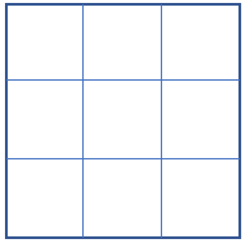
$O(n^{\log_2 7}) = O(n^{2.81})$ multiplications instead of n^3

approximation,
parallel computation of several matrix multiplications,
arithmetic progressions

$O(n^{2.372864})$ Coppersmith and Winograd 1990, Stothers 2010, V-Williams 2012, Le Gall 2014

$2 \leq \omega \leq 2.372864$ Is ω equal to 2?

Matrices



$M \leq N$ if $M = A \cdot N \cdot B$

matrix rank $R(M)$

min r $\quad M = \sum_{i=1}^{r} u_i \otimes v_i = A \cdot I_r \cdot B$

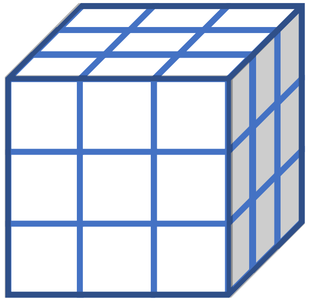
$M \leq I_r$

max r $\quad I_r = A \cdot M \cdot B$

$I_r \leq M$

Gaussian elimination!

Tensors



$S \leq T$ if $S = (A, B, C) \cdot T$

tensor rank $\mathrm{R}(S)$

min r $\quad S = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$

$S \leq \langle r \rangle$ $\quad r \times r \times r$ identity tensor

subrank $\mathrm{Q}(S)$ $\quad$ different notion!

max r $\quad \langle r \rangle \leq S$

1.2 Matrix multiplication tensor

collection of bilinear forms

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

r multiplications

$$\underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{e_{ij}} \cdot \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}}_{e_{jk}} = \underbrace{\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}_{e_{ik}}$$

$$\sum_{i,j,k=1}^{n} e_{ij} \otimes e_{jk} \otimes e_{ik} =: \langle n, n, n \rangle$$

$$\in \mathbb{F}^{n^2} \otimes \mathbb{F}^{n^2} \otimes \mathbb{F}^{n^2}$$

$$\langle n, n, n \rangle = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$$

1.3 "Universal method"

$\mathrm{R}(\langle n, n, n\rangle) \le r$

i.e. $\langle n, n, n\rangle \le \langle r\rangle$ $\quad$ r small $\quad \Rightarrow \quad$ fast matrix multiplication

Universal method:

$\langle n, n, n\rangle \le T^{\otimes k} \le \langle r\rangle$ $\quad$ r small $\quad \Rightarrow \quad$ fast matrix multiplication

$T^{\otimes k}$ is the Kronecker product analogous to the Kronecker product $M^{\otimes k}$ for matrices

1.4 Popular and successful T

$\langle n,n,n\rangle \leq T^{\otimes k}$	T
$\omega \leq 2.8$ [Strassen 1969]	$\langle r\rangle = \sum_{i=1}^{r} e_i e_i e_i$
$\omega \leq 2.48$ [Strassen 1986]	$S_q = \sum_{i=1}^{q} e_i e_0 e_i + e_0 e_i e_i$
$\omega \leq 2.41$ [CW 1990]	$\mathrm{cw}_q = \sum_{i=1}^{q}(e_i e_0 e_i + e_0 e_i e_i + e_i e_i e_0)$
$\omega \leq 2.372864$ [CW 1990,…,Le Gall 2014]	$\mathrm{CW}_q = \sum_{i=1}^{q}(e_i e_0 e_i + e_0 e_i e_i + e_i e_i e_0)$ $+ e_{q+1} e_0 e_0 + e_0 e_{q+1} e_0 + e_0 e_0 e_{q+1}$

Barriers for fast matrix multiplication

2. Barrier

2.1 Barrier theorem

Universal method:

$\langle n, n, n \rangle \leq T^{\otimes m} \leq \langle r \rangle$ r small $\Rightarrow$ fast matrix multiplication

The *universal method* with $T = \mathrm{CW}_q$ can *at best* prove

$$\omega \leq 2.16$$

[Alman]
[Christandl, Vrana and Zuiddam]

Compare with: $\omega \leq 2.372864$

2.2 Source of barriers: subrank

Amazing and crucial subrank fact [Strassen]

$\mathrm{Q}(\langle n,n,n\rangle) = n^2$ i.e. $\langle n^2\rangle \leq \langle n,n,n\rangle$ roughly

Proof: Salem–Spencer set

Intuition of barrier:

- Clearly $\mathrm{Q}(\langle n^2\rangle) = \mathrm{R}(\langle n^2\rangle)$
- Imagine $\omega = 2$, then $\mathrm{Q}(\langle n,n,n\rangle) = \mathrm{R}(\langle n,n,n\rangle) = n^2$ roughly
- If $\mathrm{Q}\left(T^{\otimes m}\right) \ll \mathrm{R}\left(T^{\otimes m}\right)$, then T does not have enough *quality* to satisfy

$$\langle n,n,n\rangle \leq T^{\otimes m} \leq \langle n^2\rangle$$

2.3 General barrier theorem

[Christandl, Vrana and Zuiddam]
[Alman]

Universal method:

$\langle n,n,n\rangle \leq T^{\otimes m} \leq \langle r\rangle$ $\quad r$ small $\quad\Rightarrow\quad$ fast matrix multiplication

The *universal method* with T can *at best* prove*

$$\omega \leq 2 \cdot \frac{\log_2 \widetilde{\mathrm{R}}(T)}{\log_2 \widetilde{\mathrm{Q}}(T)}$$

$$\widetilde{\mathrm{R}}(T) \coloneqq \inf_n \mathrm{R}\left(T^{\otimes n}\right)^{1/n}$$

$$\widetilde{\mathrm{Q}}(T) \coloneqq \sup_n \mathrm{Q}\left(T^{\otimes n}\right)^{1/n}$$

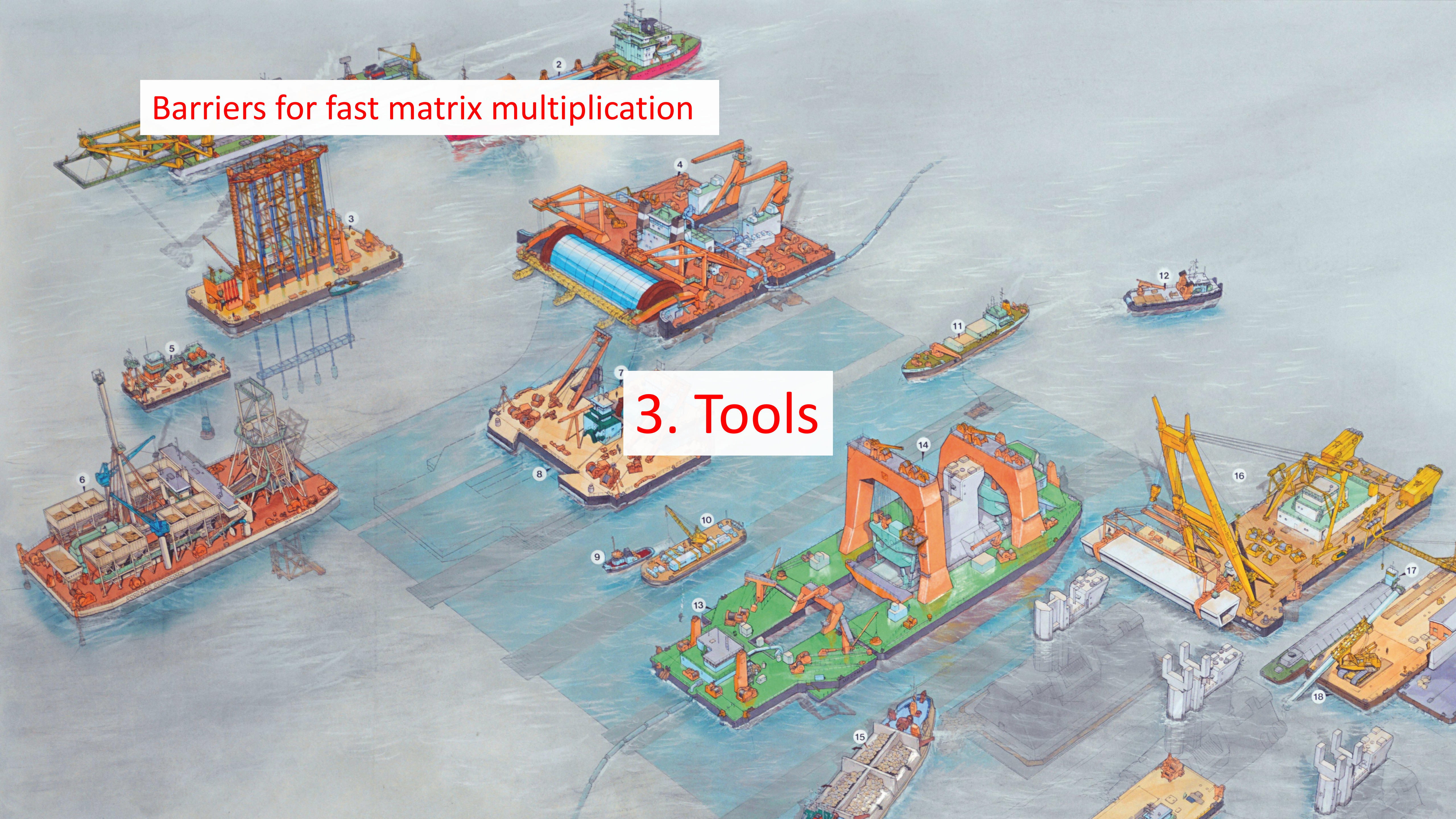
Proof sketch:

$\langle n^2\rangle \leq \langle n,n,n\rangle \leq T^{\otimes m}$

$\widetilde{\mathrm{Q}}(T)$ small $\Rightarrow$ $\mathrm{Q}(T^{\otimes m})$ small $\Rightarrow$ n small $\Rightarrow$ bad bound on ω

Barriers for fast matrix multiplication

3. Tools



3.1 Tools to compute barriers

$$\text{barrier}(T) := 2 \cdot \frac{\log_2 \widetilde{\mathrm{R}}(T)}{\log_2 \widetilde{\mathrm{Q}}(T)}$$

e.g. $T = \mathrm{CW}_q$

$$\widetilde{\mathrm{Q}}(T) \quad \leq \quad f(T) \qquad < \qquad g(T) \quad \leq \quad \widetilde{\mathrm{R}}(T)$$

- **support functionals** [Strassen 1986]
- **quantum functionals** [Christandl—Vrana—Zuiddam 2018]
- instability (from GIT) [Blasiak et al.]
- slice rank (from cap set problem) [Tao, Alman—V-Williams]

- flattening ranks

3.2 Tools for $\widetilde{\mathrm{Q}}(T) \le f(T)$

i.e. $k \to \infty$

Information-theoretic study of:

support functional $Z(T)$ — support of $T^{\otimes k}$ [Strassen]

quantum functional $F(T)$ / instability — representation-theoretic support of $T^{\otimes k}$

over $\mathbb{C}$, related to moment polytopes, scaling algorithms

[Christandl—Vrana—Zuiddam 2018] [Blasiak et al.]

$\widetilde{\mathrm{Q}}(T) \le Z(T)$

$\widetilde{\mathrm{Q}}(T) \le F(T) \le Z(T)$

no separations known

Best upper bound tools for $\widetilde{\mathrm{Q}}(T)$ that we know of

3.3 General theory of tools

Rich theory of asymptotic properties of tensors:

"Asymptotic spectrum of tensors"

[Strassen 1986]

Multiplicative, additive, normalized, ≤-monotone real functions F

$$\widetilde{\mathrm{Q}}(T) \leq F(T) \leq \widetilde{\mathrm{R}}(T)$$

Duality theorem

Analogous theory in study of Shannon capacity of graphs:

"Asymptotic spectrum of graphs"

[Zuiddam 2019]

e.g. Lovász theta number, fractional clique cover number, …

3.4 Example: barrier for big CW_q

$$\mathrm{CW}_q := e_0 e_0 e_{q+1} + e_0 e_{q+1} e_0 + e_{q+1} e_0 e_0 + \sum_{i=1}^{q} e_0 e_i e_i + e_i e_0 e_i + e_i e_i e_0$$

flattening

$$\widetilde{\mathrm{R}}(\mathrm{CW}_q) = q + 2$$

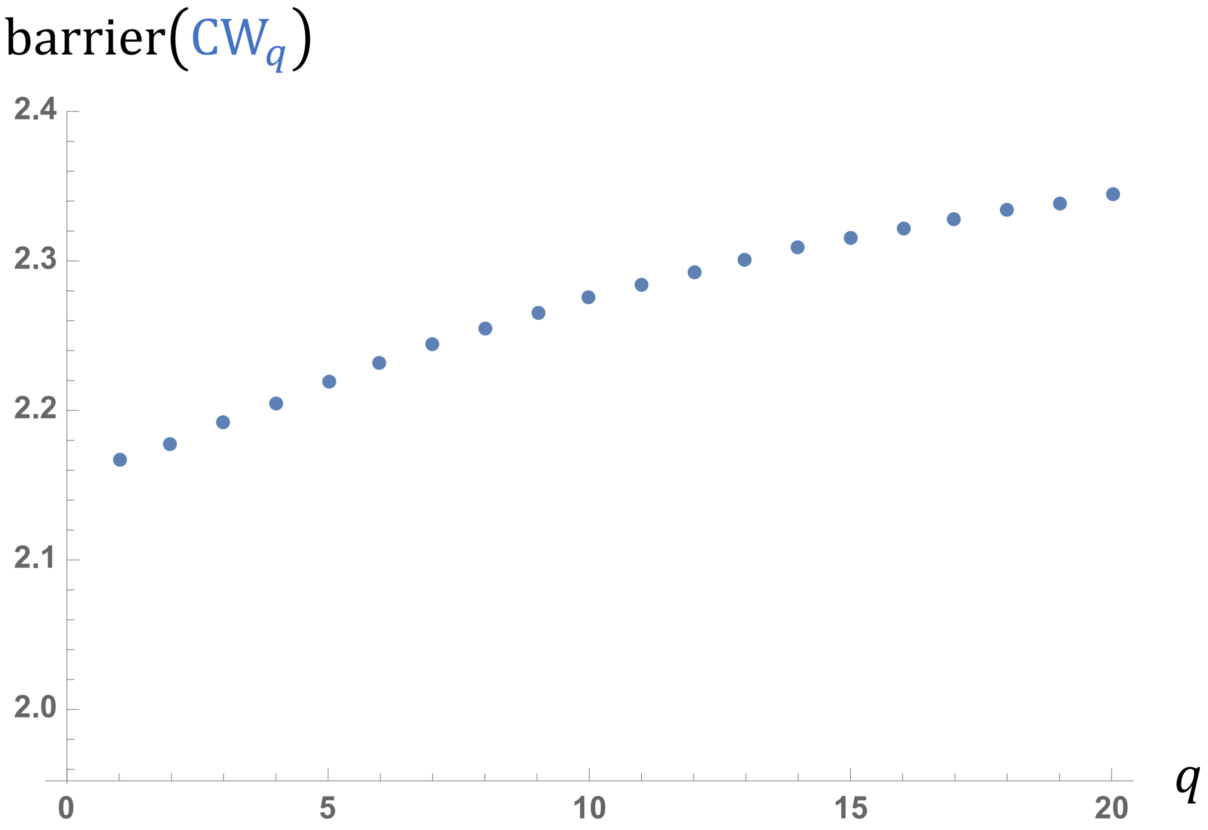
support functional

$$\widetilde{\mathrm{Q}}(\mathrm{CW}_q) \leq Z(\mathrm{CW}_q)$$

$$\mathrm{barrier}(\mathrm{CW}_q) \geq 2.16$$

minimum at $q = 2$

Josh proves inequality is tight
via Laser method



3.5 Example: barrier for small cw_q

$$\mathrm{cw}_q \coloneqq \sum_{i=1}^{q} e_0 e_i e_i + e_i e_0 e_i + e_i e_i e_0$$

flattening [CW90]

$$q + 1 \leq \widetilde{\mathrm{R}}(\mathrm{cw}_q) \leq q + 2$$

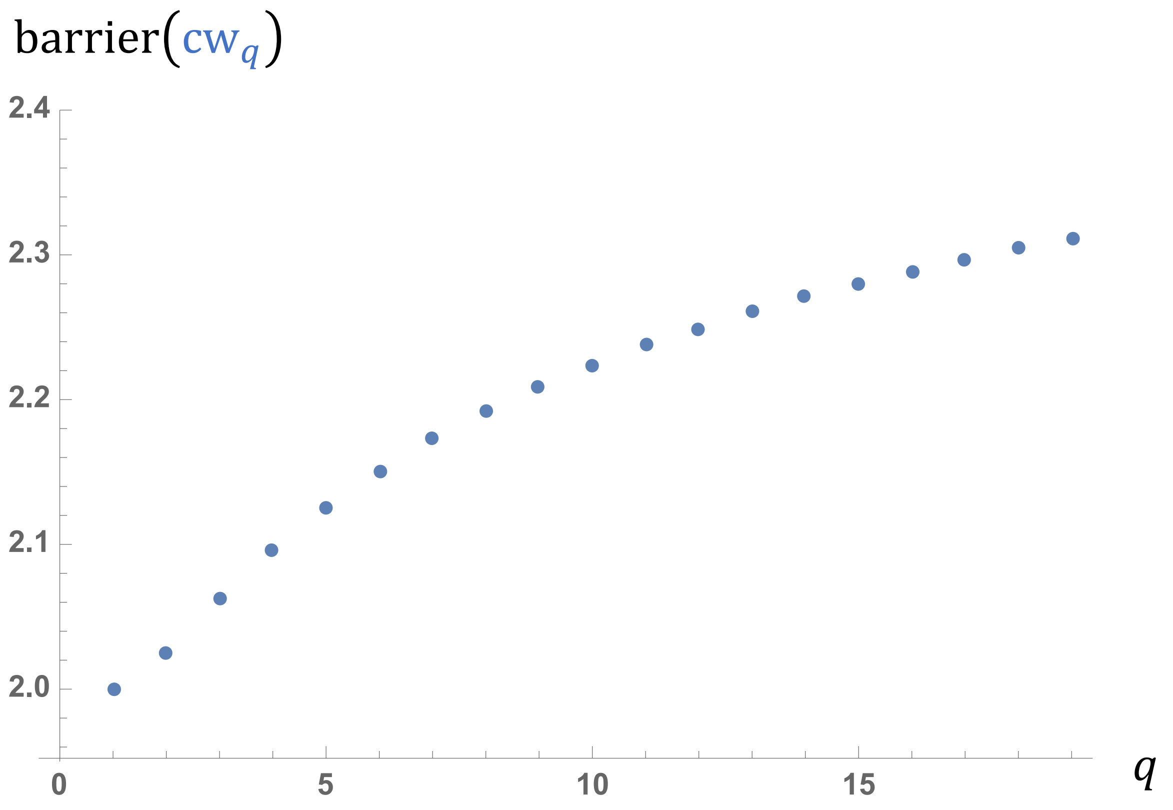
support functional

$$\widetilde{\mathrm{Q}}(\mathrm{cw}_q) \leq Z(\mathrm{cw}_q)$$

$$\mathrm{barrier}(\mathrm{cw}_q) \begin{cases} \geq 2.02 & q > 2 \\ = 2 & q = 2 \end{cases}$$

$$\widetilde{\mathrm{R}}(\mathrm{cw}_2) = 3 \Rightarrow \omega = 2$$

$$\widetilde{\mathrm{R}}(\mathrm{cw}_q) = q + 2 \Rightarrow \mathrm{barrier}(\mathrm{cw}_q) \geq 2.27$$



Conclusion: promising T to prove $\omega = 2$

T with $\widetilde{\mathrm{Q}}(T) = \widetilde{\mathrm{R}}(T)$

Example
$\widetilde{\mathrm{Q}}(\langle n \rangle) = \widetilde{\mathrm{R}}(\langle n \rangle) = n$

Example
$\omega = 2 \;\Rightarrow\; \widetilde{\mathrm{Q}}(\langle n,n,n \rangle) = \widetilde{\mathrm{R}}(\langle n,n,n \rangle)$
$= n^2$

Problem 1
other T with $\widetilde{\mathrm{Q}}(T) = \widetilde{\mathrm{R}}(T)$?

$T_1, T_2, \ldots$ with $\widetilde{\mathrm{R}}(T_i) \,/\, \widetilde{\mathrm{Q}}(T_i) \to 1$

Examples exist

Problem 2
use those T_i to upper bound ω

e.g. with group-theoretic method