Powers of Tensors and Fast Matrix Multiplication

François Le Gall

Department of Computer Science
Graduate School of Information Science and Technology
The University of Tokyo

Simons Institute, 12 November 2014

Overview of our Results

Algebraic Complexity of Matrix Multiplication

Compute the product of two $n \times n$ matrices A and B over a field $\mathbb{F}$

- Model: algebraic circuits
  - gates: +, −, ×, ÷ (operations on two elements of the field)
  - input: a_{ij}, b_{ij} $(2n^2$ inputs)
  - output: $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$ $(n^2$ outputs)

$\mathcal{C}_M(n) =$ minimal number of algebraic operations needed to compute the product

note: may depend on the field $\mathbb{F}$

Exponent of matrix multiplication

$$\omega = \inf \left\{ \alpha \,|\, \mathcal{C}_M(n) \leq n^\alpha \text{ for all large enough } n \right\}$$

note: may depend on the field $\mathbb{F}$

Obviously, $2 \leq \omega \leq 3$.

History of the main improvements on the exponent of square matrix multiplication

Upper bound	Year	Authors
$\omega \leq 3$		
$\omega < 2.81$	1969	Strassen
$\omega < 2.79$	1979	Pan
$\omega < 2.78$	1979	Bini, Capovani, Romani and Lotti
$\omega < 2.55$	1981	Schönhage
$\omega < 2.53$	1981	Pan
$\omega < 2.52$	1982	Romani
$\omega < 2.50$	1982	Coppersmith and Winograd
$\omega < 2.48$	1986	Strassen
$\omega < 2.376$	1987	Coppersmith and Winograd
$\omega < 2.373$	2010	Stothers
$\omega < 2.3729$	2012	Vassilevska Williams
$\omega < 2.3728639$	2014	Le Gall

This work

History of the main improvements on the exponent of square matrix multiplication

Upper bound	Year	Authors
$\omega \leq 3$		
$\omega < 2.81$	1969	Strassen
$\omega < 2.79$	1979	Pan
$\omega < 2.78$	1979	Bini, Capovani, Romani and Lotti
$\omega < 2.55$	1981	Schönhage
$\omega < 2.53$	1981	Pan
$\omega < 2.52$	1982	Romani
$\omega < 2.50$	1982	Coppersmith and Winograd
$\omega < 2.48$	1986	Strassen
$\omega < 2.376$	1987	Coppersmith and Winograd
$\omega < 2.373$	2010	Stothers
$\omega < 2.3729$	2012	Vassilevska Williams
$\omega < 2.3728639$	2014	Le Gall

analysis of a tensor by the laser method

History of the main improvements on the exponent of square matrix multiplication

analysis of the tensor by the laser method (LM)

$\omega < 2.48$	1986	Strassen	LM-based analysis v1
$\omega < 2.376$	1987	Coppersmith and Winograd	LM-based analysis v2.0
$\omega < 2.373$	2010	Stothers	LM-based analysis v2.1
$\omega < 2.3729$	2012	Vassilevska Williams	LM-based analysis v2.2
$\omega < 2.3728639$	2014	Le Gall	LM-based analysis v2.3

The tensors considered become more difficult to analyze
(technical difficulties appear + the "size" of the tensor increases)

Previous versions (up to v2.2):

analyzing the tensor required solving a complicated optimization problem (difficult when the size of the tensor increases)

Our new technique (v2.3):

analyzing the tensor (i.e., obtaining an upper bound on ω from it) can be done in time polynomial in the size of the tensor

- analysis based on convex optimization

Applications of our method

any tensor from which an upper bound on ω can be obtained from the laser method

Laser-method-based analysis v2.3

polynomial time

corresponding upper bound on ω

which tensor? powers of the basic tensor from Coppersmith and Winograd's paper

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)

$\omega < 2.48$	1986	Strassen	LM-based analysis v1
$\omega < 2.376$	1987	Coppersmith and Winograd	LM-based analysis v2.0

Applications of our method

any tensor from which an upper bound on ω can be obtained from the laser method

Laser-method-based analysis v2.3

$\longrightarrow$

polynomial time

corresponding upper bound on ω

which tensor? powers of the basic tensor from Coppersmith and Winograd's paper

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)

$\omega < 2.48$	1986	Strassen	LM-based analysis v1
$\omega < 2.376$	1987	Coppersmith and Winograd	LM-based analysis v2.0
$\omega < 2.373$	2010	Stothers	LM-based analysis v2.1

Applications of our method

any tensor from which an upper bound on ω can be obtained from the laser method → (Laser-method-based analysis v2.3, polynomial time) → corresponding upper bound on ω

which tensor? powers of the basic tensor from Coppersmith and Winograd's paper

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)

$\omega < 2.48$	1986	Strassen	LM-based analysis v1
$\omega < 2.376$	1987	Coppersmith and Winograd	LM-based analysis v2.0
$\omega < 2.373$	2010	Stothers	LM-based analysis v2.1
$\omega < 2.3729$	2012	Vassilevska Williams	LM-based analysis v2.2

Applications of our method

any tensor from which an upper bound on ω can be obtained from the laser method

Laser-method-based analysis v2.3

polynomial time

corresponding upper bound on ω

which tensor? powers of the basic tensor from Coppersmith and Winograd's paper

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

$\omega < 2.48$	1986	Strassen	LM-based analysis v1
$\omega < 2.376$	1987	Coppersmith and Winograd	LM-based analysis v2.0
$\omega < 2.373$	2010	Stothers	LM-based analysis v2.1
$\omega < 2.3729$	2012	Vassilevska Williams	LM-based analysis v2.2
$\omega < 2.3728639$	2014	Le Gall	LM-based analysis v2.3

How to Obtain Upper Bounds on ω ?

Strassen's algorithm (for the product of two 2×2 matrices)

Goal: compute the product of $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ by $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$

1. Compute:

$$\begin{aligned}
m_1 &= a_{11} * (b_{12} - b_{22}), \\
m_2 &= (a_{11} + a_{12}) * b_{22}, \\
m_3 &= (a_{21} + a_{22}) * b_{11}, \\
m_4 &= a_{22} * (b_{21} - b_{11}), \\
m_5 &= (a_{11} + a_{22}) * (b_{11} + b_{22}), \\
m_6 &= (a_{12} - a_{22}) * (b_{21} + b_{22}), \\
m_7 &= (a_{11} - a_{21}) * (b_{11} + b_{12}).
\end{aligned}$$

2. Output:

$$\begin{aligned}
-m_2 + m_4 + m_5 + m_6 &= c_{11}, \\
m_1 + m_2 &= c_{12}, \\
m_3 + m_4 &= c_{21}, \\
m_1 - m_3 + m_5 - m_7 &= c_{22}.
\end{aligned}$$

7 multiplications   18 additions/substractions

Strassen's algorithm (for the product of two $2^k \times 2^k$ matrices)

Goal: compute the product of $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ by $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$

1. Compute:

$$\begin{aligned}
m_1 &= a_{11} * (b_{12} - b_{22}),\\
m_2 &= (a_{11} + a_{12}) * b_{22},\\
m_3 &= (a_{21} + a_{22}) * b_{11},\\
m_4 &= a_{22} * (b_{21} - b_{11}),\\
m_5 &= (a_{11} + a_{22}) * (b_{11} + b_{22}),\\
m_6 &= (a_{12} - a_{22}) * (b_{21} + b_{22}),\\
m_7 &= (a_{11} - a_{21}) * (b_{11} + b_{12}).
\end{aligned}$$

2. Output:

$$\begin{aligned}
-m_2 + m_4 + m_5 + m_6 &= c_{11},\\
m_1 + m_2 &= c_{12},\\
m_3 + m_4 &= c_{21},\\
m_1 - m_3 + m_5 - m_7 &= c_{22}.
\end{aligned}$$

7 multiplications 18 additions/substractions

Recursive application gives $\mathcal{C}_M(2^k) = O(7^k) = O((2^k)^{\log_2 7})$

$\Longrightarrow \boxed{\omega \leq \log_2(7) = 2.807\ldots}$ [Strassen 69]

Strassen's algorithm (for the product of two $2^k \times 2^k$ matrices)

More generally:

Suppose that the product of two $m \times m$ matrices can be computed with t multiplications. Then

$$\omega \leq \log_m(t) \text{ or, equivalently, } m^\omega \leq t.$$

Strassen's algorithm is the case $m = 2$ and $t = 7$

7 multiplications   18 additions/substractions

Recursive application gives $\mathcal{C}_M(2^k) = O(7^k) = O((2^k)^{\log_2 7})$

$$\Longrightarrow \boxed{\omega \leq \log_2(7) = 2.807\ldots}$$ [Strassen 69]

The tensor of matrix multiplication

Definition

The tensor corresponding to the multiplication of an $m \times n$ matrix by an $n \times p$ matrix is

$$\langle m, n, p \rangle = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} a_{ik} \otimes b_{kj} \otimes c_{ij}.$$

intuitive interpretation:

- this is a formal sum
- when the a_{ik} and the b_{kj} are replaced by the corresponding entries of matrices, the coefficient of c_{ij} becomes $\sum_{k=1}^{n} a_{ik} b_{kj}$

General 3-tensors

Consider three vector spaces U, V and W over $\mathbb{F}$

Take bases of U, V and W:

$$U = span\{x_1, \ldots, x_{\dim(U)}\}$$
$$V = span\{y_1, \ldots, y_{\dim(V)}\}$$
$$W = span\{z_1, \ldots, z_{\dim(W)}\}$$

A tensor over (U, V, W) is an element of $U \otimes V \otimes W$

i.e., a formal sum $T = \sum_{u=1}^{\dim(U)} \sum_{v=1}^{\dim(V)} \sum_{w=1}^{\dim(W)} \underbrace{d_{uvw}}_{\in \mathbb{F}} x_u \otimes y_v \otimes z_w$

"a three-dimension array with $\dim(U) \times \dim(V) \times \dim(W)$ entries in $\mathbb{F}$"

General 3-tensors

A tensor over (U, V, W) is an element of $U \otimes V \otimes W$

i.e., a formal sum $T = \sum_{u=1}^{\dim(U)} \sum_{v=1}^{\dim(V)} \sum_{w=1}^{\dim(W)} \underbrace{d_{uvw}}_{\in \mathbb{F}} x_u \otimes y_v \otimes z_w$

$\dim(U) = mn$, $\dim(V) = np$ and $\dim(W) = mp$

$U = span\left\{\{a_{ik}\}_{1 \le i \le m, 1 \le k \le n}\right\}$

$V = span\left\{\{b_{k'j}\}_{1 \le k' \le n, 1 \le j \le p}\right\}$

$W = span\left\{\{c_{i'j'}\}_{1 \le i' \le m, 1 \le j' \le p}\right\}$

$$d_{ikk'ji'j'} = \begin{cases} 1 & \text{if } i = i', j = j', k = k' \\ 0 & \text{otherwise} \end{cases}$$

Definition

The tensor corresponding to the multiplication of an $m \times n$ matrix by an $n \times p$ matrix is

$$\langle m, n, p \rangle = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} a_{ik} \otimes b_{kj} \otimes c_{ij}.$$

Rank

Definition

The tensor corresponding to the multiplication of an $m \times n$ matrix by an $n \times p$ matrix is

$$\langle m, n, p \rangle = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} a_{ik} \otimes b_{kj} \otimes c_{ij}.$$

$$R(\langle m, n, p \rangle) \leq mnp$$

$$\begin{aligned}
\langle 2,2,2 \rangle = & a_{11} \otimes (b_{12} - b_{22}) \otimes (c_{12} + c_{22}) \\
& + (a_{11} + a_{12}) \otimes b_{22} \otimes (-c_{11} + c_{12}) \\
& + (a_{21} + a_{22}) \otimes b_{11} \otimes (c_{21} - c_{22}) \\
& + a_{22} \otimes (b_{21} - b_{11}) \otimes (c_{11} + c_{21}) \\
& + (a_{11} + a_{22}) \otimes (b_{11} + b_{22}) \otimes (c_{11} + c_{22}) \\
& + (a_{12} - a_{22}) \otimes (b_{21} + b_{22}) \otimes c_{11} \\
& + (a_{11} - a_{21}) \otimes (b_{11} + b_{12}) \otimes (-c_{22})
\end{aligned}$$

Strassen's algorithm gives

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

rank = # of multiplications of the best (bilinear) algorithm

How to obtain upper bounds on ω ?

Remember:

Suppose that the product of two $m \times m$ matrices can be computed with t multiplications. Then

$$\omega \leq \log_m(t) \text{ or, equivalently, } m^\omega \leq t.$$

In our terminology: $R(\langle m, m, m \rangle) \leq t \Longrightarrow m^\omega \leq t$

First generalization:

Theorem

$$R(\langle m, n, p \rangle) \leq t \Longrightarrow (mnp)^{\omega/3} \leq t$$

Second generalization:
[Bini et al. 1979]

Theorem

$$\underline{R}(\langle m, n, p \rangle) \leq t \Longrightarrow (mnp)^{\omega/3} \leq t$$

border rank

$\underline{R}(\langle m, n, p \rangle) \leq R(\langle m, n, p \rangle)$

How to obtain upper bounds on ω ?

Third generalization:

Theorem (the asymptotic sum inequality, special case) [Schönhage 1981]

$$\underline{R}(\langle m_1, n_1, p_1\rangle \oplus \langle m_2, n_2, p_2\rangle) \le t \implies (m_1 n_1 p_1)^{\omega/3} + (m_2 n_2 p_2)^{\omega/3} \le t$$

direct sum

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$\underline{R}\left(\bigoplus_{i=1}^{k} \langle m_i, n_i, p_i\rangle\right) \le t \implies \sum_{i=1}^{k} (m_i n_i p_i)^{\omega/3} \le t$$

First generalization:

Theorem

$$R(\langle m, n, p\rangle) \le t \implies (mnp)^{\omega/3} \le t$$

Second generalization:

[Bini et al. 1979]

Theorem

$$\underline{R}(\langle m, n, p\rangle) \le t \implies (mnp)^{\omega/3} \le t$$

border rank

$\underline{R}(\langle m, n, p\rangle) \le R(\langle m, n, p\rangle)$

History of the main improvements on the exponent of square matrix multiplication

Upper bound	Year	Authors
$\omega \leq 3$		
$\omega < 2.81$	1969	Strassen
$\omega < 2.79$	1979	Pan
$\omega < 2.78$	1979	Bini et al.
$\omega < 2.55$	1981	Schönhage
$\omega < 2.53$	1981	Pan
$\omega < 2.52$	1982	Romani
$\omega < 2.50$	1982	Coppersmith and Winograd
$\omega < 2.48$	1986	Strassen
$\omega < 2.376$	1987	Coppersmith and Winograd
$\omega < 2.373$	2010	Stothers
$\omega < 2.3729$	2012	Vassilevska Williams
$\omega < 2.3728639$	2014	Le Gall

upper bound on ω from the analysis of the rank of a tensor

analysis of the border rank of a tensor

analysis of a tensor by the asymptotic sum inequality

analysis of a tensor by the laser method

The Laser Method on a Simpler Example

The "laser method"

Why this is called the "laser method"?

> limited by our ignorance about ω. Surprisingly, the exact knowledge of the left end of Δ_c can be used to obtain an improved estimate for its right end, namely $\omega < 2.48$. The method employed is called *laser method* [27], since it is reminiscent of the generation of coherent light.

from V. Strassen.
Algebra and Complexity.
Proceedings of the first European Congress of Mathematics, pp. 429-446, 1994.

	Upper bound	Year	Authors
	$\omega \leq 3$		
	$\omega < 2.81$	1969	Strassen
	$\omega < 2.79$	1979	Pan
	$\omega < 2.78$	1979	Bini, Capovani, Romani and Lotti
	$\omega < 2.55$	1981	Schönhage
	$\omega < 2.53$	1981	Pan
	$\omega < 2.52$	1982	Romani
	$\omega < 2.50$	1982	Coppersmith and Winograd
Ref. [27]	$\omega < 2.48$	1986	Strassen
	$\omega < 2.376$	1987	Coppersmith and Winograd
	$\omega < 2.373$	2010	Stothers
	$\omega < 2.3729$	2012	Vassilevska Williams
	$\omega < 2.3728639$	2014	Le Gall

variants (improvements) of the laser method

The first CW construction

Let q be a positive integer.

Consider three vector spaces U, V and W of dimension $q+1$ over $\mathbb{F}$.

$$U = span\{x_0, \ldots, x_q\}$$

$$V = span\{y_0, \ldots, y_q\} \qquad W = span\{z_0, \ldots, z_q\}$$

Coppersmith and Winograd (1987) introduced the following tensor:

$$\boxed{T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110},}$$ tensor over (U, V, W)

where

$$T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i \cong \langle 1, 1, q \rangle \qquad T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_{00} \otimes y_{0i} \otimes z_{0i}$$

1×1 matrix by 1×q matrix

$$T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i \cong \langle q, 1, 1 \rangle \qquad T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_{i0} \otimes y_{00} \otimes z_{i0}$$

q×1 matrix by 1×1 matrix

$$T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0 \cong \langle 1, q, 1 \rangle \qquad T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_{0i} \otimes y_{i0} \otimes z_{00}$$

1×q matrix by q×1 matrix

The first CW construction

$U = span\{x_0, \ldots, x_q\}$

$V = span\{y_0, \ldots, y_q\}$ $\quad W = span\{z_0, \ldots, z_q\}$

$$U = U_0 \oplus U_1, \quad \text{where } U_0 = span\{x_0\} \text{ and } U_1 = span\{x_1, \ldots, x_q\}$$
$$V = V_0 \oplus V_1, \quad \text{where } V_0 = span\{y_0\} \text{ and } V_1 = span\{y_1, \ldots, y_q\}$$
$$W = W_0 \oplus W_1, \quad \text{where } W_0 = span\{z_0\} \text{ and } W_1 = span\{z_1, \ldots, z_q\}$$

Coppersmith and Winograd (1987) introduced the following tensor:

$$T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110},$$ ← This is not a direct sum

where

$T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i$ tensor over (U_0, V_1, W_1)

$T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i$ tensor over (U_1, V_0, W_1)

$T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0$ tensor over (U_1, V_1, W_0)

The first CW construction

$$T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}$$

$$\underline{R}(T_{\text{easy}}) \leq q + 2$$

Actually, $\underline{R}(T_{\text{easy}}) = q + 2$

Since the sum in not direct, we cannot use the asymptotic sum inequality

Consider $T_{\text{easy}}^{\otimes 2} = (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}) \otimes (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110})$

$= T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{011} + T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{101} + \cdots + T_{\text{easy}}^{110} \otimes T_{\text{easy}}^{110}$ (9 terms)

Consider $T_{\text{easy}}^{\otimes N} = T_{\text{easy}}^{011} \otimes \cdots \otimes T_{\text{easy}}^{011} + \cdots + T_{\text{easy}}^{110} \otimes \cdots \otimes T_{\text{easy}}^{110}$ (3^N terms)

Note: $\underline{R}(T_{\text{easy}}^{\otimes N}) = (q+1)^{N+o(N)}$ would imply $\omega = 2$

Coppersmith and Winograd showed how to select $\approx \left(\frac{3}{2^{2/3}}\right)^N$ terms that do not share variables (i.e., form a direct sum)

by zeroing variables
(i.e., without increasing the rank)

The first CW construction: Analysis

$$H\left(\tfrac{1}{3},\tfrac{2}{3}\right) = -\tfrac{1}{3}\log\left(\tfrac{1}{3}\right) - \tfrac{2}{3}\log\left(\tfrac{2}{3}\right) \qquad \text{(entropy)}$$

$$= \log\left(3^{1/3}\times\left(\tfrac{3}{2}\right)^{2/3}\right) = \log\left(\tfrac{3}{2^{2/3}}\right)$$

Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^{\otimes N}$ can be converted into a direct sum of

by zoroing variables
(i.e., without increasing the rank)

$$\exp\left(\left(H\left(\frac{1}{3},\frac{2}{3}\right) - o(1)\right)N\right) = \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N}$$

terms, each containing $\boxed{\frac{N}{3} \text{ copies of } T_{\text{easy}}^{011},\ \frac{N}{3} \text{ copies of } T_{\text{easy}}^{101} \text{ and } \frac{N}{3} \text{ copies of } T_{\text{easy}}^{110}.}$

Consider $T_{\text{easy}}^{\otimes N} = \underbrace{T_{\text{easy}}^{011}\otimes\cdots\otimes T_{\text{easy}}^{011}}_{} + \cdots + \underbrace{T_{\text{easy}}^{110}\otimes\cdots\otimes T_{\text{easy}}^{110}}_{}$ (3^N terms)

N copies of T_{easy}^{011}
0 copies of T_{easy}^{101}
0 copies of T_{easy}^{110}

0 copies of T_{easy}^{011}
0 copies of T_{easy}^{101}
N copies of T_{easy}^{110}

The first CW construction: Analysis

Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^{\otimes N}$ can be converted into a direct sum of

$$\exp\left(\left(H\left(\frac{1}{3},\frac{2}{3}\right)-o(1)\right)N\right) = \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N}$$

terms, each containing $\boxed{\frac{N}{3} \text{ copies of } T_{\text{easy}}^{011},\ \frac{N}{3} \text{ copies of } T_{\text{easy}}^{101} \text{ and } \frac{N}{3} \text{ copies of } T_{\text{easy}}^{110}.}$

isomorphic to $\left[T_{\text{easy}}^{011}\right]^{\otimes N/3} \otimes \left[T_{\text{easy}}^{101}\right]^{\otimes N/3} \otimes \left[T_{\text{easy}}^{110}\right]^{\otimes N/3} \cong \langle q^{N/3}, q^{N/3}, q^{N/3}\rangle$

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$\underline{R}\left(\bigoplus_{i=1}^{k}\langle m_i, n_i, p_i\rangle\right) \le t \Longrightarrow \sum_{i=1}^{k}(m_i n_i p_i)^{\omega/3} \le t$$

Consequence: $\left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N} \times q^{N\omega/3} \le \underline{R}(T_{\text{easy}}^{\otimes N}) \le \underline{R}(T_{\text{easy}})^N = (q+2)^N$

$$\Longrightarrow \frac{3}{2^{2/3}} \times q^{\omega/3} \le q+2 \Longrightarrow \boxed{\omega \le 2.403...} \text{ for } q=8$$

Idea behind the proof

$$T_{\text{easy}} = T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}$$

$$T_{\text{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i$$

$$T_{\text{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i$$

$$T_{\text{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0$$

Consider $N = 2$

$$T_{\text{easy}}^{\otimes 2} = (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110}) \otimes (T_{\text{easy}}^{011} + T_{\text{easy}}^{101} + T_{\text{easy}}^{110})$$

$$= T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{011} + \boxed{T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{101}} + \cdots + T_{\text{easy}}^{110} \otimes T_{\text{easy}}^{110} \quad \text{(9 terms)}$$

001111 011011 111100

011110 100111 110011

110110 101101 111001

$$T_{\text{easy}}^{011} \otimes T_{\text{easy}}^{101} = \sum_{i,i'=0}^{q} (x_0 \otimes x_{i'}) \otimes (y_i \otimes y_0) \otimes (z_i \otimes z_{i'}) \longrightarrow \text{label } 011011$$

tensor over $(U_0 \otimes U_1) \otimes (V_1 \otimes V_0) \otimes (W_1 \otimes W_1)$

Idea behind the proof

$$T^{011}_{\text{easy}} \otimes T^{011}_{\text{easy}} = \sum_{i,i'=0}^{q} (x_0 \otimes x_0) \otimes (y_i \otimes y_{i'}) \otimes (z_i \otimes z_{i'})$$

tensor over $(U_0 \otimes U_0) \otimes (V_1 \otimes V_1) \otimes (W_1 \otimes W_1)$

remove this term

(e.g., by setting all variables in $V_1 \otimes V_1$ to zero)

note: this removes more than one term!

SHARE VARIABLES

Consider $N = 2$

$$T^{\otimes 2}_{\text{easy}} = (T^{011}_{\text{easy}} + T^{101}_{\text{easy}} + T^{110}_{\text{easy}}) \otimes (T^{011}_{\text{easy}} + T^{101}_{\text{easy}} + T^{110}_{\text{easy}})$$

$$= \boxed{\cancel{T^{011}_{\text{easy}} \otimes T^{011}_{\text{easy}}}} + \boxed{T^{011}_{\text{easy}} \otimes T^{101}_{\text{easy}}} + \cdots + \cancel{T^{110}_{\text{easy}} \otimes T^{110}_{\text{easy}}}$$ (9 terms)

~~001111~~ 011011 ~~111100~~

~~011110~~ ~~101101~~ ~~100111~~ ~~111001~~ ~~110011~~

$\boxed{110110}$

$$T^{011}_{\text{easy}} \otimes T^{101}_{\text{easy}} = \sum_{i,i'=0}^{q} (x_0 \otimes x_{i'}) \otimes (y_i \otimes y_0) \otimes (z_i \otimes z_{i'})$$

by setting all variables in $U_1 \otimes U_0$, $V_0 \otimes V_0$ and $W_0 \otimes W_1$ to zero

label 011011

tensor over $(U_0 \otimes U_1) \otimes (V_1 \otimes V_0) \otimes (W_1 \otimes W_1)$

Idea behind the proof

Conclusion: we can convert $T_{\text{easy}}^{\otimes 2}$ (a sum of 9 terms) into a direct sum of 2 terms

NEXT STEP

Consider $T_{\text{easy}}^{\otimes N} = T_{\text{easy}}^{011} \otimes \cdots \otimes T_{\text{easy}}^{011} + \cdots + T_{\text{easy}}^{110} \otimes \cdots \otimes T_{\text{easy}}^{110}$ (3^N terms)

labels: $\underbrace{0 \cdots 01 \cdots 11 \cdots 1}_{3N}$ $\qquad$ $\underbrace{1 \cdots 11 \cdots 10 \cdots 0}_{3N}$

Idea behind the proof

Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^{\otimes N}$ can be converted into a direct sum of

$$\exp\left(\left(H\left(\frac{1}{3},\frac{2}{3}\right)-o(1)\right)N\right) = \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N}$$

terms, each containing $\boxed{\frac{N}{3} \text{ copies of } T_{\text{easy}}^{011},\ \frac{N}{3} \text{ copies of } T_{\text{easy}}^{101} \text{ and } \frac{N}{3} \text{ copies of } T_{\text{easy}}^{110}.}$

We can obtain $\left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N}$ labels of the form

$$\overbrace{\underbrace{0\cdots 1}_{\substack{\#0 = N/3\\ \#1 = 2N/3}}\ \underbrace{1\cdots 0}_{\substack{\#0 = N/3\\ \#1 = 2N/3}}\ \underbrace{0\cdots 1}_{\substack{\#0 = N/3\\ \#1 = 2N/3}}}^{3N}$$

number of possibilities

$$\binom{N}{\frac{N}{3},\frac{2N}{3}} \approx \exp\left(H\left(\frac{1}{3},\frac{2}{3}\right)N\right)$$

that do not share a blue part, a red part or a green part

The proof of this theorem is based on a complicated construction using the existence of dense sets of integers with no three-term arithmetic progression and Reinterpretation

General Formulation of the Laser Method

The laser method: general formulation

For any tensor T, any $N \geq 1$ and any $\rho \in [2,3]$ define $V_{\rho,N}(T)$ as the maximum of $\sum_{i=1}^{k}(m_i n_i p_i)^{\rho/3}$ over all restrictions of $T^{\otimes N}$ isomorphic to $\bigoplus_{i=1}^{k}\langle m_i, n_i, p_i\rangle$

$$V_\rho(T) = \lim_{N\to\infty} V_{\rho,N}(T)^{1/N}$$

The value of T

This is the definition for symmetric tensors. Otherwise we use $V_\rho(T) = V_\rho(T \otimes \pi T \otimes \pi^2 T)^{1/3}$

$$V_\rho(\langle m,n,p\rangle) = (mnp)^{\rho/3}$$

This is an increasing function of ρ

$$V_\rho(T \oplus T') \geq V_\rho(T) + V_\rho(T') \qquad V_\rho(T \otimes T') \geq V_\rho(T) \times V_\rho(T')$$

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$\sum_{i=1}^{k}(m_i n_i p_i)^{\omega/3} \leq \underline{R}\left(\bigoplus_{i=1}^{k}\langle m_i, n_i, p_i\rangle\right)$$

Example: The first CW construction

For any tensor T, any $N \geq 1$ and any $\rho \in [2,3]$ define $V_{\rho,N}(T)$ as the maximum of $\sum_{i=1}^{k}(m_i n_i p_i)^{\rho/3}$ over all restrictions of $T^{\otimes N}$ isomorphic to $\bigoplus_{i=1}^{k}\langle m_i, n_i, p_i\rangle$

$$V_\rho(T) = \lim_{N\to\infty} V_{\rho,N}(T)^{1/N}$$

The value of T

Theorem [Coppermith and Winograd 87]

The tensor $T_{\text{easy}}^{\otimes N}$ can be converted into a direct sum of

$$\exp\left(\left(H\left(\frac{1}{3},\frac{2}{3}\right) - o(1)\right)N\right) = \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N}$$

terms, each containing $\frac{N}{3}$ copies of T_{easy}^{011}, $\frac{N}{3}$ copies of T_{easy}^{101} and $\frac{N}{3}$ copies of T_{easy}^{110}.

isomorphic to $[T_{\text{easy}}^{011}]^{\otimes N/3} \otimes [T_{\text{easy}}^{101}]^{\otimes N/3} \otimes [T_{\text{easy}}^{110}]^{\otimes N/3} \cong \langle q^{N/3}, q^{N/3}, q^{N/3}\rangle$

$$V_{\rho,N}(T_{\text{easy}}) \geq \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N} \times q^{\rho N/3} \longrightarrow V_\rho(T_{\text{easy}}) \geq \frac{3}{2^{2/3}} \times q^{\rho/3}$$

The laser method: general formulation

For any tensor T, any $N \geq 1$ and any $\rho \in [2,3]$ define $V_{\rho,N}(T)$ as the maximum of $\sum_{i=1}^{k}(m_i n_i p_i)^{\rho/3}$ over all restrictions of $T^{\otimes N}$ isomorphic to $\bigoplus_{i=1}^{k}\langle m_i, n_i, p_i\rangle$

$$V_\rho(T) = \lim_{N\to\infty} V_{\rho,N}(T)^{1/N}$$

The value of T

for instance, $V_\omega(\langle m,n,p\rangle) = (mnp)^\rho$

Theorem (the asymptotic sum inequality, general form) [Schönhage 1981]

$$\sum_{i=1}^{k}(m_i n_i p_i)^{\omega/3} \leq \underline{R}\left(\bigoplus_{i=1}^{k}\langle m_i, n_i, p_i\rangle\right)$$

Theorem (simple generalization of the asymptotic sum inequality)

$$V_\omega(T) \leq \underline{R}(T)$$

The laser method: general formulation

Consider three vector spaces U, V and W over $\mathbb{F}$

A tensor T over (U, V, W) is an element of $U \otimes V \otimes W$

Assume that U, V and W are decomposed as

$$U = \bigoplus_{i \in I} U_i \quad V = \bigoplus_{j \in J} V_j \quad W = \bigoplus_{k \in K} W_k \quad \text{for some } I, J, K \subseteq \mathbb{Z}$$

The tensor T is a partitioned tensor (with respect to this decomposition) if it can be written as $T = \sum_{(i,j,k) \in I \times J \times K} T_{ijk}$

where $T_{ijk} \in U_i \otimes V_j \otimes W_k$ for each $(i, j, k) \in I \times J \times K$

support of the tensor: $\mathrm{supp}(T) = \{(i, j, k) \in I \times J \times K \mid T_{ijk} \neq 0\}$

each non-zero T_{ijk} is called a component of T

We say that the tensor is <u>tight</u> if there exists some integer d such that

$i + j + k = d$ for all $(i, j, k) \in \mathrm{supp}(T)$

Example: The first CW construction

$$U = U_0 \oplus U_1, \quad \text{where } U_0 = span\{x_0\} \text{ and } U_1 = span\{x_1, \ldots, x_q\} \qquad I = \{0,1\}$$

$$V = V_0 \oplus V_1, \quad \text{where } V_0 = span\{y_0\} \text{ and } V_1 = span\{y_1, \ldots, y_q\} \qquad J = \{0,1\}$$

$$W = W_0 \oplus W_1, \quad \text{where } W_0 = span\{z_0\} \text{ and } W_1 = span\{z_1, \ldots, z_q\} \qquad K = \{0,1\}$$

$$T_{\mathsf{easy}} = T_{\mathsf{easy}}^{011} + T_{\mathsf{easy}}^{101} + T_{\mathsf{easy}}^{110},$$

where

$$T_{\mathsf{easy}}^{011} = \sum_{i=1}^{q} x_0 \otimes y_i \otimes z_i$$ tensor over (U_0, V_1, W_1)

$$T_{\mathsf{easy}}^{101} = \sum_{i=1}^{q} x_i \otimes y_0 \otimes z_i$$ tensor over (U_1, V_0, W_1)

$$T_{\mathsf{easy}}^{110} = \sum_{i=1}^{q} x_i \otimes y_i \otimes z_0$$ tensor over (U_1, V_1, W_0)

$$\operatorname{supp}(T_{\mathsf{easy}}) = \{(0,1,1), (1,0,1), (1,1,0)\}$$

it is tight, since $i + j + k = 2$ for all $(i,j,k) \in \operatorname{supp}(T_{\mathsf{easy}})$

The laser method: general formulation

Main Theorem [LG 14] (reinterpretation of prior works)

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in \mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk})) - \Gamma(P).$$

H: entropy
P_ℓ: projection of P along the ℓ-th coordinate ($=$ marginal distribution)
$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

Conclusion: we can compute a lower bound on the value of T if we know a lower bound on the value of each component

we can then obtain an upper bound on ω via $\boxed{V_\omega(T) \leq \underline{R}(T)}$

concretely, we use $\boxed{V_\rho(T) \geq \underline{R}(T) \Longrightarrow \omega \leq \rho}$ and do a binary search on ρ

Example: The first CW construction

Main Theorem [LG 14] (reinterpretation of prior works)

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in \mathrm{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

H: entropy
P_ℓ: projection of P along the ℓ-th coordinate ($=$ marginal distribution)
$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

$$\mathrm{supp}(T_{\text{easy}}) = \{(0,1,1),(1,0,1),(1,1,0)\}$$

$$P(0,1,1) = P(1,0,1) = P(1,1,0) = 1/3$$

$$\Gamma(P) = 0$$

$P_1(0) = 1/3,\ P_1(1) = 2/3$ and $P_2 = P_3 = P_1$

$$V_\rho(T^{011}_{\text{easy}}) = V_\rho(\langle 1,1,q\rangle) = q^{\rho/3}$$
$$V_\rho(T^{101}_{\text{easy}}) = V_\rho(\langle q,1,1\rangle) = q^{\rho/3}$$
$$V_\rho(T^{110}_{\text{easy}}) = V_\rho(\langle 1,q,1\rangle) = q^{\rho/3}$$

$$\log(V_\rho(T_{\text{easy}})) \geq H\left(\frac{1}{3},\frac{2}{3}\right) + \frac{1}{3}\log\left(q^{\rho/3}\right) + \frac{1}{3}\log\left(q^{\rho/3}\right) + \frac{1}{3}\log\left(q^{\rho/3}\right)$$

Theorem [Coppersmith and Winograd 87]

The tensor $T_{\text{easy}}^{\otimes N}$ can be converted into a direct sum of

$$\exp\left(\left(H\left(\frac{1}{3},\frac{2}{3}\right)-o(1)\right)N\right) = \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N}$$

terms, each containing $\frac{N}{3}$ copies of T_{easy}^{011}, $\frac{N}{3}$ copies of T_{easy}^{101} and $\frac{N}{3}$ copies of T_{easy}^{110}.

$$V_{\rho,N}(T_{\text{easy}}) \geq \left(\frac{3}{2^{2/3}}\right)^{(1-o(1))N} \times q^{\rho N/3} \longrightarrow V_\rho(T_{\text{easy}}) \geq \frac{3}{2^{2/3}} \times q^{\rho/3}$$

$$\text{supp}(T_{\text{easy}}) = \{(0,1,1),(1,0,1),(1,1,0)\}$$

$$P(0,1,1) = P(1,0,1) = P(1,1,0) = 1/3$$

$$\Gamma(P) = 0$$

$P_1(0) = 1/3,\ P_1(1) = 2/3$ and $P_2 = P_3 = P_1$

$$V_\rho(T_{\text{easy}}^{011}) = V_\rho(\langle 1,1,q\rangle) = q^{\rho/3}$$
$$V_\rho(T_{\text{easy}}^{101}) = V_\rho(\langle q,1,1\rangle) = q^{\rho/3}$$
$$V_\rho(T_{\text{easy}}^{110}) = V_\rho(\langle 1,q,1\rangle) = q^{\rho/3}$$

$$\log(V_\rho(T_{\text{easy}})) \geq H\left(\frac{1}{3},\frac{2}{3}\right) + \frac{1}{3}\log\left(q^{\rho/3}\right) + \frac{1}{3}\log\left(q^{\rho/3}\right) + \frac{1}{3}\log\left(q^{\rho/3}\right)$$

The laser method: general formulation

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk})) - \Gamma(P).$$

Interpretation: the laser method enables us to convert (by zeroing variables)

$T^{\otimes N}$ into a direct sum of $\exp\left(\left(\sum_{\ell=1}^{3}\frac{H(P_\ell)}{3} - \Gamma(P) - o(1)\right)N\right)$

terms, each isomorphic to $\bigotimes_{(i,j,k)\in\mathrm{supp}(T)} [T^{ijk}]^{\otimes P(i,j,k)N}$

The second CW construction

Let q be a positive integer.

Consider three vector spaces U, V and W of dimension $q+2$ over $\mathbb{F}$.

$U = span\{x_0, \ldots, x_q, x_{q+1}\}$ $\quad$ $W = span\{z_0, \ldots, z_q, z_{q+1}\}$

$V = span\{y_0, \ldots, y_q, y_{q+1}\}$

Coppersmith and Winograd (1987) considered the following tensor:

$$T_{\text{CW}} = T_{\text{easy}} + \boxed{T_{\text{CW}}^{002} + T_{\text{CW}}^{020} + T_{\text{CW}}^{200}} \qquad \underline{R}(T_{\text{CW}}) = q+2$$

$$T_{\text{CW}} = T_{\text{CW}}^{011} + T_{\text{CW}}^{101} + T_{\text{CW}}^{110} + \boxed{T_{\text{CW}}^{002} + T_{\text{CW}}^{020} + T_{\text{CW}}^{200},}$$

where
$$T_{\text{CW}}^{011} = T_{\text{easy}}^{011}$$
$$T_{\text{CW}}^{101} = T_{\text{easy}}^{101}$$
$$T_{\text{CW}}^{110} = T_{\text{easy}}^{110}$$
and
$$T_{\text{CW}}^{002} = x_0 \otimes y_0 \otimes z_{q+1} \cong \langle 1,1,1\rangle$$
$$T_{\text{CW}}^{020} = x_0 \otimes y_{q+1} \otimes z_0 \cong \langle 1,1,1\rangle$$
$$T_{\text{CW}}^{200} = x_{q+1} \otimes y_0 \otimes z_0 \cong \langle 1,1,1\rangle.$$

The second CW construction

$U = span\{x_0, \ldots, x_q, x_{q+1}\}$ $\quad$ $W = span\{z_0, \ldots, z_q, z_{q+1}\}$

$V = span\{y_0, \ldots, y_q, y_{q+1}\}$

$U = U_0 \oplus U_1 \oplus U_2,$ where $U_0 = span\{x_0\}$, $U_1 = span\{x_1, \ldots, x_q\}$ and $U_2 = span\{x_{q+1}\}$

$V = V_0 \oplus V_1 \oplus V_2,$ where $V_0 = span\{y_0\}$, $V_1 = span\{y_1, \ldots, y_q\}$ and $V_2 = span\{y_{q+1}\}$

$W = W_0 \oplus W_1 \oplus W_2,$ where $W_0 = span\{z_0\}$, $W_1 = span\{z_1, \ldots, z_q\}$ and $W_2 = span\{z_{q+1}\}$

$$T_{\mathsf{CW}} = T^{011}_{\mathsf{CW}} + T^{101}_{\mathsf{CW}} + T^{110}_{\mathsf{CW}} + T^{002}_{\mathsf{CW}} + T^{020}_{\mathsf{CW}} + T^{200}_{\mathsf{CW}}$$

This is not a direct sum

T^{011}_{CW} tensor over (U_0, V_1, W_1)

T^{101}_{CW} tensor over (U_1, V_0, W_1)

T^{110}_{CW} tensor over (U_1, V_1, W_0)

$T^{002}_{\mathsf{CW}} = x_0 \otimes y_0 \otimes z_{q+1} \cong \langle 1, 1, 1 \rangle$ tensor over (U_0, V_0, W_2)

$T^{020}_{\mathsf{CW}} = x_0 \otimes y_{q+1} \otimes z_0 \cong \langle 1, 1, 1 \rangle$ tensor over (U_0, V_2, W_0)

$T^{200}_{\mathsf{CW}} = x_{q+1} \otimes y_0 \otimes z_0 \cong \langle 1, 1, 1 \rangle$ tensor over (U_2, V_0, W_0)

The second CW construction: laser method

$$\mathrm{supp}(T_{\mathsf{CW}}) = \{(0,1,1),(1,0,1),(1,1,0),(0,0,2),(0,2,0),(2,0,0)\}$$

$$V_\rho(T_{\mathsf{CW}}^{002}) = V_\rho(T_{\mathsf{CW}}^{020}) = V_\rho(T_{\mathsf{CW}}^{200}) = 1 \qquad V_\rho(T_{\mathsf{CW}}^{011}) = V_\rho(T_{\mathsf{CW}}^{101}) = V_\rho(T_{\mathsf{CW}}^{110}) = q^{\rho/3}$$

take $P(0,1,1) = P(1,0,1) = P(1,1,0) = \alpha$

$P(0,0,2) = P(0,2,0) = P(2,0,0) = (1/3-\alpha)$ with $0 \le \alpha \le 1/3$

$P_1(0) = \alpha + 2(1/3-\alpha),\ P_1(1) = 2\alpha,\ P_1(2) = (1/3-\alpha)$

$$T_{\mathsf{CW}} = T_{\mathsf{CW}}^{011} + T_{\mathsf{CW}}^{101} + T_{\mathsf{CW}}^{110} + T_{\mathsf{CW}}^{002} + T_{\mathsf{CW}}^{020} + T_{\mathsf{CW}}^{200}$$

T_{CW}^{011} tensor over (U_0, V_1, W_1)

T_{CW}^{101} tensor over (U_1, V_0, W_1)

T_{CW}^{110} tensor over (U_1, V_1, W_0)

$T_{\mathsf{CW}}^{002} = x_0 \otimes y_0 \otimes z_{q+1} \cong \langle 1,1,1\rangle$ tensor over (U_0, V_0, W_2)

$T_{\mathsf{CW}}^{020} = x_0 \otimes y_{q+1} \otimes z_0 \cong \langle 1,1,1\rangle$ tensor over (U_0, V_2, W_0)

$T_{\mathsf{CW}}^{200} = x_{q+1} \otimes y_0 \otimes z_0 \cong \langle 1,1,1\rangle$ tensor over (U_2, V_0, W_0)

The second CW construction: laser method

$$\mathrm{supp}(T_{\mathsf{CW}}) = \{(0,1,1),(1,0,1),(1,1,0),(0,0,2),(0,2,0),(2,0,0)\}$$

$$V_\rho(T^{002}_{\mathsf{CW}}) = V_\rho(T^{020}_{\mathsf{CW}}) = V_\rho(T^{200}_{\mathsf{CW}}) = 1 \qquad V_\rho(T^{011}_{\mathsf{CW}}) = V_\rho(T^{101}_{\mathsf{CW}}) = V_\rho(T^{110}_{\mathsf{CW}}) = q^{\rho/3}$$

take
$$P(0,1,1) = P(1,0,1) = P(1,1,0) = \alpha$$
$$P(0,0,2) = P(0,2,0) = P(2,0,0) = (1/3-\alpha)$$
with $0 \le \alpha \le 1/3$
$$P_1(0) = \alpha + 2(1/3-\alpha),\ P_1(1) = 2\alpha,\ P_1(2) = (1/3-\alpha)$$

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \ge \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk})) - \Gamma(P).$$

$$\Longrightarrow \log(V_\rho(T_{\mathsf{CW}})) \ge H\left(\frac{2}{3}-\alpha, 2\alpha, \frac{1}{3}-\alpha\right) + \log(q^{\alpha\omega})$$

combined with $V_\omega(T_{\mathsf{CW}}) \le \underline{R}(T_{\mathsf{CW}}) = q+2$

this gives $\boxed{\omega \le 2.38718\ldots}$ for $q = 6$ and $\alpha = 0.3173$

Analysis of the second construction

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

Analysis of the second power

$$T_{\mathsf{CW}}^{\otimes 2} = (T_{\mathsf{CW}}^{011} + T_{\mathsf{CW}}^{101} + T_{\mathsf{CW}}^{110} + T_{\mathsf{CW}}^{002} + T_{\mathsf{CW}}^{020} + T_{\mathsf{CW}}^{200})^{\otimes 2} \quad \text{(36 terms)}$$

$\underline{R}(T_{\mathsf{CW}}^{\otimes 2}) \leq (q+2)^2$

Idea: rewrite it as a (non-direct) sum of 15 terms by regrouping terms

$$\begin{aligned} T_{\mathsf{CW}}^{\otimes 2} = {} & T^{400} + T^{040} + T^{004} + T^{310} + T^{301} + T^{103} + T^{130} + T^{013} \\ & + T^{031} + T^{220} + T^{202} + T^{022} + T^{211} + T^{121} + T^{112}, \end{aligned}$$

where

$$\begin{aligned} T^{400} &= T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{200}, \\ T^{310} &= T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{110} + T_{\mathsf{CW}}^{110} \otimes T_{\mathsf{CW}}^{200}, \\ T^{220} &= T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{020} + T_{\mathsf{CW}}^{020} \otimes T_{\mathsf{CW}}^{200} + T_{\mathsf{CW}}^{110} \otimes T_{\mathsf{CW}}^{110}, \\ T^{211} &= T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{011} + T_{\mathsf{CW}}^{011} \otimes T_{\mathsf{CW}}^{200} + T_{\mathsf{CW}}^{110} \otimes T_{\mathsf{CW}}^{101} + T_{\mathsf{CW}}^{101} \otimes T_{\mathsf{CW}}^{110}, \end{aligned}$$

MERGING

and the other 11 terms are obtained by permuting the variables (e.g., $T^{040} = T_{\mathsf{CW}}^{020} \otimes T_{\mathsf{CW}}^{020}$).

Analysis of the second power

$\mathrm{supp}(T_{\mathsf{CW}}^{\otimes 2}) = \{\underbrace{(4,0,0),\ldots,(0,0,4)}_{\text{3 permutations}},\underbrace{(3,1,0),\ldots,(0,1,3)}_{\text{6 permutations}},\underbrace{(2,2,0),\ldots,(0,2,2)}_{\text{3 permutations}},\underbrace{(2,1,1),\ldots,(1,1,2)}_{\text{3 permutations}}\}$

lower bounds on the values of each component can be computed (recursively)

choice of distribution: (4-1=3 parameters)

$$P(4,0,0) = \ldots = P(0,0,4) = \alpha, \quad P(3,1,0) = \ldots = P(0,1,3) = \beta$$
$$P(2,2,0) = \ldots = P(0,2,2) = \gamma, \quad P(2,1,1) = \ldots = P(1,1,2) = \delta$$

$$T_{\mathsf{CW}}^{\otimes 2} = T^{400} + T^{040} + T^{004} + T^{310} + T^{301} + T^{103} + T^{130} + T^{013} + T^{031} + T^{220} + T^{202} + T^{022} + T^{211} + T^{121} + T^{112},$$

where

$$T^{400} = T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{200},$$
$$T^{310} = T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{110} + T_{\mathsf{CW}}^{110} \otimes T_{\mathsf{CW}}^{200},$$
$$T^{220} = T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{020} + T_{\mathsf{CW}}^{020} \otimes T_{\mathsf{CW}}^{200} + T_{\mathsf{CW}}^{110} \otimes T_{\mathsf{CW}}^{110},$$
$$T^{211} = T_{\mathsf{CW}}^{200} \otimes T_{\mathsf{CW}}^{011} + T_{\mathsf{CW}}^{011} \otimes T_{\mathsf{CW}}^{200} + T_{\mathsf{CW}}^{110} \otimes T_{\mathsf{CW}}^{101} + T_{\mathsf{CW}}^{101} \otimes T_{\mathsf{CW}}^{110},$$

and the other 11 terms are obtained by permuting the variables (e.g., $T^{040} = T_{\mathsf{CW}}^{020} \otimes T_{\mathsf{CW}}^{020}$).

Analysis of the second power

$$\mathrm{supp}(T_{\mathsf{CW}}^{\otimes 2}) = \{\underbrace{(4,0,0),\ldots,(0,0,4)}_{\text{3 permutations}}, \underbrace{(3,1,0),\ldots,(0,1,3)}_{\text{6 permutations}}, \underbrace{(2,2,0),\ldots,(0,2,2)}_{\text{3 permutations}}, \underbrace{(2,1,1),\ldots,(1,1,2)}_{\text{3 permutations}}\}$$

lower bounds on the values of each component can be computed (recursively)

choice of distribution:
(4-1=3 parameters)

$$P(4,0,0) = \ldots = P(0,0,4) = \alpha, \quad P(3,1,0) = \ldots = P(0,1,3) = \beta$$
$$P(2,2,0) = \ldots = P(0,2,2) = \gamma, \quad P(2,1,1) = \ldots = P(1,1,2) = \delta$$

we have $\Gamma(P) = 0$

Analysis of the second power

$$\mathrm{supp}(T_{\mathsf{CW}}^{\otimes 2}) = \{\underbrace{(4,0,0),\ldots,(0,0,4)}_{\text{3 permutations}},\underbrace{(3,1,0),\ldots,(0,1,3)}_{\text{6 permutations}},\underbrace{(2,2,0),\ldots,(0,2,2)}_{\text{3 permutations}},\underbrace{(2,1,1),\ldots,(1,1,2)}_{\text{3 permutations}}\}$$

lower bounds on the values of each component can be computed (recursively)

choice of distribution: (4-1=3 parameters)

$$P(4,0,0) = \ldots = P(0,0,4) = \alpha, \quad P(3,1,0) = \ldots = P(0,1,3) = \beta$$
$$P(2,2,0) = \ldots = P(0,2,2) = \gamma, \quad P(2,1,1) = \ldots = P(1,1,2) = \delta$$

we have $\Gamma(P) = 0$

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk})) - \Gamma(P).$$

Theorem

$$V_\omega(T) \leq \underline{R}(T)$$

$\Longrightarrow \boxed{\omega \leq 2.3755\ldots}$ for $q = 6$ and $\alpha = 0.00023$, $\beta = 0.0125$, $\gamma = 0.10254$ and $\delta = 0.2056$

Analysis of the second power

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

What about the third power (using similar merging schemes)?

$\longrightarrow$ this does not give any improvement

Analysis of the fourth power

$$T_{\mathsf{CW}}^{\otimes 4} = (T_{\mathsf{CW}}^{011} + T_{\mathsf{CW}}^{101} + T_{\mathsf{CW}}^{110} + T_{\mathsf{CW}}^{002} + T_{\mathsf{CW}}^{020} + T_{\mathsf{CW}}^{200})^{\otimes 4} \quad (6^4 \text{ terms})$$

$$\underline{R}(T_{\mathsf{CW}}^{\otimes 4}) \leq (q+2)^4$$

Idea: rewrite it as a (non-direct) sum of a smaller number of terms by regrouping terms

$$T_{\mathsf{CW}}^{\otimes 4} = T^{800} + T^{710} + T^{620} + T^{611} + T^{530} + T^{521} + T^{440} + T^{431} + T^{422} + T^{332} + \text{permutations of these terms}$$

$$T^{080}, T^{008}, T^{701}, T^{107}, T^{170}, T^{017}, T^{071}, \ldots$$

10-1=9 parameters for the probability distribution

this time $\Gamma(P) \neq 0$

The laser method: general formulation

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in \mathrm{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

H: entropy
P_ℓ: projection of P along the ℓ-th coordinate (= marginal distribution)
$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

The laser method: general formulation

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in \mathrm{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk})) - \Gamma(P).$$

H: entropy

P_ℓ: projection of P along the ℓ-th coordinate ($=$ marginal distribution)

$\Gamma(P)$: to be defined later (zero in the case of simple tensors)

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

when the structure of support is simple, we typically have

$P_1 = Q_1,\ P_2 = Q_2,\ P_3 = Q_3 \Longrightarrow P = Q$ and thus $\Gamma(P) = 0$

The laser method: general formulation

Interpretation: the laser method enables us to convert (by zeroing variables)

$T^{\otimes N}$ into a direct sum of $\exp\left(\left(\sum_{\ell=1}^{3}\frac{H(P_\ell)}{3}-\Gamma(P)-o(1)\right)N\right)$

terms, each $\boxed{\text{isomorphic to } \bigotimes_{(i,j,k)\in\mathrm{supp}(T)}[T^{ijk}]^{\otimes P(i,j,k)N}}$ "type P"

we can control only the choice of the marginal distributions P_1, P_2 and P_3

what we obtain is a (non-direct) sum of all "type Q" terms

the most frequent terms are those with Q maximizing H(Q)

the fact that "type P" are not the most frequent introduces the penalty term $-\Gamma(P)$

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

The laser method: computing the bound

Main Theorem [LG 14]

For any tight partitioned tensor T, any $\boxed{\text{probability distribution}}$ P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \underbrace{\boxed{\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3}}}_{\text{concave}} + \underbrace{\boxed{\sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk}))}}_{\text{linear}} - \underbrace{\boxed{\Gamma(P)}}_{=0}.$$

How to find the best distribution for a given ρ?

assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known

If $\Gamma(P)$ = 0 for all distributions P, the best distribution can be done efficiently (numerically) using convex optimization

maximization of a $\boxed{\text{concave function}}$ under $\boxed{\text{linear constraints}}$

The laser method: computing the bound

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \underbrace{\boxed{\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3}}}_{\text{concave}} + \underbrace{\boxed{\sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk}))}}_{\text{linear}} - \underbrace{\boxed{\Gamma(P)}}_{???}.$$

How to find the best distribution for a given ρ?

assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]

The laser method: computing the bound

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$
such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]

The laser method: computing the bound

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \underbrace{\boxed{\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3}}}_{\text{concave}} + \underbrace{\boxed{\sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk}))}}_{\text{linear}} - \underbrace{\boxed{\Gamma(P)}}_{???}.$$

How to find the best distribution for a given ρ?

assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]

Simplification: restrict the search to the set of distributions P such that $\Gamma(P) = 0$

still hard to solve, but can be done up to the 8th power of the CW tensor [Vassilevska-Williams 12]

The laser method: computing the bound

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

Simplification: restrict the search to the set of distributions P such that $\Gamma(P) = 0$

still hard to solve, but can be done up to the 8th power of the CW tensor [Vassilevska-Williams 12]

The laser method: computing the bound

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

$$\log(V_\rho(T)) \geq \underbrace{\boxed{\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3}}}_{\text{concave}} + \underbrace{\boxed{\sum_{(i,j,k)\in \mathrm{supp}(T)} P(i,j,k) \log(V_\rho(T_{ijk}))}}_{\text{linear}} - \underbrace{\boxed{\Gamma(P)}}_{???}.$$

How to find the best distribution for a given ρ?

assume that (a lower bound on) each $V_\rho(T_{ijk})$ is known

In general:

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

hard to solve, but can be done up to the 4th power of the CW tensor [Stothers 10]

Simplification: restrict the search to the set of distributions P such that $\Gamma(P) = 0$

still hard to solve, but can be done up to the 8th power of the CW tensor [Vassilevska-Williams 12]

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2, 3]$, we have

call this expression $f(P)$

$$\log(V_\rho(T)) \geq \boxed{\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk}))} - \Gamma(P).$$

Efficient method to find a solution [LG 14] (close to the optimal solution):

$$\Gamma(P) = \max[H(Q)] - H(P)$$

where the max is over all distributions Q over $\mathrm{supp}(T)$ such that $P_1 = Q_1$, $P_2 = Q_2$ and $P_3 = Q_3$

Main Theorem [LG 14]

For any tight partitioned tensor T, any probability distribution P over $\mathrm{supp}(T)$, and any $\rho \in [2,3]$, we have

call this expression $f(P)$

$$\log(V_\rho(T)) \geq \boxed{\sum_{\ell=1}^{3} \frac{H(P_\ell)}{3} + \sum_{(i,j,k)\in\mathrm{supp}(T)} P(i,j,k)\log(V_\rho(T_{ijk}))} - \Gamma(P).$$

Efficient method to find a solution [LG 14] (close to the optimal solution):

1. find a distribution P that maximizes $f(P)$, and call it $\hat{P}$
   concave objective function, linear constraints
2. find the distribution Q that maximizes $H(Q)$ under the constraints $Q_1 = \hat{P}_1$, $Q_2 = \hat{P}_2$ and $Q_3 = \hat{P}_3$. Call it $\hat{Q}$.
   concave objective function, linear constraints
3. output $f(\hat{Q})$

Since $\Gamma(\hat{Q}) = 0$, we have $\log(V_\rho(T)) \geq f(\hat{Q})$ from the theorem

Analysis of power 16 and 32

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

solutions to the optimization problems obtained numerically by convex optimization

Conclusion

We constructed a time-efficient implementation of the laser method

any tight partitioned tensor for which (lower bounds on) the value of each component is known

Laser-method-based analysis (v2.3)

convex optimization polynomial time

upper bound on ω

We applied it to study higher powers of the basic tensor by CW

analysis of the m-th power of the tensor by CW

m	Upper bound	Number of variables in in the optimization problem	Authors
1	$\omega < 2.3871900$	1	CW (1987)
2	$\omega < 2.3754770$	3	CW (1987)
4	$\omega < 2.3729269$	9	Stothers (2010)
8	$\omega < 2.3729$	29	Vassilevska Williams (2012)
16	$\omega < 2.3728640$	101	Le Gall (2014)
32	$\omega < 2.3728639$	373	Le Gall (2014)

recent result [Ambainis, Filmus, LG 14]:

studying higher powers (using the same approach) cannot give an upper bound better than 2.3725