Matrix Multiplication (MM)

- Given two matrices $x = (x_{ij})_{n \times n}$ & $y = (y_{ij})_{n \times n}$,
  we want to compute their product $xy = (z_{ij})_{n \times n}$, over ring R.

- By definition, $z_{ij} = \sum_{k=1}^{n} x_{ik} \cdot y_{kj}$.

▷ Naively, MM requires n^3 multiplications & $n^2(n-1)$ additions.

- Could we reduce the number of multiplications at the cost of additions, for a fixed n?

▷ Strassen (1969) showed how to multiply 2×2 matrices using 7 mult. but 18 additions!

The 7 products:

- We want to compute $(z_{ij})_{2\times 2} = x \cdot y$.

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

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

- Since, the above holds for any ring R, we can apply this to design a <u>recursive</u> algorithm for MM.

- <u>Idea</u>: Block MM of general matrices x, y.

Theorem (Strassen, 1969): MM can be done in $O(n^{\lg 7})$ R-operations.

Pf:

- Let $x, y \in R^{n \times n}$, with $n = 2^{\ell}$, $\ell \in \mathbb{N}$.

- We will show, by induction on ℓ, that we can do MM in 7^{ℓ} R-mult. & $6 \cdot (7^{\ell} - 4^{\ell})$ R-addn.

- Base case ($\ell = 1$): As above.

- Induction ($\ell - 1 \to \ell$): We use the following block structure of x & y:

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

where, x_{ij}, y_{ij}, z_{ij}'s are $2^{\ell-1} \times 2^{\ell-1}$ matrices.

- Clearly, Strassen's eqns. (for 2×2)

hold for these matrices as well.

- By induction:

$\# R\text{-mult.} = 7 \times (7^{\ell-1}) = 7^{\ell}$.

$\# R\text{-addn.} = 7 \times (6 \cdot 7^{\ell-1} - 6 \cdot 4^{\ell-1}) +$ for the recursive calls

matrix addns. $18 \times (2^{\ell-1})^2$

$= 6 \cdot (7^{\ell} - 4^{\ell})$

$\Rightarrow$ Overall, $O(7^{\ell}) = O(n^{\lg 7})$ R-operations.

$\lg 7 \sim 2.8$

$\square$

– After decades of work, the current best algorithm for MM has complexity $O(n^{2.3728639})$ (Le Gall, 2014).

<u>Conjecture</u>: MM has complexity $O(n^{2+\varepsilon})$, for any $\varepsilon > 0$.

The exponent of MM

- Let us denote the exponent of MM by ω.

▷ It is known that $2 \le \omega < 2.3728639$.

- All the upper bound methods for ω use the notion of tensor rank.

Definition: The MM tensor is a polynomial in $R[X_{ij}, Y_{ij}, Z_{ij} \mid 1 \le i \le j \le n]$, namely:

$$T_{n,R} := \sum_{i,j,k \in [n]} X_{ik} \cdot Y_{kj} \cdot Z_{ij} .$$

- Eg. $T_{2,R} = Z_{11} \cdot (X_{11} Y_{11} + X_{12} Y_{21}) + Z_{12} \cdot (X_{11} Y_{12} + X_{12} Y_{22}) + Z_{21} \cdot (X_{21} Y_{11} + X_{22} Y_{21}) + Z_{22} \cdot (X_{21} Y_{12} + X_{22} Y_{22})$.

- There are clever notions of decomposition:

<u>Definition</u>: <u>Rank</u> $r(T)$ of the tensor T is the least $r \in \mathbb{N}$ s.t. $\exists$ <u>linear</u> forms $L_i \in R[\bar{X}]$, $M_i \in R[\bar{Y}]$, $N_i \in R[\bar{Z}]$, $i \in [r]$ satisfying,

$$T = \sum_{i \in [r]} L_i \cdot M_i \cdot N_i \quad .$$

$\triangleright \quad n^2 \leq r(T_{n,R}) \leq n^3 .$

<u>Pf</u>: • By evaluating $T_{n,R}$ at suitable points, we can make it <u>zero</u>.

$$\Rightarrow \quad r(T_{n,R}) \geq n^2 . \quad \text{(Exercise)}$$

• By the definition of $T_{n,R}$, we have $r(T_{n,R}) \leq n^3$. $\square$

- It is easy to see that $r(T_{n,R})$ upper bounds the <u>mult.</u>-complexity of MM. (this is crucial in recursive MM)

Recursively going from n_0 to n gives $\omega \lesssim \log_{n_0} r(T_{n_0})$.

▷ MM can be done in $r(T_{n,R})$ R-multiplications.

Pf sketch:

- Tensor T & its rank $r(T)$ is defined in a way that each entry z_{ij} could be computed by using the same set of $r(T)$ products. □

- Eg. Strassen's algorithm is inspired from the decomposition:

$$T_{2,R} = p_1(\bar{x},\bar{y})\cdot(z_{11}+z_{22}) + p_2\cdot(z_{21}-z_{22}) + p_3\cdot(z_{12}+z_{22}) + p_4\cdot(z_{11}+z_{21}) + p_5\cdot(-z_{11}+z_{12}) + p_6\cdot(z_{22}) + p_7\cdot(z_{11}).$$

- In fact, it can be shown that $r(T_{2,R}) = 7$.

[Håstad '90]: Tensor rank computation is NP-hard.

OPEN: $r(T_{3,R})$ not known. $(19 \le r \le 23)$