Matrix Multiplication (MM)

- Given two matrices $x = (x_{ij})_{i,j\in[n]}$ & $y = (y_{ij})_{i,j\in[n]}$ over R.

We want to compute $x \cdot y =: z = (z_{ij})_{n\times n}$.

▷ $z_{ij} = \sum_{k\in[n]} x_{ik} \cdot y_{kj}$.

▷ Naively, MM requires n^3 R-mult. & $n^2(n-1)$ R-addn.

- Qn: Could we reduce R-mult. at the cost of R-addn. ? (Say, n fixed ?)

▷ Strassen (1969) showed how to multiply 2×2 matrices using $2^3-1 = 7$ mult. (but 18 addn.).

The 7 products:

$p_1 = (x_{11}+x_{22})(y_{11}+y_{22})$

$p_2 = (x_{21}+x_{22})\,y_{11}$

$p_3 = x_{11}(y_{12}-y_{22})$

$p_4 = x_{22}(-y_{11}+y_{21})$

$p_5 = (x_{11}+x_{12})\,y_{22}$

$p_6 = (-x_{11}+x_{21})(y_{11}+y_{12})$

$p_7 = (x_{12}-x_{22})(y_{21}+y_{22})$

$$\triangleright \quad x \cdot y = z = \begin{pmatrix} p_1 + p_4 - p_5 + p_7 & p_3 + p_5 \\ p_2 + p_4 & p_1 + p_3 - p_2 + p_6 \end{pmatrix}$$

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

<u>Theorem</u> [Strassen '69]: MM takes $O(n^{\lg 7})$ R-ops.

Pen Eraser Text Undo Redo SAVE MORE

$$\triangleright \quad x \cdot y = z = \begin{pmatrix} p_1+p_4-p_5+p_7 & p_3+p_5 \\ p_2+p_4 & p_1+p_3-p_2+p_6 \end{pmatrix}$$

- Since, the above holds for any ring R, we can apply this to design a recursive algorithm for MM. [Use halving of n.]

Theorem [Strassen'69]: MM takes $O(n^{\lg 7})$ R-ops.

Pf: • Let $x, y \in R^{n \times n}$, $n =: 2^{\ell}$.

• We'll show by induction on ℓ that we can do MM in

7^ℓ R-mult. & $6\cdot(7^\ell - 4^\ell)$ R-addns.

Base case: $[\ell=1]$: already seen.

Ind. step $[\ell-1 \to \ell]$: Use block-structure mult.:

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \cdot \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} =: \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix}_{2^\ell \times 2^\ell}$$

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

So, 1) Use strassen's eqns. of 2×2 matrices (over $R^{\frac{n}{2}\times\frac{n}{2}}$)

2) & recursion to $\frac{n}{2}$.

lec8.1-20200128.mp4 Pen Eraser Text Undo Redo SAVE LIST MORE

9:05 1:00:00 5 / 55

Time taken: (by induction)

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

$\#R\text{-addns} = \underset{\uparrow}{7} \times 6(7^{\ell-1} - 4^{\ell-1}) + \underset{\uparrow}{18} \times (2^{\ell-1})^2$

7 recursive calls   Strassen's eqns

$= 6 \cdot (7^{\ell} - 4^{\ell})$ ✓

$\Rightarrow$ Overall, $O(7^{\ell}) = O(n^{\log 7})$ R-ops. $\square$

$\log 7 \approx 2.8$

- After decades of work, the current best for MM is $O(n^{2.3728639})$ [Le Gall, 2014]

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

The exponent of MM

- Let's denote it by $\underline{\omega}$.

  $\triangleright \quad 2 \leq \omega < 2.3728639$.

- All the known upper bounds on ω use the notion of <u>tensor rank</u>.

<u>Defn</u>: The <u>MM-tensor</u> is a polynomial in $R[X_{ij}, Y_{ij}, Z_{ij} \mid i \leq j \in [n]]$ namely:

$$\underline{T_{n,R}} := \sum_{i,j,k \in [n]} X_{ik} Y_{kj} Z_{ij} \qquad \text{(cubic poly.)}$$

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

Defn: Rank $r(T)$ of tensor T is the least r s.t. $\exists$ linear forms $L_i \in R[\bar{x}]$, $M_i \in R[\bar{y}]$, $N_i \in R[\bar{z}]$, $i \in [r]$ satisfying:

$$T = \sum_{i\in[r]} \underbrace{L_i \cdot M_i \cdot N_i}_{\text{simple tensor of } rk=1}.$$

$\to$ order-2 tensor; $T(\bar{x},\bar{y}) = \sum_{i\in[r]} L_i(\bar{x}) \cdot M_i(\bar{y})$

[Exercise: Matrix Rk?]

Pen Eraser Text Undo Redo SAVE MORE

▷ $n^2 \leq r(T_{n,R}) \leq n^3$.

Pf sketch: $[\leq n^3]$: by defn of rk.

$[\geq n^2]$: Fixing variables in $T_{n,R}$ gives a contradiction. $\square$

Claim: MM_n can be done in $r(T_{n,R})$ many R-multiplications.

Pf: • Say, $T_{n,R} = \sum_{i\in[r]} L_i(\bar{x}) \cdot M_i(\bar{y}) \cdot N_i(\bar{z})$

• Extract the coefficient of z_{11} both sides & so on.

• Products $\{L_i \cdot M_i \mid i \in [r]\}$ are enough. $\square$

Pen Eraser Text Undo Redo SAVE MORE

— Strassen's eqns show $r(T_{2,R}) \leq 7$.

OPEN: $r(T_{3,R})$ is unknown.

▷ " $\in [19, 23]$.

[Håstad '90]: (order-3) tensor rank computation is NP-hard.

▷ $r(T_{n_0}) =: r_0 \implies \omega \leq \log_{n_0} r_0$.

→ Known results define tensors "related to" $T_{n,R}$.

60/60