Codes and Expansions in Algorithms for Matrix Multiplication

Codes & Expansion (CodEx) Seminar
February, 2021

Joshua A. Grochow

University of Colorado **Boulder**

Main Message

Algorithms for Matrix Multiplication

Are

Code-Like Objects

Multiplying matrices

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

n multiplications and $n-1$ additions per dot product

n^2 dot products

$\Rightarrow O(n^3)$ steps

Multiplying matrices

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

n multiplications and $n-1$ additions per dot product
n^2 dot products
$\Rightarrow O(n^3)$ steps

Theorem [Klyuyev & Kokovkin-Scherbak '65]: Optimal if only allowed to work on rows and columns as a whole.

Theorem [Strassen '69]: Can do better! $O\left(n^{\log_2 7}\right) = O(n^{2.81})$.

Strassen's Algorithm I: A magical 2x2 trick

Ordinary 2x2 product: 8 products, 4 sums

$$I = (a_{11}+a_{22})(b_{11} + b_{22}) \qquad II = (a_{21} + a_{22})(b_{11})$$

$$III = (a_{11})(b_{12} - b_{22}) \qquad IV = (a_{22})(-b_{11} + b_{21})$$

$$VI = (-a_{11} + a_{21})(b_{11} + b_{12}) \qquad V = (a_{11} + a_{12})(b_{22})$$

$$VII = (a_{12} - a_{22})(b_{21} + b_{22})$$

$$\begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} I \\ II \\ \vdots \\ VII \end{pmatrix}$$

Strassen 2x2 product: 7 products, 18 sums

Strassen's Algorithm II: Recurse

Works correctly even if entries are from a noncommutative ring

Suppose the entries are from $M_n(\mathbb{C})$

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

Strassen's Algorithm II: Recurse

Works correctly even if entries are from a noncommutative ring

Suppose the entries are from $M_n(\mathbb{C})$

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

$$\begin{pmatrix} \begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix} & \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nn} \end{pmatrix} \\ \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{pmatrix} & \begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \delta_{n2} & \cdots & \delta_{nn} \end{pmatrix} \end{pmatrix}$$

$$M_2\big(M_n(\mathbb{C})\big) \cong M_{2n}(\mathbb{C})$$

Strassen's Algorithm II: Recurse

Works correctly even if entries are from a noncommutative ring

Suppose the entries are from $M_n(\mathbb{C})$

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

$$M_2\big(M_n(\mathbb{C})\big) \cong M_{2n}(\mathbb{C})$$

Multiply $2n \times 2n$ matrices using 7 $n \times n$ products, and 18 $n \times n$ additions

Let $R(n) =$ # products ("rank"), $A(n) =$# additions to mult two $n \times n$ matrices

$$R(2n) \leq 7R(n) \qquad\qquad R(n) \leq n^{\log_2 7}$$

$$A(2n) \leq 18n^2 + 7A(n) \quad \rightarrow \quad A(n) \leq 6\left(7n^{\log_2 7} - 4n^2\right) = \Theta(n^{\log_2 7})$$

Multiplying matrices

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

Theorem [Strassen '69]: Can do better! $O\left(n^{\log_2 7}\right) = O(n^{2.81})$.

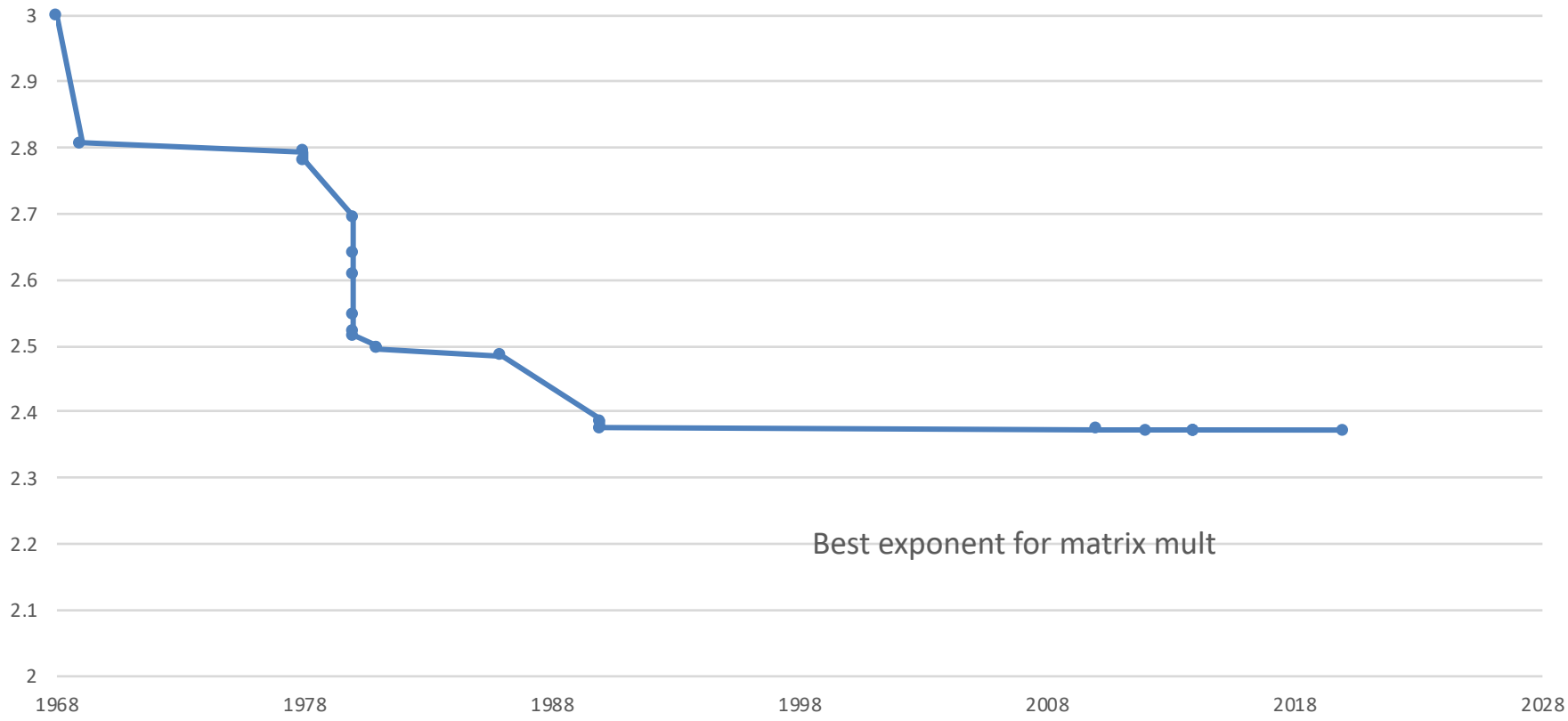
Definition: The exponent of matrix multiplication is

$$\omega = \inf\{e : MM_n \; in \; O(n^{e+\varepsilon}) \forall \varepsilon > 0\}$$

Conjecture (folklore): $\omega = 2$.

In principle, ω depends on the characteristic.

Improvements (in theory)



Current record [Alman-Williams '20]: $\omega < 2.372859$

Rephrasing Matrix Multiplication Symmetrically

It is a bilinear map $M_n \otimes M_n \to M_n$, so we can treat it as an element of $M_n \otimes M_n \otimes M_n^*$, namely

$$T_{MM} = \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

$$\langle T_{MM} | A \otimes B \otimes C \rangle = tr(ABC)$$

Note $(AB)_{ik} = tr(ABE_{ki})$

Rephrasing Matrix Multiplication Symmetrically

$$\sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

If we write $M_n \cong U \otimes V^*$, then this lives in

$$M_n \otimes M_n \otimes M_n^*$$
$$\cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*)$$
$$\cong U \otimes (V \otimes V^*) \otimes (W^* \otimes W) \otimes U^*$$

In this decomposition, MM is

$$Id_U \otimes Id_V \otimes Id_W$$

Rephrasing Matrix Multiplication Symmetrically

$$\sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

If we write $M_n \cong U \otimes V^*$, then this lives in

$$M_n \otimes M_n \otimes M_n^*$$
$$\cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*)$$
$$\cong U \otimes (V^* \otimes V) \otimes (W^* \otimes W) \otimes U^*$$

In this decomposition, MM is

$$Id_U \otimes Id_V \otimes Id_W$$

Rephrasing Matrix Multiplication Symmetrically

$$\sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

If we write $M_n \cong U \otimes V^*$, then this lives in

$$M_n \otimes M_n \otimes M_n^*$$
$$\cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*)$$
$$\cong U \otimes (V^* \otimes V) \otimes (W^* \otimes W) \otimes U^*$$

In this decomposition, MM is

$$Id_U \otimes Id_V \otimes Id_W$$

Rephrasing Matrix Multiplication Symmetrically

$$\sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

If we write $M_n \cong U \otimes V^*$, then this lives in

$$M_n \otimes M_n \otimes M_n^*$$
$$\cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*)$$
$$\cong U \otimes (V^* \otimes V) \otimes (W^* \otimes W) \otimes U^*$$

In this decomposition, MM is

$$Id_U \otimes Id_V \otimes Id_W$$

Tensor Rank

$$T_{MM} = \sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

Def: A rank-1 tensor is $A \otimes B \otimes C$.

A tensor t has rank $\leq r$ if it can be written

$$T = \sum_{i=1}^{r} A_i \otimes B_i \otimes C_i$$

(Generalizes matrix rank)

Tensor Rank

Def: A tensor t has rank $\leq r$ if it can be written

$$T = \sum_{i=1}^{r} A_i \otimes B_i \otimes C_i$$

$\mathrm{mult}(T)$:= # of non-scalar multiplications needed to compute T.

Thm (classical): $mult(T) \leq rk(T) \leq 2mult(T)$

What is an algorithm for computing matrix multiplication?

Def: A tensor t has rank $\leq r$ if it can be written

$$T = \sum_{i=1}^{r} A_i \otimes B_i \otimes C_i$$

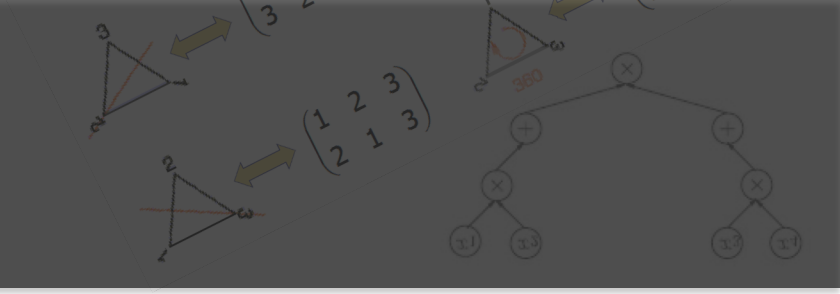
By an "algorithm for matrix multiplication" we mean

$$\{A_i \otimes B_i \otimes C_i \mid i = 1, \dots, r\}$$

s.t. $T_{MM} = \sum A_i \otimes B_i \otimes C_i$.

This is (somehow) like a code!

Strassen's Algorithm: A magical 2x2 trick



Strassen 2x2 product: **7** products, 18 sums

$$I = (a_{11}+a_{22})(b_{11} + b_{22}) \qquad II = (a_{21} + a_{22})(b_{11})$$

$$III = (a_{11})(b_{12} - b_{22}) \qquad IV = (a_{22})(-b_{11} + b_{21})$$

$$VI = (-a_{11} + a_{21})(b_{11} + b_{12}) \qquad V = (a_{11} + a_{12})(b_{22})$$

$$VII = (a_{12} - a_{22})(b_{21} + b_{22})$$

$$\begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} I \\ II \\ \vdots \\ VII \end{pmatrix}$$

Strassen: $rk(T_{MM,2}) \leq 7$

Multiplying matrices in practice

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

Used in scientific computing*, graphics, GPUs, combinatorial algorithms, other algebraic algorithms, deep learning

Matrix multiplication can even be the bottleneck

Multiplying matrices in practice

In practice

- Often have sparse or structured matrices → better algorithms available
- Even with dense, unstructured matrices, memory and communication are more frequently the bottleneck than number of operations

Multiplying matrices in practice

In practice

- Often have sparse or structured matrices → better algorithms available
- Even with dense, unstructured matrices, memory and communication are more frequently the bottleneck than number of operations
- Strassen's algorithm actually used in practice
- Other algorithms today aren't, but future ones could be!

Multiplying matrices

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

Used in scientific computing*, graphics, GPUs, combinatorial algorithms, other algebraic algorithms, deep learning

Matrix multiplication can even be the bottleneck

More importantly: gives us insight into the nature of computing & complexity!

Strassen's Algorithm II: Origin story

How did Strassen find this?

Trying to prove it couldn't be done mod 2, by hand, (clever) brute force.

Found solution mod 2.

Figured out signs to work over $\mathbb{Z}$.

Mystery:
Math behind the magic?

Where does the magical 2×2 trick

“really” come from?

Math behind the magic I
Clausen's Construction

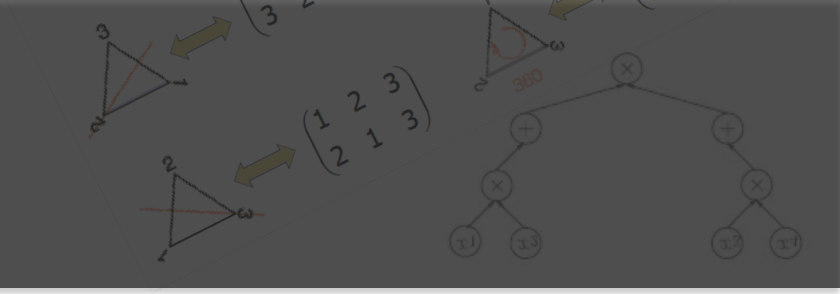
Multiplication table

	E_{11}	E_{12}	E_{21}	E_{22}
E_{11}	E_{11}	E_{12}	0	0
E_{21}	E_{21}	E_{22}	0	0
E_{12}	0	0	E_{11}	E_{12}
E_{22}	0	0	E_{21}	E_{22}

Leads to tensor rank decomposition

$$T_{MM,2} =$$
$$E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots$$
$$+E_{22} \otimes E_{22} \otimes E_{22}$$

Math behind the magic I
Clausen's Construction



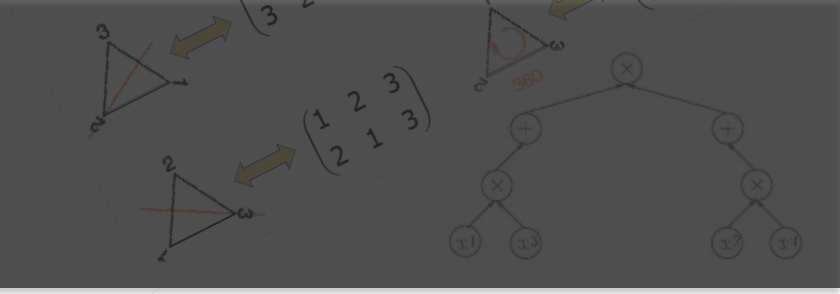
Multiplication table

	E_{11}	E_{12}	E_{21}	E_{22}
E_{11}	E_{11}	E_{12}	0	0
E_{21}	E_{21}	E_{22}	0	0
E_{12}	0	0	E_{11}	E_{12}
E_{22}	0	0	E_{21}	E_{22}

Leads to tensor rank decomposition

$$T_{MM,2} =$$
$$E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots$$
$$+E_{22} \otimes E_{22} \otimes E_{22}$$

Math behind the magic I
Clausen's Construction



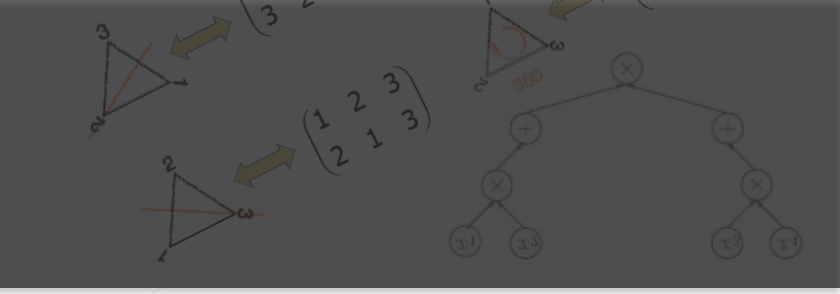
Multiplication table

	E_{11}	E_{12}	E_{21}	E_{22}
E_{11}	E_{11}	E_{12}	0	0
E_{21}	E_{21}	E_{22}	0	0
E_{12}	0	0	E_{11}	E_{12}
E_{22}	0	0	E_{21}	E_{22}

Leads to tensor rank decomposition

$$T_{MM,2} =$$
$$E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots$$
$$+E_{22} \otimes E_{22} \otimes E_{22}$$

Math behind the magic I
Clausen's Construction



Multiplication table

	E_{11}	E_{12}	E_{21}	E_{22}
E_{11}	E_{11}	E_{12}	0	0
E_{21}	E_{21}	E_{22}	0	0
E_{12}	0	0	E_{11}	E_{12}
E_{22}	0	0	E_{21}	E_{22}

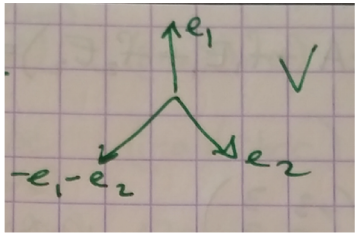
Leads to tensor rank decomposition

$$T_{MM,2} =$$
$$E_{11} \otimes E_{11} \otimes E_{11} + E_{21} \otimes E_{11} \otimes E_{21} + \cdots$$
$$+ E_{22} \otimes E_{22} \otimes E_{22}$$

Math behind the magic I
Clausen's Construction

Consider the action of S_3 on $\mathbb{C}^2$ given by

$$(12) \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad (123) \mapsto \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$



Let $X_0 = Y_0 = Id_2$

Let $X_1 = E_{11}$.

Let X_1, X_2, X_3 be the orbit of X_1 under conjugation by $C_3 \leq S_3$.

Let $Y_i = (12)X_i(12)$.

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Claim: $\{X_{0,}X_1, X_2, X_3\}$ is a basis of M_2. (So is $\{Y_{0,}Y_1, Y_2, Y_3\}$.

Their multiplication table is

	Y_0	Y_1	Y_2	Y_3
X_0	Y_0	Y_1	Y_2	Y_3
X_1	X_1	0	Y_2	X_1
X_2	X_2	Y_1	X_2	0
X_3	X_3	X_3	0	Y_3

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Claim: $\{X_0, X_1, X_2, X_3\}$ is a basis of M_2. (So is $\{Y_0, Y_1, Y_2, Y_3\}$.

Their multiplication table is

	Y_0	Y_1	Y_2	Y_3
X_0	Y_0	Y_1	Y_2	Y_3
X_1	X_1	0	Y_2	X_1
X_2	X_2	Y_1	X_2	0
X_3	X_3	X_3	0	Y_3

The rank decomposition has terms that combine:

$$T_{MM,2} =$$
$$X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots$$
$$+X_3 \otimes (Y_0 + Y_1) \otimes X_3$$

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Claim: $\{X_0, X_1, X_2, X_3\}$ is a basis of M_2. (So is $\{Y_0, Y_1, Y_2, Y_3\}$.

Their multiplication table is

	Y_0	Y_1	Y_2	Y_3
X_0	Y_0	Y_1	Y_2	Y_3
X_1	X_1	0	Y_2	X_1
X_2	X_2	Y_1	X_2	0
X_3	X_3	X_3	0	Y_3

The rank decomposition has terms that combine:

$$T_{MM,2} =$$
$$X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots$$
$$+X_3 \otimes (Y_0 + Y_1) \otimes X_3$$

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Claim: $\{X_0, X_1, X_2, X_3\}$ is a basis of M_2. (So is $\{Y_0, Y_1, Y_2, Y_3\}$.

Their multiplication table is

	Y_0	Y_1	Y_2	Y_3
X_0	Y_0	Y_1	Y_2	Y_3
X_1	X_1	0	Y_2	X_1
X_2	X_2	Y_1	X_2	0
X_3	X_3	X_3	0	Y_3

The rank decomposition has terms that combine:

$$T_{MM,2} =$$
$$X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots$$
$$+ X_3 \otimes (Y_0 + Y_1) \otimes X_3$$

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Claim: $\{X_0, X_1, X_2, X_3\}$ is a basis of M_2. (So is $\{Y_0, Y_1, Y_2, Y_3\}$.

Their multiplication table is

	Y_0	Y_1	Y_2	Y_3
X_0	Y_0	Y_1	Y_2	Y_3
X_1	X_1	0	Y_2	X_1
X_2	X_2	Y_1	X_2	0
X_3	X_3	X_3	0	Y_3

The rank decomposition has terms that combine:

$$T_{MM,2} =$$
$$X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots$$
$$+ X_3 \otimes (Y_0 + Y_1) \otimes X_3$$

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Claim: $\{X_0, X_1, X_2, X_3\}$ is a basis of M_2. (So is $\{Y_0, Y_1, Y_2, Y_3\}$.

Their multiplication table is

	Y_0	Y_1	Y_2	Y_3
X_0	Y_0	Y_1	Y_2	Y_3
X_1	X_1	0	Y_2	X_1
X_2	X_2	Y_1	X_2	0
X_3	X_3	X_3	0	Y_3

The rank decomposition has terms that combine:

$$T_{MM,2} =$$
$$X_0 \otimes Y_0 \otimes Y_0 + X_1 \otimes (Y_0 + Y_3) \otimes X_1 + \cdots$$
$$+ X_3 \otimes (Y_0 + Y_1) \otimes X_3$$

Only 7 terms!

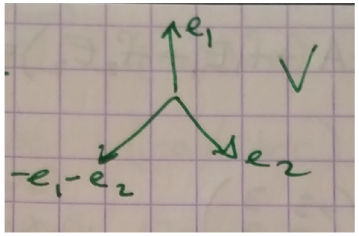
Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Math behind the magic I
Clausen's Construction

Consider the action of S_3 on $\mathbb{C}^2$ given by

$$(12) \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad (123) \mapsto \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$



Let $X_0 = Y_0 = Id_2$

Let $X_1 = E_{11}$.

Let X_1, X_2, X_3 be the orbit of X_1 under conjugation by $C_3 \leq S_3$.

Let $Y_i = (12)X_i(12)$.

Reminiscent of construction of Golay code

Clausen '88 thesis

See Bürgisser-Clausen-Shokrollahi Ch. 1

Similar (if you squint)
Construction of Golay code

Similar (if you squint)
Construction of Golay code

		Y_2	Y_3	Y_4	Y_5	Y_6
1	*AB*	0	0	1	1	1
2	*AC*	0	1	0	1	1
3	*AD*	0	1	1	0	1
4	*AE*	0	1	1	1	0
5	*BC*	1	0	0	1	1
6	*BD*	1	0	1	0	1
7	*BE*	1	0	1	1	0
8	*CD*	1	1	0	0	1
9	*CE*	1	1	0	1	0
10	*DE*	1	1	1	0	0

FIG. 1.20

intersections *AB*, *BE*, *ED*, *DC* and *CA*, which are neighboring symbols in the 5-cycle

(*ABEDC*)

and 1's in the other five row positions. From the set of $4!-1$ such 5-cycles beginning with *A*, other than *ABEDC* itself, pick the eleven that differ from *ABEDC* by an even number of interchanges of elements. In this set of twelve, including *ABEDC*, there are six pairs which differ only in their order. One such pair is

ABEDC and *ACDEB*.

One representative of each pair is matched with one of the six Y_i's, $i = 7, 8, \ldots, 12$ as follows:

Y_7	Y_8	Y_9
(*ABEDC*)	(*ABCED*)	(*ABDCE*)
Y_{10}	Y_{11}	Y_{12}
(*ACEBD*)	(*ACBDE*)	(*ADCBE*)

The columns associated with the eleven Y_i's, $i = 2, 3, \ldots, 12$, are displayed in Figure 1.21.

Thompson, *From Error-Correcting Codes Through Sphere Packings to Simple Groups*, MAA 1984

Similar (if you squint)
Construction of Golay code

		Y_2	Y_3	Y_4	Y_5	Y_6
1	*AB*	0	0	1	1	1
2	*AC*	0	1	0	1	1
3	*AD*	0	1	1	0	1
4	*AE*	0	1	1	1	0
5	*BC*	1	0	0	1	1
6	*BD*	1	0	1	0	1
7	*BE*	1	0	1	1	0
8	*CD*	1	1	0	0	1
9	*CE*	1	1	0	1	0
10	*DE*	1	1	1	0	0

FIG. 1.20

intersections *AB*, *BE*, *ED*, *DC* and *CA*, which are neighboring symbols in the 5-cycle

(*ABEDC*)

and 1's in the other five row positions. From the set of $4!-1$ such 5-cycles beginning with *A*, other than *ABEDC* itself, pick the eleven that differ from *ABEDC* by an even number of interchanges of elements. In this set of twelve, including *ABEDC*, there are six pairs which differ only in their order. One such pair is

ABEDC and *ACDEB*.

One representative of each pair is matched with one of the six Y_i's, $i = 7, 8, \ldots, 12$ as follows:

Y_7	Y_8	Y_9
(*ABEDC*)	(*ABCED*)	(*ABDCE*)
Y_{10}	Y_{11}	Y_{12}
(*ACEBD*)	(*ACBDE*)	(*ADCBE*)

The columns associated with the eleven Y_i's, $i = 2, 3, \ldots, 12$, are displayed in Figure 1.21.

Thompson, *From Error-Correcting Codes Through Sphere Packings to Simple Groups*, MAA 1984

Aside

Second Connection With Codes

Second connection with codes
Lempel-Winograd ’76 (not squinty)

Strassen 2x2 product: 7 products, 18 sums

$$I = (a_{11}+a_{22})(b_{11} + b_{22}) \qquad II = (a_{21} + a_{22})(b_{11})$$

$$III = (a_{11})(b_{12} - b_{22}) \qquad IV = (a_{22})(-b_{11} + b_{21})$$

$$VI = (-a_{11} + a_{21})(b_{11} + b_{12}) \qquad V = (a_{11} + a_{12})(b_{22})$$

$$VII = (a_{12} - a_{22})(b_{21} + b_{22})$$

$$\begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix} = \boxed{\begin{pmatrix} 1 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}} \begin{pmatrix} I \\ II \\ \vdots \\ VII \end{pmatrix}$$

Second connection with codes
Lempel-Winograd ’76 (not squinty)

$$\begin{pmatrix} c_{11} \\ c_{12} \\ c_{21} \\ c_{22} \end{pmatrix} = \boxed{\begin{pmatrix} 1 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}} \begin{pmatrix} I \\ II \\ \vdots \\ VII \end{pmatrix}$$

Thm [Lempel-Winograd]: Rank n decompositions of a tensor with k outputs correspond to $[n, k, d]$ codes with

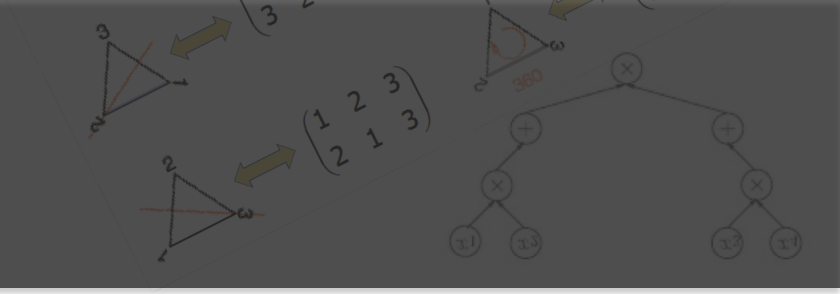
$d \geq$ (something having to do with the tensor).

For $n \times n$ matrix multiplication, these are $[n^{\omega}, n^2, \mathrm{n}]$ codes.

/End Aside

Back to the Story

Math behind the magic II
Unitary 2-Design



Matrix multiplication is characterized by its symmetries (important in Geometric Complexity Theory approach to P versus NP).

Let $\mu: M_n \otimes M_n \to M_n$ be matrix multiplication. Then:

$$\mu(A,B) = X^{-1}\mu(XAY^{-1}, YBZ^{-1})Z$$

for all $X, Y, Z \in GL_n$.

Any bilinear map $f: M_n \otimes M_n \to M_n$ such that

$$f(A,B) = X^{-1}f(XAY^{-1}, YBZ^{-1})Z$$

for all $X, Y, Z \in GL_n$ is a scalar multiple of μ.

Our starting point: For V an irrep of G, write $M_n = V \otimes V^*$, then invariance under G^3 (rather than GL_n^3) suffices!

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Remember This?

$$\sum_{ijk} E_{ij} \otimes E_{jk} \otimes E_{ki}$$

If we write $M_n \cong U \otimes V^*$, then this lives in

$$M_n \otimes M_n \otimes M_n^*$$
$$\cong (U \otimes V^*) \otimes (V \otimes W^*) \otimes (W \otimes U^*)$$
$$\cong U \otimes (V \otimes V^*) \otimes (W^* \otimes W) \otimes U^*$$

In this decomposition, MM is

$$Id_U \otimes Id_V \otimes Id_W$$

2×2 matrix multiplication
Using group orbits

G: Finite group, acting irreducibly on $\mathbb{C}^n$.
S: G-orbit of in $\mathbb{C}^n$. Then (Schur's Lemma)

$$\frac{n}{|S|}\sum_{v\in S}|v\rangle\langle v| = Id$$

Then we have

$$\frac{n^3}{|S|^3}\sum_{u,v,w\in S}|u\rangle\langle u|\otimes|v\rangle\langle v|\otimes|w\rangle\langle w| = Id\otimes Id\otimes Id$$

$$\frac{n^3}{|S|^3}\sum_{u,v,w\in S}|u\rangle\langle v|\otimes|v\rangle\langle w|\otimes|w\rangle\langle u| = MM$$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

Suppose $S \subseteq \mathbb{C}^n$ satisfies

$$\frac{1}{|S|}\sum_{v\in S}|v\rangle\langle v| = \frac{1}{n}Id$$

Then we have

$$\frac{n^3}{|S|^3}\sum_{u,v,w\in S}|u\rangle\langle u|\otimes|v\rangle\langle v|\otimes|w\rangle\langle w| = Id\otimes Id\otimes Id$$

$$\frac{n^3}{|S|^3}\sum_{u,v,w\in S}|u\rangle\langle v|\otimes|v\rangle\langle w|\otimes|w\rangle\langle u| = MM$$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

Suppose $S \subseteq \mathbb{C}^n$ satisfies

$$\frac{1}{|S|}\sum_{v\in S} |v\rangle\langle v| = \frac{1}{n} Id$$

Then we have

$$\frac{n^3}{|S|^3}\sum_{u,v,w\in S} |u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w| = Id \otimes Id \otimes Id$$

$$\frac{n^3}{|S|^3}\underbrace{\sum_{u,v,w\in S}}_{|S|^3 \text{ summands}} |u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u| = MM$$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

Consider

$$\sum_{u,v,w \in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

Consider

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Why consider this?

- Looks kinda like matrix multiplication. Almost G^3-invariant.
- Only need to sum of triples of *distinct* $u, v, w \in S$. $\rightarrow |S|(|S|-1)(|S|-2)$ summands < $|S|^3$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$ $\rightarrow Id^{\otimes 3}$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$ $\rightarrow Id^{\otimes 3}$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$ $\rightarrow$MM

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$ $\rightarrow Id^{\otimes 3}$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$ $\rightarrow$0?
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$ $\rightarrow$0?
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$ $\rightarrow$MM

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

$$\sum_{u,v,w\in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

Expand out, giving 4 kinds of terms:

1. Matched $-|u\rangle\langle u| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$ $\rightarrow Id^{\otimes 3}$
2. 1 mismatch $|u\rangle\langle v| \otimes |v\rangle\langle v| \otimes |w\rangle\langle w|$ $\rightarrow$0?
3. 2 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle w|$ $\rightarrow$0?
4. 3 mismatches $|u\rangle\langle v| \otimes |v\rangle\langle w| \otimes |w\rangle\langle u|$ $\rightarrow$MM

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

If

$$\sum_{u \in S} u = 0 \qquad (*)$$

Then those terms vanish, as desired.

Given $u \in \mathbb{C}^n$, (*) is its projection onto trivial rep.

→If $\mathbb{C}^n$ is a *nontrivial irrep*, the sum must be 0

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

If S is a G-orbit in a nontrivial irrep,

then

$$MM = Id^{\otimes 3} + \frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle v-u| \otimes |v\rangle\langle w-v| \otimes |w\rangle\langle u-w|$$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Proof

If

Unitary 2-Design!

$$\sum_{u \in S} u = 0$$

$$\frac{1}{|S|} \sum_{v \in S} |v\rangle\langle v| = \frac{1}{n} id$$

Then

$$Id^{\otimes 3} + \frac{n^3}{|S|^3} \sum_{u,v,w \in S} |u\rangle\langle v - u| \otimes |v\rangle\langle w - v| \otimes |w\rangle\langle u - w|$$

$$= MM$$

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

The math behind the magic II

Theorem [G.-Moore]: For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s-1)(s-2)+1$ multiplications.

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

The math behind the magic II

Theorem [G.-Moore]: For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s-1)(s-2)+1$ multiplications.

Observe: If G finite group, V nontrivial irrep, $v \in V$ w/ $|v| = 1$, then the orbit Gv is a unitary 2-design.

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

The math behind the magic II

Theorem [G.-Moore]: For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s-1)(s-2)+1$ multiplications.

The action of S_3 on $\mathbb{C}^2$ has an orbit of size 3 (equilaterial triangle).

Corollary [G.-Moore]: $3(3-1)(3-2)+1 = 7$.

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

The math behind the magic II

Theorem [G.-Moore]: For $S \subseteq \mathbb{C}^n$ a unitary 2-design, $s = |S|$, $n \times n$ matrices can be multiplied using at most $s(s-1)(s-2)+1$ multiplications.

The action of S_{n+1} on $\mathbb{C}^n$ has an orbit of size n+1. Smallest possible, since must sum to zero.

Corollary [G.-Moore]: $(n+1)n(n-1)+1 = n^3 - n + 1 < n^3$.

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Summary of
The math behind the magic

Okay, but is that the same as Strassen's algorithm?

- Yes, by a theorem of de Groote '78
- But…we don't care! It gives a conceptual explanation of the upper bound of 7.

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

Summary of
The math behind the magic

Okay, but is that the same as Strassen's algorithm?

- Yes, by a theorem of de Groote '78
- But…we don't care! It gives a conceptual explanation of the upper bound of 7.

Third Connection: Uniqueness, similar to uniqueness of Golay code

Summary of
The math behind the magic

Gave conceptual explanation of Strassen's 7.

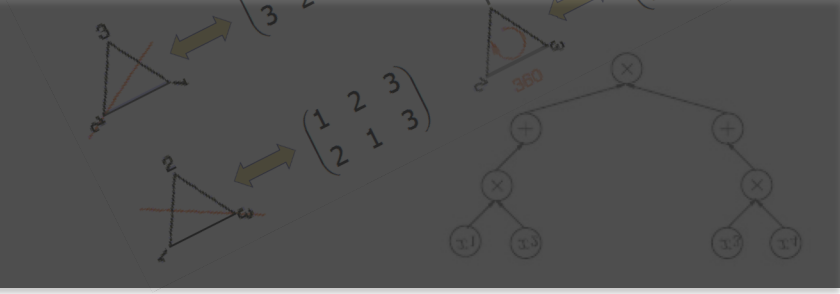
Open: Find a similar explanation that works over arbitrary rings (as Strassen's algorithm does; ours needs ½, $1/3$, and $\sqrt{3}$). (Arithmetic geometry)

Further Ideas:

- Use of unitary d-designs for $d > 2$? (cf. iterated matrix multiplication)
- Other symmetric algorithms? See also Burichenko; Chiantini-Ikenmeyer-Landsberg-Ottaviani.

G. & Moore arXiv:1612.01527
and arXiv:1708.09398

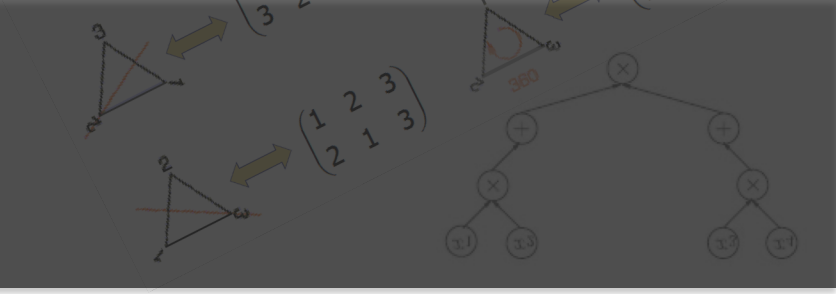
Other explanations
Of Strassen's 2×2 algorithm



Gastinel 1971
Yuval 1978
Chatelin 1986
Clausen 1988 (thesis)
Alekseyev 1996
Gates & Kreinovich 2000 (*Bull. EATCS*)
Paterson 2009 (Valiant's 60th)
Minz 2015
Chiantini, Ikenmeyer, Landsberg, Ottaviani 2016
Ikenmeyer & Lysikov 2017

Uses/reveals symmetries

Additional connections
With code-like things



- Some algorithms highly symmetric, some not. Symmetry, optimality, and uniqueness.
- Arithmetic progression-free sets (Coppersmith-Winograd '90)
- Cohn-Umans '03 group-theoretic approach has a packing problem at its core
- Coherent configurations (Cohn-Umans '13)
- Can phrase MM as a minimization problem. Similar (squint) to spherical codes / the Thomson problem
- Combine algorithms for small MM into algorithms for larger MM (e.g. Hopcroft-Kerr '71, Strassen's laser method). Similar (squint!) to construction of Leech lattice

Additional references

(Note hyperlinks throughout the document – the ones here are *additional*)

Symmetries in algorithms for matrix multiplication

- Burichenko arXiv:1408.6273
- Landsberg & Ryder, Exp. Math. 2017
- Landsberg & Michałek, arXiv:1601.08229

Cap sets & Cohn-Umans approach

- Blasiak, Church, Cohn, G. Naslund, Sawin, Umans, Disc. Analysis 2017
- Blasiak, Church, Cohn, G., Umans arXiv:1712.02302