Degree-restricted depth-4

- Recall that a depth-4 circuit of the type $\Sigma\Pi^a\Sigma\Pi^b$ has the form $f = \sum_{i\in[s]} Q_{i1}\cdots Q_{ia}$ in $\mathbb{F}[x_1,..,x_n]$, where $\deg(Q_{ij}) \le b$.

- We know that a size-s deg-d f has a $\Sigma\Pi^{O(\sqrt{d})}\Sigma\Pi^{\sqrt{d}}$ circuit of size $s^{O(\sqrt{d})}$.

  $\omega(\cdot)$ is small omega → Conversely, if f requires $s^{\omega(\sqrt{d})}$ size $\Sigma\Pi^{O(\sqrt{d})}\Sigma\Pi^{\sqrt{d}}$ circuits then it requires $s^{\omega(1)}$ size arbitrary circuits.

- To study this model (Kayal '12) modified the partial derivative based measures.

Definition: Let $\partial^{=k}(f)$ be the set of order-k partial derivatives of f & $x^{\le l}$ be the monomials of $\deg \le l$.

The shifted partials of f, denoted

by $\langle \partial^{=k} f \rangle_{\leq \ell}$, is the $\mathbb{F}$-vector space spanned by $\{ x^{\bar{e}} \cdot \partial_{\bar{d}} f \mid |\bar{e}| \leq \ell, |\bar{d}| = k \}$.

The dimension of shifted partials is denoted by $\Gamma_{k,\ell}(f)$.

- The matrix, wrt f, whose rank we are interested in is:

$$\begin{array}{ccc} & m & \\ x^{\bar{\alpha}} \cdot \partial_{\bar{\beta}} & \underbrace{\begin{pmatrix} \operatorname{coef}(m)(x^{\bar{\alpha}} \cdot \partial_{\bar{\beta}} f) \end{pmatrix}}_{\text{n-var. monomials of deg} \leq \ell + d - k} & \Big\} \, x^{\leq \ell} \partial^{=k} \end{array}$$

▷ Clearly, $\Gamma_{k,\ell}$ is <u>sub-additive</u>.

<u>Pf:</u> Derivation is an $\mathbb{F}$-linear operation. □

<u>Lemma 1:</u> Let f be an n-variate computed by a $\Sigma^{s} \Pi^{a} \Sigma \Pi^{b}$ circuit. Then,

$$\Gamma_{k,\ell}(f) \leq s \cdot \binom{a+k}{k} \cdot \binom{n + (b-1)k + \ell}{n}.$$

<u>Proof:</u>

- By subadditivity, it suffices to

consider a product gate $f = Q_1 \cdots Q_a$ with $\deg Q_i \leq b$.

- For a $\bar{\beta}$, $|\bar{\beta}| = k$, $\partial_{\bar{\beta}} := \partial_{x^{\bar{\beta}}}(Q_1 \cdots Q_a)$ can be expanded using the product rule of derivation.
- The number of summands there is $\leq \binom{a+k}{k}$.
- Now, by subadditivity, we reduce to cases of the type: $\partial_1 Q_1 \cdots \partial_k Q_k$.

in this summand we can ignore $Q_{k+1} \cdots Q_a$ due to $\binom{a+k}{k}$

$\Rightarrow$ after monomial multiplication we have products like $x^{\bar{\alpha}} \cdot \prod_{i \in [k]} \partial_i Q_i$, $|\bar{\alpha}| \leq \ell$.

- The number of monomials here is $\leq$
$$\binom{n + \deg}{n} \leq \binom{n + (b-1)k + \ell}{n}.$$

$$\Rightarrow \quad T_{k,\ell}(f) \leq \binom{a+k}{k} \cdot \binom{n + (b-1)k + \ell}{n}.$$

compare this with $\binom{n+k}{k} \cdot \binom{n+\ell}{\ell}$

$\square$

wrt RHS

- Thus, we want an f with a "large" $T_{k,\ell}(f)$ for some parameters k & ℓ.

- We will now lower bound $T_{k,\ell}$ for $\det_n$ (& similarly per_n).

<u>Lemma 2:</u> [Gupta, Kamath, Kayal, Saptharishi '14]:

$$T_{k,\ell}(\det_n) \geq \binom{n+k}{2k} \cdot \binom{n^2-2k+\ell}{\ell}.$$

<u>Proof:</u>

- Say $\det_n$ has variables x_{ij}, $i,j \in [n]$.
- Let us fix a monomial ordering as:

$$x_{11} > x_{12} > \dots > x_{1n} > \dots > x_{n1} > \dots > x_{nn}.$$

- Under this ordering we want to estimate the number of <u>leading</u> monomials in the polynomials in the set

$$\{ x^{\bar{\alpha}} \cdot \partial_{\bar{\beta}} \det_n \mid |\bar{\alpha}| \leq \ell, |\bar{\beta}| = k \}.$$

- Clearly, that estimate is a lower bound on $T_{k,\ell}(\det_n)$.

- Note that $\partial_{\bar{\beta}} \det_n$ is either zero or an $(n-k)$-minor of $\det_n$.

The leading monomial of this minor is merely the product of the variables in its principal diagonal.

leading monomial

nonzero

$\Rightarrow LM(\partial_{\bar{p}} \det_n) = x_{i_1 j_1} \cdots\cdots x_{i_{n-k} j_{n-k}}$

where $i_1 < \cdots < i_{n-k}$ & $j_1 < \cdots < j_{n-k}$.

- Let us call such indices an $(n-k)$-increasing sequence in $[n] \times [n]$.

▷ They are in bijection with $(n-k)$-minors.

$\Rightarrow \Gamma_{k,\ell}(\det_n) \geq$ # monomials of deg $\leq (n+\ell-k)$ that contain an $(n-k)$-increasing seq.

- To lower bound RHS we consider:

<u>Defn:</u> Let $D_2 := \{x_{11}, x_{22}, \ldots, x_{nn}\} \cup \{x_{12}, x_{23}, \ldots, x_{n-1,n}\}$ be the diagonal & the vars above it.

For monomial m define its

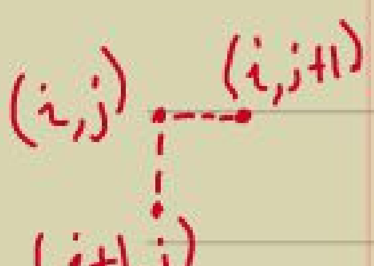
canonical increasing seq. $\chi(m)$ as the $(n-k)$-increasing seq. in m that is entirely contained in D_2 (& highest wrt $>$).

If the latter does not exist then define $\chi(m) := \phi$.

▷ Let S be an $(n-k)$-increasing seq. entirely contained in D_2 and m_S be its product. There are $\geqslant 2(n-k)-1$ variables in D_2 s.t. any monomial m in them satisfies:

$$\chi(m \cdot m_S) = \chi(m_S).$$

<u>Proof:</u>



- Note that for $(i,j) \neq (n,n)$, x_{ij} has a <u>companion</u> in D_2 of the type $x_{i+1,j}$ or $x_{i,j+1}$.
- Clearly, the variables in m_S, or their companions, do not alter $\chi(\cdot)$ when multiplied to m_S. $\square$

▷ # $(n-k)$-increasing sequences, contained in D_2, is $\binom{n+k}{2k}$.

<u>Proof:</u> • We want to pick $(n-k)$ elements from

$x_{11}\ x_{12}\ x_{22}\ x_{23} \cdots\cdots x_{n-1,n}\ x_{nn}$

in a way that no two <u>adjacent</u> elements are picked.

• Consider the remaining $(2n-1) - (n-k) = n+k-1$ elements.

• Associate them with a string of $(n+k-1)$ 1's.

or at the two ends • We want to choose $(n-k)$ places in the <u>middle</u> of these 1's.

$$\Rightarrow \# \text{such choices} = \binom{(n+k-1)+1}{n-k}$$

$$= \binom{n+k}{n-k}.$$

□

$\chi(\cdot)$

• Note that this type of $(n-k)$-increasing sequence does not change if we multiply by $|X \setminus D_2| = (n^2-2n+1)$ many variables.

Moreover, we can multiply by at least $2(n-k)-1$ variables in D_2 without changing $\chi(\cdot)$.

Note: $m m_S = m' m_{S'}$ $\Rightarrow \chi(m m_S) = \chi(m' m_{S'})$ $\Leftrightarrow S = S'$.

$\Rightarrow$ We get the following lower bound on the number of distinct leading monomials in $\{x^{\bar{\alpha}} \cdot \partial_{\bar{\beta}} \det_n \mid |\bar{\alpha}| \leq \ell, |\bar{\beta}| = k\}$:

$$\binom{n+k}{2k} \cdot \binom{n^2-2n+1+2(n-k)-1+\ell}{\ell}$$

$$= \binom{n+k}{2k} \cdot \binom{n^2-2k+\ell}{\ell}.$$

$\square$

- Now we have upper bounded $T_{k,\ell}$ for $\Sigma^s \Pi^a \Sigma \Pi^b$ & lower bounded for $\det_n$.

It is time to compare the two.

c is a constant

- For the applications $a = cn/b$ is of interest.
- For technical reasons, we use $k = \varepsilon n/b$ & $\ell = n^2 b$ (small enough constant $\varepsilon > 0$).

- By the two lemmas we get:

$$s \geqslant \binom{n+k}{2k} \cdot \binom{n^2-2k+\ell}{\ell} \Big/ \binom{cn/b+k}{k} \cdot \binom{n^2+(b-1)k+\ell}{n^2}.$$

<u>Claim 1:</u> $\ln \binom{n+k}{2k} = 2\varepsilon \frac{n}{b}\left(\ln \frac{b}{2\varepsilon} + 1\right) \pm O(n/b^2)$.

<u>Claim 2:</u> $\ln \binom{n^2-2k+\ell}{\ell} \Big/ \binom{n^2+(b-1)k+\ell}{n^2} = -2\varepsilon \frac{n}{b}\left(\ln b + \frac{1}{2}\right) \pm O(1)$

<u>Claim 3:</u> $\ln \binom{cn/b+k}{k} = (c+\varepsilon)\cdot \frac{n}{b} \cdot H_e\left(\frac{\varepsilon}{c+\varepsilon}\right) - O(\ln n)$.

- These claims, after some calculations, imply:

$$\ln s \geqslant -\varepsilon \cdot \ln(4\varepsilon(c+\varepsilon)) \cdot n/b \pm O(n/b^2)$$
$$= \Omega(n/b), \text{ for small } \varepsilon.$$

- The claims could be proved using the following binomial estimates:

$\underline{h^{f+g}}$

$$\ln \frac{(h+f)!}{(h-g)!} = (f+g)\ln h \pm O\left(\frac{(f+g)^2}{h}\right), \text{ if } f+g = o(h),$$

& $\ln\binom{\alpha n}{\beta n} = \alpha n \cdot H_e(\beta/\alpha) - O(\ln n)$,

for constants $\alpha \geq \beta > 0$.

- The proofs are left as exercises.

- This completes the proof of:

<u>Theorem</u> [GKKS'14]: Any $\Sigma^s \Pi^{O(n/b)} \Sigma \Pi^b$ circuit computing $\det_n$ or per_n requires $s = \exp(\Omega(n/b))$.

- For $b = \sqrt{n}$, this shows that the depth reduction to depth-4 is <u>almost optimal</u>. ($\because$ $\det_n$ has such a circuit of size $n^{O(\sqrt{n})}$.)

This was further clarified by:

<u>Thm</u> [Fournier, Limaye, Malod, Srinivasan '14]: For a small $\delta > 0$ & $d \leq n^\delta$, any $\Sigma^s \Pi^{O(\sqrt{d})} \Sigma \Pi^{\sqrt{d}}$ circuit computing $IMM_{n,d}$ has $s = n^{\Omega(\sqrt{d})}$.

optimal

Homogeneous depth-4

- Homogeneity is a restriction for constant-depth circuits.

  (Not so for general circuits.)

- If a homogeneous $\Sigma\Pi^a\Sigma\Pi^b$ computes a degree d polynomial f, then we get the degree restriction $a, b \le d$.

  Can this be used in shifted partials?

Defn: In a homogeneous depth-4 circuit $f(x_1,..,x_n) = \sum_{i\in[s]} Q_{i1}\cdots\cdots Q_{ia_i}$, each Q_{ij} is a homogeneous sparse poly.

& $\sum_{j\in[a_i]} \deg Q_{ij} = \deg f$, $\forall i \in [s]$.

($\Rightarrow f$ is homogeneous too.)

- In homogeneous $\Sigma\Pi^a\Sigma\Pi^b$, b can be as high as the degree d of a polynomial f.

  So, we need to utilize the sparsity of the Q_{ij}'s.

- We will show, using random restrictions, that Q_{ij}'s can be "reduced" to a sum of $\sqrt{d}$-support-monomials.
low-support

Lemma: Let f be an n-variate d-deg polynomial computable by a size $s \le n^{c\sqrt{d}}$ (constant $c>0$) homogeneous depth-4 C. Let ρ be a random restriction that sets each variable to 0 with probability $1-n^{-2c}$.

Then, with prob $\ge 1-\frac{1}{s}$, the polynomial $\rho(f)$ is computable by a homogeneous depth-4 C' with bottom support $\le \sqrt{d}$ & size $\le s$.

Proof:

- Among all Q_{ij} consider the monomials $\{m_1, \ldots, m_r\}$ that have support $> \sqrt{d}$. Clearly, $r \le s$.

$$\forall i \in [r],\ \Pr[\rho(m_i) \ne 0] < (n^{-2c})^{\sqrt{d}}$$

$$\Rightarrow \Pr[\exists i,\ \rho(m_i) \ne 0] < r n^{-2c\sqrt{d}} \le 1/s.$$

$\Rightarrow$ With prob $> 1-\frac{1}{s}$ all the large support monomials vanish. $\square$

- Now, we need to find a measure that is "small" for such $\Sigma\Pi\Sigma\Pi$.

Since we will prove a lower bound for a multilinear f, we can pick a measure that ignores the non-multilinear monomials.

<u>Defn:</u> For any $k, \ell \in \mathbb{N}$ & polynomial $f(\bar{x})$, define <u>projected shifted partials</u> $PSP_{k,\ell}(f)$ as the $\mathbb{F}$-span of the set of polynomials:

$$\{ mult(m_1 \cdot \partial_{m_2} f) \mid \deg m_1 = \ell,\ \deg m_2 = k \ \& \ m_1, m_2 \text{ are multilinear monomials} \}$$

where <u>mult(·)</u> refers to the projection to the multilinear part (eg. remainder modulo $\langle x_1^2, \ldots, x_n^2 \rangle$).

The measure $T^{PSP}_{k,\ell}(f)$ is the dimension of $PSP_{k,\ell}(f)$.

Lemma 1 (Upper bd.): Let f be an n-variate d-degree polynomial computed by a homogeneous $\Sigma\Pi\Sigma\Pi$ of bottom-support $\leq r$ & size $\leq s$. Then, for any k, ℓ with $\ell + 4rk \leq \frac{n}{2}$ we have

$$\Gamma^{PSP}_{k,\ell}(f) \leq s \cdot \binom{d/r + k}{k} \cdot \binom{n}{\ell + 4rk}.$$

Proof:

- Consider a product gate $Q_{i1} \cdots Q_{ia}$.
- We could assume that the individual deg of any variable in Q_{ij} is ≤ 2.

  Otherwise, there is a monomial say x_1^3 which can never contribute to the polynomials $\mathrm{mult}(m_1 \partial_{m_2} f)$, as multilinear $m_2 \Rightarrow \partial_{m_2}(x_1^3)$ is non-multilr.

- Also, by multiplying out the Q_{ij}'s if needed, we can assume that $\deg Q_{ij} \in [r, 4r]$.
- Thus, we reduce to the case of $\Sigma\Pi^a\Sigma\Pi^b$, $a \leq d/r$ & $b \leq 4r$.

- Further, by using the multilinearity restrictions in the definition of $PSP_{k,\ell}(f)$, we get the upper bound. $\square$

- The lower bound of the measure is trickier.

  Because to get a result for a polynomial f one has to prove a measure lower bound for the various projections of f (under random restrictions ρ).

- Currently, such results are known for two types of polynomials:

Defn: • [Iterated matrix multiplication polynomial]

$IMM_{n,d}(\bar{x}) := (M_1 \cdots M_d)_{1,1}$

$n^2 d$-variate d-degree

where, $M_k = (x_{k,ij} \mid i,j \in [n])$ for $k \in [d]$.

• [Nisan-Wigderson polynomial] Let $\mathbb{F}_m$ be the finite field with m elements (identified with the elements $1,2,\ldots,m$).

$\forall\, 0 \le k \le n$, $NW_{n,m,k}(x_{11},\ldots,x_{nm}) :=$

nm-variate
n-degree

$$\sum_{\substack{p(t)\in \mathbb{F}_m[t] \\ \deg p \le k}} x_{1,p(1)} \cdots\cdots x_{n,p(n)}$$

▷ $IMM_{n,d} \in VP$. (<u>OPEN</u>: $NW_{h,m,k} \overset{\in VNP.}{\in} VP$?)

<u>Thm</u> [KLSS'14]: Over char <u>zero</u>, the homogeneous depth4 complexity of $NW_{d,d^3,d/3}$ is $d^{\Omega(\sqrt{d})}$.

It's in VNP.

<u>Thm</u> [KS'14]: The above holds for all $\mathbb{F}$.

It's in VP. Further, $IMM_{n,d}$ has homogeneous depth4 complexity $d^{\Omega(\sqrt{d})}$.

▷ Further, generalized to $\Sigma\Gamma\Sigma\Pi$ by PSS'16.

– Proofs are left as reading exercises (from [Saptharishi '16]).

Limitations of measures?

- Our lower bound proofs were all rank-based.

- In other words, we design a determinant based polynomial $M(\cdot)$ that takes as input — the coefficient-vector of f.

technically, family of f

- To show $f \in \mathcal{D} \setminus \mathcal{C}$, for algebraic complexity classes $\mathcal{C}$ & $\mathcal{D}$, we show:
  1) $\forall g \in \mathcal{C}$, $M(g) = 0$,
  2) $M(f) \neq 0$, &
  3) $f \in \mathcal{D}$.

  $\Rightarrow f \notin \mathcal{C}$

[FSV'17] If $\exists g \in \mathcal{C}$ s.t. coefficient-vector of g is a non-root of determinant,

g hits det

then "$f \notin \mathcal{C}$" cannot be shown by rank-based measures!