Multilinear models

- Multilinearity is a natural restriction on arithmetic circuits.

  Eg. det & per are multilinear polynomials. Can they be computed without computing monomials like x_i^j, $j>1$?

Defn: A circuit C is multilinear if every gate computes a multilinear polynomial.

- We will first focus on multilinear depth-3 & prove exponential lower bounds.

- The measure, on polynomials, that will succeed is as follows.

  Partition the variables X as $Y \sqcup Z$. For a monomial m_Y (resp. m_Z) in the Y (resp. Z) variables, let $\mathrm{coef}(m_Y \cdot m_Z)(f)$ denote the coefficient of $m_Y m_Z$ in $f(X)$.

Define matrix $M_{Y,Z}(f)$ as:

$$\begin{array}{cc} & m_Z \\ m_Y & \underbrace{\begin{pmatrix} & & \\ & \mathrm{coef}(m_Y m_Z)(f) & \\ & & \end{pmatrix}}_{\text{monomials in } Z} \left.\vphantom{\begin{pmatrix} \\ \\ \\ \end{pmatrix}}\right\} \text{monomials in } Y \end{array}$$

<u>Defn</u>: $\Gamma_{Y,Z}(f) := \mathrm{rk}\ M_{Y,Z}(f)$.

For multilinear f, $\mathrm{coef}(m_Y m_Z)(f) = (\partial_{m_Y m_Z} f)(\bar{0})$. So, $M_{Y,Z}(f)$ is also called the partial derivative matrix of f.

- It behaves well under ring operations:

<u>Lemma</u> (Sub-additivity): $\Gamma_{Y,Z}(f_1+f_2) \leq \Gamma(f_1) + \Gamma(f_2)$.

<u>Pf</u>: Follows from the rank property of $A+B$ for matrices. $\square$

<u>Lemma</u> (Multiplicativity): For $f_1 \in \mathbb{F}[Y_1, Z_1]$, $f_2 \in \mathbb{F}[Y_2, Z_2]$ with $Y = Y_1 \sqcup Y_2$ & $Z = Z_1 \sqcup Z_2$,

we have $\Gamma_{Y,Z}(f_1 \cdot f_2) = \Gamma_{Y_1,Z_1}(f_1) \cdot \Gamma_{Y_2,Z_2}(f_2)$.

<u>Proof:</u>

- Note that $M_{Y,Z}(f_1 \cdot f_2)$ equals the tensor product $M_{Y_1,Z_1}(f_1) \otimes M_{Y_2,Z_2}(f_2)$.

- This follows from the disjointness of these subsets, which allows
$\mathrm{coef}(m_Y m_Z)(f) = \mathrm{coef}(m_{Y_1} m_{Z_1})(f) \cdot \mathrm{coef}(m_{Y_2} m_{Z_2})(f)$.

- Rank property of a tensor product gives
$\Gamma_{Y,Z}(f) = \prod_{i \in [2]} \Gamma_{Y_i,Z_i}(f_i)$. $\square$

<u>Lemma</u> (Mult. by Z-free): For any $g \in \mathbb{F}[Y]^*$,
$\Gamma_{Y,Z}(g \cdot f) = \Gamma_{Y,Z}(f)$.

<u>Proof:</u>

- If we consider the function field $\mathbb{F}(Z)$, then g, f can be considered as polynomials in $\mathbb{F}(Z)[Y]$.

- By "coefficient" extraction we can prove:

$$\Gamma_{Y,Z}(gf) = rk_{\mathbb{F}}\{(\partial_m(gf))_{Z=0} \mid m \text{ is a monomial in } Z\}$$ result in $\mathbb{F}[Y]$.

[Pf idea: column m_Z in $M_{Y,Z}(gf)$ exactly represents the polynomial in $\mathbb{F}[Y]$ which is the "coefficient" of m_Z in (gf).
$\Rightarrow$ the rank of these Y-polynomials equals the column rank of $M_{Y,Z}(gf)$.]

- Now, using the Z-freeness of g, we get:

$$\Gamma_{Y,Z}(gf) = rk_{\mathbb{F}}\{g\cdot(\partial_m f)_{Z=0} \mid m \text{ is a monomial in } Z\}$$
$$= \Gamma_{Y,Z}(f). \qquad \square$$

<u>Lemma</u>: For any multilinear f, we have
$$\Gamma_{Y,Z}(f) \leq 2^{\min(|Y|,|Z|)}.$$

<u>Proof</u>:

- Follows from the size of $M_{Y,Z}(f)$. $\square$

- Eg. $f(Y,Z) = \prod_{i\in[n]} (y_i + z_i)$ proves the optimality of the above upper bound as:
$$\Gamma_{Y,Z}(f) = \prod_{i\in[n]} \Gamma_{y_i,z_i}(y_i + z_i) = 2^n.$$

- The above example is merely depth-2! Raz showed that the measure for $\Pi\Sigma$ can be significantly reduced if we consider a random partition of X.

Theorem (Upper bound) [Raz'09]: Let $f(X) = \ell_1 \cdots \ell_d$ be an n-variate multilinear. For a random partition $X = Y \sqcup Z$, $|Y| = |Z| = n/2$, we have whp:
$$\Gamma_{Y,Z}(f) \leq 2^{n/2 - n/32}.$$

Proof: • Wlog each ℓ_i has support-size ≥ 2, as univariate ℓ_i's do not change the measure wrt <u>any</u> partition.

(by multiplicativity)

- Clearly, $T_{Y,Z}(f) \le 2^d$. Hence, we are done if $d < n/3$.

Assume $d \geq n/3$.

- Since l_i's are disjoint-support & many ($\geq n/3$) we get by an averaging argument that:

l_i's with support-size 2 or 3 is $\geq d/4$.

[Otherwise, $> 3d/4$ l_i's have support-size $\geq 4 \Rightarrow n > 3d$, a #.]

- We call these l_i's <u>small</u>.
- Now for a small l_i, we have

$$\Pr_{Y,Z}[\text{support}(l_i) \subseteq Y \text{ or } Z] \geq 2 \cdot \frac{1}{2^3} = \frac{1}{4}.$$

$$\Rightarrow \operatorname{Exp}_{Y,Z}[\#i \mid \text{small } l_i \text{ is in } \mathbb{F}[Y], \mathbb{F}[Z]] \geq d/16.$$

- These l_i's stop contributing to $T_{Y,Z}(f)$.

$\Rightarrow \Gamma_{Y,Z}(f) \leq 2^{d - d/16} \leq 2^{n/2} \cdot 2^{-n/32}$.

(as $n/3 \leq d \leq n/2$) $\square$

$\underline{\text{Det}_n \text{ \& } \text{Per}_n \text{ have high } \Gamma_{Y,Z}(\cdot)}$

- Note that $\det_n$ has n^2 variables. We will first reduce its variables to a random X, $|X| = 2m := 2 \cdot \frac{\sqrt{n}}{5}$, & then use a random partition $X = Y \sqcup Z$.

<u>Theorem</u> (Lower bound) [Raz'09]: With probability $\geq 1/2$, a random restriction σ of $\{x_{11}, \ldots, x_{nn}\}$ to $X = Y \sqcup Z$, $|Y| = |Z| = m := \lfloor \sqrt{n}/5 \rfloor$, yields $\Gamma_{Y,Z}(\sigma \circ \det_n) = 2^m$.

<u>Proof:</u>

- The map σ will fix $n^2 - 2m$ variables to $\mathbb{F}$ values, in a certain way.
- The remaining $2m$ variables are X.
- Let us compute the probability that

these variables <u>do not</u> share the same row or column.

- $\Pr_\sigma$ [X have diff. rows/cols]

$$= \frac{n^2}{n^2} \cdot \frac{(n-1)^2}{n^2} \cdot \ldots \cdot \frac{(n-2m+1)^2}{n^2} = \left(1-\frac{1}{n}\right)^2 \cdots \left(1-\frac{2m-1}{n}\right)^2$$

$$> \left(1 - \sum_{i=1}^{2m-1} \frac{i}{n}\right)^2 > 1 - \frac{2m \cdot (2m-1)}{n} > 1 - \frac{4}{25} .$$

- For such a σ the determinant shares properties with:

$$\begin{vmatrix} y_1 & 1 & & & & & \\ 1 & z_1 & & & & & 0 \\ & & \ddots & & & & \\ & & & y_m & 1 & & \\ & & & 1 & z_m & & \\ & 0 & & & & 1 & \\ & & & & & & \ddots \end{vmatrix}$$

$$= (y_1 z_1 - 1) \cdots (y_m z_m - 1) =: D_m .$$

error prob $\leq \frac{1}{3}$ if $|\mathbb{F}| \geq 3$. → • In particular, we have whp, for a random σ : $T_{Y,Z}(\sigma(\det_n)) = T_{Y,Z}(D_m)$. (Exercise!)

- Clearly, $T_{y,z}(D_m) = 2^m$. (Multiplicativity)

$$\Rightarrow \Pr_{\sigma}\left[T_{y,z}(\sigma \circ \det_n) = 2^m\right] > 1 - \frac{4}{25} - \frac{1}{3}$$
$$> 1/2. \quad \square$$

- Finally, we deduce an exponential lower bound against multilinear depth-3.

Corollary: $\det_n$ or per_n require $2^{\Omega(\sqrt{n})}$ size multilinear depth-3 circuits.

Proof:

- Suppose $\det_n = C(\bar{x})$, for a multilinear $\Sigma^s \Pi \Sigma$ circuit C.
- We apply, as before, a random variable reduction σ both sides
  $\Rightarrow \sigma \circ \det_n = \sigma \circ C$ (2m-variate now).
- The two theorems imply that
  $$2^m \leq s \cdot 2^{2m/2 - 2m/32}$$

$$\Rightarrow \quad s \geq 2^{m/16} = 2^{\sqrt{n}/80}. \qquad \square$$

- Note that this almost matches the best depth-3 complexity of $\det_n$.

Exercise: The same argument holds for per_n.

Generalizing to constant-depth multilr

- (Raz, Yehudayoff '09) generalized the above ideas to get a result for multilinear depth-Δ circuits.

- Here, instead of a product of linear polynomials we work with the following:

Defn: A multilinear polynomial $f = g_1 \cdots g_t$ is called a t-product if:
each g_i depends on $\geq t$ variables.

Lemma: Let f be a multilinear n-variate d-degree polynomial that has a size-s multilinear (product) depth-Δ formula ϕ. Then, f can be written as a sum of: $\leq s$ multilinear t-products ($t = (n/100)^{1/2\Delta}$) & a multilinear polynomial of degree $\leq n/100$.

Proof:

- If $d \leq n/100$, then it is clear.
- Let $d > n/100$. Since ϕ is a formula of product-depth Δ, there is a product gate v of fanin $\geq (n/100)^{1/\Delta} =: t^2$.
- Let us expand the formula wrt this gate:

$$f = \underbrace{\phi_v}_{\text{output at } v} + \phi_{v=0}$$

- As ϕ_v is a product of t^2 polynomials we can group them to see that ϕ_v is multilinear t-product.
- As $\phi_{v=0}$ is of smaller size, we can recurse.

$\square$

- Now we need to study the effect of a random partitioning on a t-product.

<u>Lemma:</u> Let $f(X)$ be n-variate & computable by a size-s multilinear depth-Δ formula. If $X = Y \sqcup Z$, $|Y| = |Z| = n/2$, is random then with probability $1 - s \cdot \exp(-n^{\Omega(1/\Delta)})$:

the constants in Ω behave well

$$\Gamma_{Y,Z}(f) = s \cdot 2^{n/2} \cdot \exp(-n^{\Omega(1/\Delta)}).$$

<u>Proof:</u>

- By the previous lemma, write $f = g_0 + \sum_{i=1}^{s} g_i$ where $\deg g_0 \leq n/100$ & $g_1, \ldots, g_s$ are multilinear t-products.
- Note that g_0's sparsity can be at most
$$\sum_{i \leq n/100} \binom{n}{i} = 2^{H_2(1/100) \cdot n - O(\lg n)} < 2^{n/10}.$$

$$\Rightarrow \Gamma_{Y,Z}(g_0) < 2^{n/10} \quad \text{(sub-additivity)}.$$

- All that remains is to bound $\Gamma_{Y,Z}(g_1)$ for a random partition $X = Y \sqcup Z$.

• Let $g = h_1 \cdots h_t$, $h_i \in \mathbb{F}[X_i]$, be a t-product for $X = \sqcup X_i$.

Let $Y_i := X_i \cap Y$ & $Z_i := X_i \cap Z$.

• Let $d_i := |\#Y_i - \#Z_i|/2$ be the <u>imbalance</u> between Y_i, Z_i in h_i.

X_i is called <u>k-imbalanced</u> if $d_i \geq k$.

Let $b_i := (\#Y_i + \#Z_i)/2 = \#X_i/2$.

• We have $T_{Y,Z}(g) = \prod_i T_{Y,Z}(h_i) \leq \prod_i 2^{\min(|Y_i|,|Z_i|)}$
$$= \prod_i 2^{b_i - d_i} = 2^{|X|/2} / \prod_i 2^{d_i} .$$

$\Rightarrow$ it suffices to show that one of the X_i's is imbalanced (i.e. d_i is large).

• We need to estimate $|Y_i|$ on choosing a random $Y \in \binom{[n]}{n/2}$.

• The relevant probability is that of the <u>hypergeometric distribution</u>:

<u>Claim</u>: For a fixed set $A \in \binom{[n]}{a}$, $k \le a \le 2n/3$,

$$\Pr_{R \in \binom{[n]}{n/2}} \left[|R \cap A| = k \right] = O(1/\sqrt{a}).$$

<u>Proof</u>:
- $\Pr_R [|R \cap A| = k] = \binom{a}{k} \cdot \binom{n-a}{n/2-k} / \binom{n}{n/2}$
- Call it $P(k)$.
- $P(k+1) > P(k)$ iff

$$(a-k)\left(\frac{n}{2}-k\right) > (k+1)\left(\frac{n}{2}-a+k+1\right) \text{ iff}$$

$$\frac{an}{2} - k\left(a+\frac{n}{2}\right) > \left(\frac{n}{2}-a+1\right) + k\left(\frac{n}{2}-a+2\right)$$

iff $k < \frac{a-1}{2}$.

- Thus, $P(k) \le \binom{a}{\frac{a-1}{2}} \cdot \binom{n-a}{\frac{n}{2}-\frac{a-1}{2}} / \binom{n}{n/2}$

(Stirling's approx.) $$= O\left(\sqrt{\frac{n}{a(n-a)}}\right) = O\left(\frac{1}{\sqrt{a}}\right). \quad \square$$

(k-balanced)
- Let $\underline{\mathcal{E}_i}$ denote the event that $d_i < k$.
- We have $\Pr\left[\bigwedge_{i=1}^{t} \mathcal{E}_i\right]$ equal to

$$\Pr[\mathcal{E}_1] \cdot \Pr[\mathcal{E}_2 \mid \mathcal{E}_1] \cdot \Pr[\mathcal{E}_3 \mid \mathcal{E}_1 \wedge \mathcal{E}_2] \cdots\cdots .$$

- $\Pr[\mathcal{E}_1] = \Pr_Y\left[|Y \cap X_1| \in [b_1 - k, b_1 + k] \right]$

which the above claim estimates as:

$$k \cdot O(1/\sqrt{b_1}) \quad \text{(assuming } k \leq b_1/2\text{)}.$$

- Consider the event $\mathcal{E}_i$ given $\mathcal{E}_1, \dots, \mathcal{E}_{i-1}$. Since $X_1, \dots, X_{i-1}$ have been partitioned in a fairly balanced way ($\forall j \in [i-1], d_j < k$), we deduce that $|Y \cap (X_1 \cup \dots \cup X_{i-1})^c|$

$$= |Y \cap X| - |Y \cap (X_1 \cup \dots \cup X_{i-1})|$$
$$< n/2 - (b_1 - k + \dots + b_{i-1} - k)$$
$$= (n/2 - b_1 - \dots - b_{i-1}) + (i-1)k$$

$\Rightarrow$ The partition of $X' := (X_1 \sqcup \dots \sqcup X_{i-1})^c$ by $Y \sqcup Z$ is (ik)-balanced.

- So, assuming $ik \ll n$, we can redo the calculation in the above claim & still get $\Pr[\mathcal{E}_i | \mathcal{E}_1 \wedge \dots \wedge \mathcal{E}_{i-1}] = k \cdot O(1/\sqrt{b_i})$.

$$\Rightarrow \Pr\Big[\bigwedge_{i \in [t]} \mathcal{E}_i\Big] = O\left(k^t / \sqrt{b_1 \cdots b_t}\right)$$

$$\Rightarrow \Pr_Y\left[T_{Y,Z}(g) > 2^{|X|/2} \cdot 2^{-k}\right] = O\left(k^t / \sqrt{b_1 \cdots b_t}\right)$$

• In particular, on fixing $k \leq t^{1/3}$, we get: $\Pr_Y[T_{Y,Z}(g) > 2^{n/2} \cdot 2^{-k}] = O(\prod_{i=1}^{t} t^{-1/6})$
$= O(t^{-t/6}) = \exp(-n^{\Omega(1/\Delta)})$.

Ω independent of ε if $\varepsilon \leq 1/3$ →

$\Rightarrow \Pr_Y[T_{Y,Z}(f) > s \cdot 2^{n/2-k} =: s 2^{n/2} \cdot 2^{-t^{\varepsilon}}]$
$= s \cdot \exp(-n^{\Omega(1/\Delta)})$.

$\square$

— Thus, there is an $\varepsilon \in (0, \frac{1}{3}]$ such that if $s \leq \exp(t^{\varepsilon})$ then
$\Pr[T_{Y,Z}(f) = 2^{n/2}] < 1/10$.

$\Rightarrow$ $f(X)$ could compute $\det(X)$ only if $s > 2^{t^{\varepsilon}} = \exp(n^{\Omega(1/\Delta)})$.

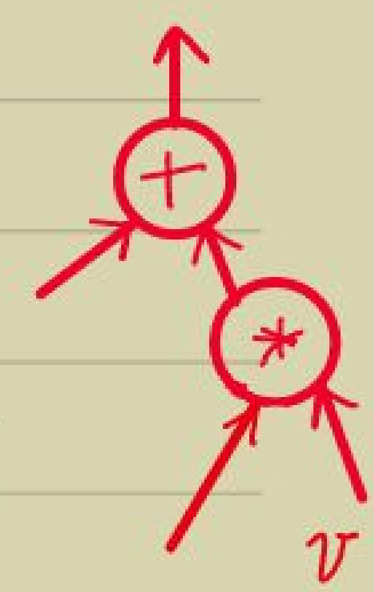
— This finishes (Raz, Yehudayoff '09) proof for $\det_n$ or per_n against constant-depth multilinear model.

- We can also say something for <u>multilinear formulas</u> using the probability calculation seen above.

- The multilinear products of interest there are:

<u>Defn</u>: Multilinear $f = \prod_{i=1}^{t} g_i$, with partition $X = \bigsqcup_{i \in [t]} X_i$, is called a <u>log-product</u> if for all i, $|X|/3^i \leq |X_i| \leq 2 \cdot |X|/3^i$ and $|X_t| = 1$.

<u>Lemma</u>: Any size-s multilinear formula ϕ can be written as a sum of $(s+1)$ log-products.



<u>Proof</u>: • Let $|X| > 2$ & ϕ compute f.

• Let v be a node in ϕ that depends on variables X_v such that

assume fanin 2 in Φ → $$|X|/3 \leq |X_v| \leq 2 \cdot |X|/3$$

• By the formula properties, we have

$$f = \phi_v \cdot g + \phi_{v=0}$$

for some $g \in \mathbb{F}[X \setminus X_v]$.

- Note that $|X|/3 \leq |X \setminus X_v| \leq 2 \cdot |X|/3$.
- Moreover, since g has size $< s$, we can use induction & write it as a sum of $\leq$ size$(g)+1$ log-products.

Similarly, for $\phi_{v=0}$.

$\Rightarrow$ f is a sum of $(s+1)$ log-products.

$\square$

– Now, we can estimate $\Gamma_{Y,Z}(h_1 \cdots h_t)$ for a log-product $h_1 \cdots h_t$, $t = O(\lg n)$.

Note that around $\frac{1}{2}\lg n$ many of these h_i's do depend on at least $\sqrt{n}$ many variables each.

$\Rightarrow$ On doing the probability calculation we will get that $\Gamma_{Y,Z}(h_1 \cdots h_t)$ is high with prob. smaller than $n^{-\Omega(\lg n)}$.

(Raz '09) $\Rightarrow$ $\det_n$ or per_n requires $n^{\Omega(\lg n)}$ size !