Computing Multilinear Polynomials by Arithmetic Circuits of Bounded Individual Degree

Suryajith Chillara[1]

אוניברסיטת חיפה
University of Haifa
جامعة حيفا

Technion Theory Lunch
28.10.2020.

[1]Research supported by PBC post doctoral fellowship from Israeli Council of Higher Education.

Strassen's matrix multiplication [Strassen, 1969]

$$\begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \times \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix}$$

Strassen's matrix multiplication [Strassen, 1969]

$$\begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{bmatrix} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \times \begin{bmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{bmatrix}$$

$$\begin{aligned}
M_1 &= (A_{1,1} + A_{2,2}) \times (B_{1,1} + B_{2,2}) \\
M_2 &= (A_{2,1} + A_{2,2}) \times B_{1,1} \\
M_3 &= A_{1,1} \times (B_{1,1} - B_{2,2}) \\
M_4 &= A_{2,2} \times (B_{2,1} - B_{1,1}) \\
M_5 &= (A_{1,1} + A_{1,2}) \times B_{1,2} \\
M_6 &= (A_{2,1} - A_{1,2}) \times (B_{1,1} + B_{1,2}) \\
M_7 &= (A_{1,2} - A_{2,2}) \times (B_{2,1} + B_{2,2})
\end{aligned}$$

$$\begin{aligned}
C_{1,1} &= M_1 + M_4 - M_5 + M_7 \\
C_{1,2} &= M_3 + M_5 \\
C_{2,1} &= M_2 + M_4 \\
C_{2,2} &= M_1 - M_2 + M_3 + M_6
\end{aligned}$$

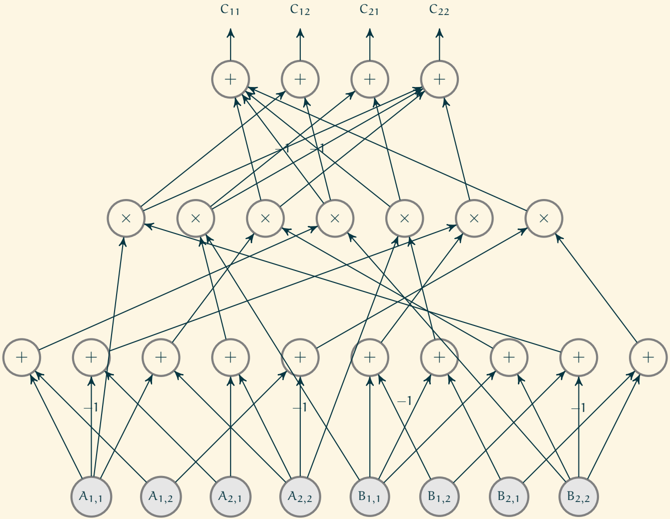


Figure: Strassen's algorithm for multiplication of two 2×2 matrices.

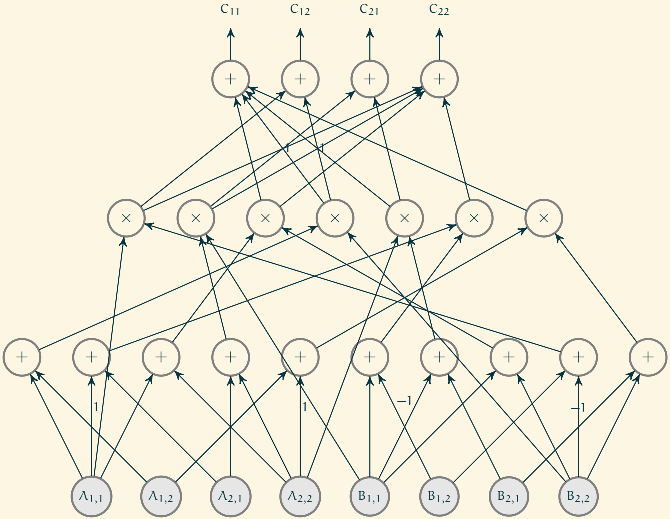


Figure: Strassen's algorithm for multiplication of two 2×2 matrices.

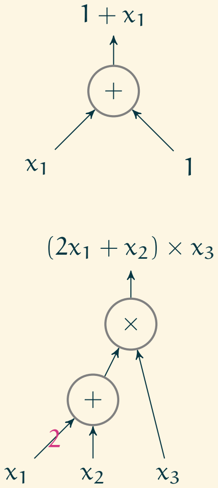
This DAG can be thought of as a "hardwired circuit" for 2×2 matrix multiplication.

Computing polynomials

Definition

An Arithmetic Circuit Φ over the field $\mathbb{F}$ and the set of variables $X = (x_1, x_2, \ldots, x_n)$ is a *directed acyclic graph* as follows:

- Leaf nodes are labelled either by a variable or a field element from $\mathbb{F}$ and the root node outputs the polynomial.
- Every other node is labelled by either $\times$ or $+$.
- The size of Φ is the number of nodes present in it.
- The depth of Φ is the length of the longest leaf to root path.

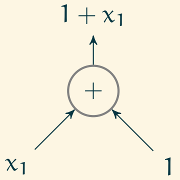


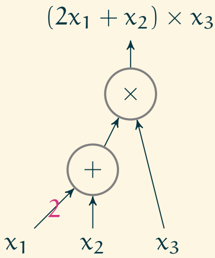
Computing polynomials

Definition

An Arithmetic Circuit Φ over the field $\mathbb{F}$ and the set of variables $X = (x_1, x_2, \ldots, x_n)$ is a *directed acyclic graph* as follows:

- Leaf nodes are labelled either by a variable or a field element from $\mathbb{F}$ and the root node outputs the polynomial.
- Every other node is labelled by either $\times$ or $+$.
- The size of Φ is the number of nodes present in it.
- The depth of Φ is the length of the longest leaf to root path.





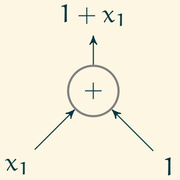
Formulas are circuits whose underlying graph is a tree.

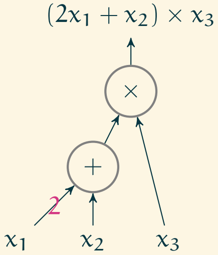
Computing polynomials

Definition

An Arithmetic Circuit Φ over the field $\mathbb{F}$ and the set of variables $X = (x_1, x_2, \ldots, x_n)$ is a *directed acyclic graph* as follows:

- Leaf nodes are labelled either by a variable or a field element from $\mathbb{F}$ and the root node outputs the polynomial.
- Every other node is labelled by either $\times$ or $+$.
- The size of Φ is the number of nodes present in it.
- The depth of Φ is the length of the longest leaf to root path.





Formulas are circuits whose underlying graph is a tree.

W.L.O.G we assume arithmetic circuits to be layered: $\Sigma\Pi \cdots \Sigma\Pi$.

Significance of size and depth

- Small circuit size $\implies$ efficient algorithms.

Significance of size and depth

- Small circuit size $\implies$ efficient algorithms.
- Small circuit depth $\implies$ efficient parallel algorithms.

Algebraic P vs Algebraic NP [Valiant, 1979]

Definition (Algebraic P/p-computable/VP)

Class VP consists of all polynomial families $\{f_n\}_{n \geqslant 0}$ of degree $n^{O(1)}$ which can be computed by $n^{O(1)}$ sized arithmetic circuits.

Algebraic P vs Algebraic NP [Valiant, 1979]

Definition (Algebraic P/p-computable/VP)

Class VP consists of all polynomial families $\{f_n\}_{n \geqslant 0}$ of degree $n^{O(1)}$ which can be computed by $n^{O(1)}$ sized arithmetic circuits.

Determinant is a "canonical" polynomial for VP.

Algebraic P vs Algebraic NP [Valiant, 1979]

Definition (Algebraic P/p-computable/VP)

Class VP consists of all polynomial families $\{f_n\}_{n \geqslant 0}$ of degree $n^{O(1)}$ which can be computed by $n^{O(1)}$ sized arithmetic circuits.

Determinant is a "canonical" polynomial for VP.

Definition (Algebraic NP/p-definable/VNP)

Class VNP consists of all polynomial families $\{F_n\}_{n \geqslant 0}$ of degree $n^{O(1)}$ which can be expressed as follows.

$$F_n(X) = \sum_{\mathbf{e} \in \{0,1\}^{m(n)}} g_{n,m(n)}(X, \mathbf{e})$$

where $g_{n,m(n)}$ is a polynomial in VP.

Algebraic P vs Algebraic NP [Valiant, 1979]

Definition (Algebraic P/p-computable/VP)

Class VP consists of all polynomial families $\{f_n\}_{n \geqslant 0}$ of degree $n^{O(1)}$ which can be computed by $n^{O(1)}$ sized arithmetic circuits.

Determinant is a "canonical" polynomial for VP.

Definition (Algebraic NP/p-definable/VNP)

Class VNP consists of all polynomial families $\{F_n\}_{n \geqslant 0}$ of degree $n^{O(1)}$ which can be expressed as follows.

$$F_n(X) = \sum_{\mathbf{e} \in \{0,1\}^{m(n)}} g_{n,m(n)}(X, \mathbf{e})$$

where $g_{n,m(n)}$ is a polynomial in VP.

Permanent is a "canonical" polynomial for VNP.

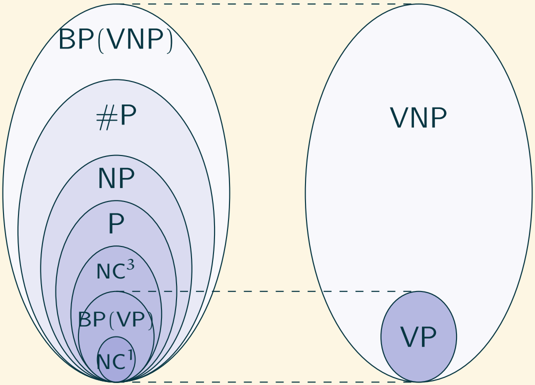
Valiant's hypothesis [Valiant, 1979]

Hypothesis

$$\mathsf{VP} \neq \mathsf{VNP}.$$

That is, Permanent of a generic $n \times n$ matrix cannot be computed by $\mathrm{poly}(n)$-sized arithmetic circuits.

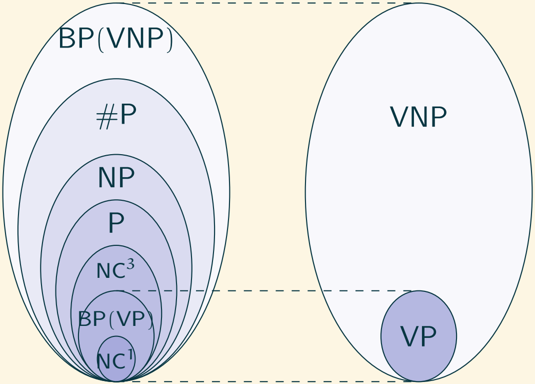
Cook's vs Valiant's hypotheses [Bürgisser, 2000]



*Not to scale.

Given a polynomial f, we can assign a corresponding Boolean function $\mathrm{BP}(f)$ to it such that f and $\mathrm{BP}(f)$ agree on evaluations over $\{0,1\}^N$.

Cook's vs Valiant's hypotheses [Bürgisser, 2000]



*Not to scale.

Given a polynomial f, we can assign a corresponding Boolean function $\mathrm{BP}(f)$ to it such that f and $\mathrm{BP}(f)$ agree on evaluations over $\{0,1\}^N$.

$\mathsf{VP} \neq \mathsf{VNP}$ can be thought of as a "coarser" separation than $\mathsf{P} \neq \mathsf{NP}$.

Cook's vs Valiant's hypotheses

Theorem [Bürgisser, 2000]

(GRH): If $\mathsf{VP} = \mathsf{VNP}$ then non-uniform $\#\mathsf{P} \subseteq$ non-uniform NC^3.

Cook's vs Valiant's hypotheses

Theorem [Bürgisser, 2000]

(GRH): If $\mathsf{VP} = \mathsf{VNP}$ then non-uniform P equals non-uniform NP.

Cook's vs Valiant's hypotheses

Theorem [Bürgisser, 2000]

(GRH): If $\mathsf{VP} = \mathsf{VNP}$ then non-uniform P equals non-uniform NP.

- $\mathsf{PH} \subseteq \Sigma_2^p$ [Karp and Lipton, 1980],
- $\mathsf{AM} = \mathsf{MA}$ [Arvind, Köbler, Schöning, and Schuler, 1995].

Cook's vs Valiant's hypotheses

Theorem [Bürgisser, 2000]

(GRH): If $\mathsf{VP} = \mathsf{VNP}$ then non-uniform P equals non-uniform NP.

- $\mathsf{PH} \subseteq \Sigma_2^p$ [Karp and Lipton, 1980],
- $\mathsf{AM} = \mathsf{MA}$ [Arvind, Köbler, Schöning, and Schuler, 1995].

Valiant's observations [Valiant, 1992]

- "Since the set of valid algebraic identities in the algebraic model form a proper subset of those in the Boolean setting, lower bound proof for the algebraic setting should be strictly easier."
- "In particular, the main power of the algebraic model derives from the possibility of cancellations."
- Example: Samuelson-Berkowitz method for computing the determinant.

Best known general circuit bounds

- Best known circuit size lower bound is $\Omega(N \log N)$ for a Power Symmetric polynomial [Baur and Strassen, 1983].

- Best known formula size lower bound is $\Omega(N^2)$ for a very simple polynomial [Kalorkoti, 1985].

Best known general circuit bounds

- Best known circuit size lower bound is $\Omega(N \log N)$ for a Power Symmetric polynomial [Baur and Strassen, 1983].
- Best known formula size lower bound is $\Omega(N^2)$ for a very simple polynomial [Kalorkoti, 1985].

Strategy: Prove lower bounds against restricted models and then extend the understanding to the general setting.

A Restricted Model

Formal degree

Formal degree of a circuit represents what the degree of the output would have been if there were no cancellations and is an upper bound on the degree of the output.

Formal degree

Formal degree of a circuit represents what the degree of the output would have been if there were no cancellations and is an upper bound on the degree of the output.

For an arithmetic circuit C and for all $x_i \in X$,

- Formal degree of a leaf node w with respect to x_i,

$$\mathrm{fdeg}_{x_i}(w) = \begin{cases} 1 & \text{if } w \text{ is labelled by variable } x_i, \\ 0 & \text{otherwise.} \end{cases}$$

Formal degree

Formal degree of a circuit represents what the degree of the output would have been if there were no cancellations and is an upper bound on the degree of the output.

For an arithmetic circuit C and for all $x_i \in X$,

- Formal degree of a leaf node w with respect to x_i,

$$\mathrm{fdeg}_{x_i}(w) = \begin{cases} 1 & \text{if } w \text{ is labelled by variable } x_i, \\ 0 & \text{otherwise.} \end{cases}$$

- Formal degree of a sum node u with inputs $u_1, \ldots, u_k$, with respect to x_i,

$$\mathrm{fdeg}_{x_i}(u) = \max_{j \in [k]} \left\{ \mathrm{fdeg}_{x_i}(u_j) \right\}.$$

Formal degree

Formal degree of a circuit represents what the degree of the output would have been if there were no cancellations and is an upper bound on the degree of the output.

For an arithmetic circuit C and for all $x_i \in X$,

- Formal degree of a leaf node w with respect to x_i,

$$\mathrm{fdeg}_{x_i}(w) = \begin{cases} 1 & \text{if } w \text{ is labelled by variable } x_i, \\ 0 & \text{otherwise.} \end{cases}$$

- Formal degree of a sum node u with inputs $u_1, \ldots, u_k$, with respect to x_i,

$$\mathrm{fdeg}_{x_i}(u) = \max_{j \in [k]} \left\{ \mathrm{fdeg}_{x_i}(u_j) \right\}.$$

- Formal degree of a product node v with inputs $v_1, \ldots, v_k$, with respect to x_i,

$$\mathrm{fdeg}_{x_i}(v) = \sum_{j \in [k]} \mathrm{fdeg}_{x_i}(v_j).$$

Multi-r-ic circuits

Definition

An arithmetic circuit C is said to be syntactically multi-r-ic if the formal degree of the output node is at most r with respect each of its variables.

Multi-r-ic circuits

Definition

An arithmetic circuit C is said to be syntactically multi-r-ic if the formal degree of the output node is at most r with respect each of its variables.

When $r = 1$, we have multilinear circuits. We could start proving results for $r = 1$ and then extend these to the setting where $r > 1$.

Lower bounds for syntactically multilinear circuits

- Formulas: $N^{\Omega(\log N)}$ [Raz, 2006; Raz and Yehudayoff, 2008; Dvir, Malod, Perifel, and Yehudayoff, 2012].

- Bounded depth formulas:
  - $2^{\Omega(N^{1/\Delta})}$ [Raz and Yehudayoff, 2009],
  - $2^{\Omega(\Delta N^{1/\Delta})}$ [Chillara, Limaye, and Srinivasan, 2019].

- Circuits:
  - $\Omega\left(N^{1.33}/\log^2 N\right)$ [Raz, Shpilka, and Yehudayoff, 2008],
  - $\Omega\left(N^2/\log^2 N\right)$ [Alon, Kumar, and Volk, 2020].

- Depth four circuits: $N^{\Omega\left(\sqrt{\frac{N}{\log N}}\right)}$ [Raz and Yehudayoff, 2009].

Separations for multilinear circuits

- **Limits of parallelization:** Depth reduction shown by [Brent, 1974] to $O(\log s)$ depth is optimal for multilinear formulas [Chillara, Limaye, and Srinivasan, 2019].

Separations for multilinear circuits

- **Limits of parallelization:** Depth reduction shown by [Brent, 1974] to $O(\log s)$ depth is optimal for multilinear formulas [Chillara, Limaye, and Srinivasan, 2019].
- **Circuits vs formulas:** Circuits are more powerful than formulas [Raz, 2006]. For all small Δ, circuits of product-depth at most Δ are more powerful than the formulas of product-depth Δ [Chillara, Limaye, and Srinivasan, 2019].

Separations for multilinear circuits

- **Limits of parallelization:** Depth reduction shown by [Brent, 1974] to $O(\log s)$ depth is optimal for multilinear formulas [Chillara, Limaye, and Srinivasan, 2019].
- **Circuits vs formulas:** Circuits are more powerful than formulas [Raz, 2006]. For all small Δ, circuits of product-depth at most Δ are more powerful than the formulas of product-depth Δ [Chillara, Limaye, and Srinivasan, 2019].
- **Branching programs vs formulas:** Algebraic branching programs are more powerful than formulas [Dvir, Malod, Perifel, and Yehudayoff, 2012].

Separations for multilinear circuits

- **Limits of parallelization:** Depth reduction shown by [Brent, 1974] to $O(\log s)$ depth is optimal for multilinear formulas [Chillara, Limaye, and Srinivasan, 2019].
- **Circuits vs formulas:** Circuits are more powerful than formulas [Raz, 2006]. For all small Δ, circuits of product-depth at most Δ are more powerful than the formulas of product-depth Δ [Chillara, Limaye, and Srinivasan, 2019].
- **Branching programs vs formulas:** Algebraic branching programs are more powerful than formulas [Dvir, Malod, Perifel, and Yehudayoff, 2012].
- **Separation from general formulas:** Over large fields, general formulas of product-depth $\Delta = o(\log s)$ are more powerful than multilinear formulas of product-depth Δ [Chillara, Limaye, and Srinivasan, 2019].

Hierarchies for multilinear circuits

- **Depth Hierarchy:** Formulas of product-depth Δ are exponentially more powerful than those of product-depth $\Delta - 1$ [Raz and Yehudayoff, 2009; Chillara, Engels, Limaye, and Srinivasan, 2018a].

- **Size Hierarchy:** Formulas of size s are more powerful than the small depth formulas at size $\sqrt{s}$ [Chillara, Limaye, and Srinivasan, 2018b].

Lower bounds for syntactically multi-r-ic circuits

- Homogeneous formulas: $N^{\Omega(\log N)}$ [Kayal, Saha, and Tavenas, 2018].
- Constant Depth Homogeneous Formulas:
  - $2^{\Omega\left(\frac{1}{r}\cdot\left(\frac{N}{2}\right)^{1/\Delta}\right)}$ [Kayal, Saha, and Tavenas, 2018],
  - $2^{\Omega\left(\frac{\Delta}{r}\cdot\left(\frac{Nr}{2}\right)^{1/\Delta}\right)}$ [Chillara, 2019].
- Depth four:
  - Multilinear polynomial: $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$ where $N = n^2 d$ [Kayal, Saha, and Tavenas, 2018].
  - Multi-r-ic polynomials: For $r = o(N)$,
    - $2^{\Omega(\sqrt{N})}$ [Kayal, Saha, and Tavenas, 2018],
    - $\exp\left(\Omega\left(\sqrt{\frac{N \log N}{r}}\right)\right)$ [Hegde and Saha, 2017].

Depth four multi-r-ic circuits

Definition

A depth four circuit C computes the polynomials of the form

$$f(x_1, \ldots, x_N) = \sum_{i=1}^{s} T_i = \sum_{i=1}^{s} \prod_{j=1}^{d_i} Q_{i,j}(x_1, \ldots, x_N).$$

A depth four circuit C is said to be syntactically multi-r-ic if the formal degree of its output node, with respect to each of its variables is at most r.

Depth four multi-r-ic circuits

Definition

A depth four circuit C computes the polynomials of the form

$$f(x_1, \ldots, x_N) = \sum_{i=1}^{s} T_i = \sum_{i=1}^{s} \prod_{j=1}^{d_i} Q_{i,j}(x_1, \ldots, x_N).$$

A depth four circuit C is said to be syntactically multi-r-ic if the formal degree of its output node, with respect to each of its variables is at most r.

For every $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$ ($i \in [s]$),

- Each variable can appear in at most r many $Q_{i,j}$'s in T_i.
-
$$\forall k \in [N], \sum_{j \in [d_i]} \deg_{x_k}(Q_{i,j}) \leqslant r.$$

A motivation to study depth four circuits

Chasm at depth four

Strong lower bounds against *restricted* depth four circuits imply strong lower bounds against general arithmetic circuits.

- $2^{\omega(\sqrt{d}\log N)}$ against bounded fan-in depth four circuits [Agrawal and Vinay, 2008; Koiran, 2012; Tavenas, 2015],
- $2^{\omega\left(\sqrt{rN\log N}\right)}$ against multi-r-ic depth four circuits [Kumar, de Oliveira, and Saptharishi, 2019].

Previous work for multi-r-ic depth four circuits

Theorem [Kayal, Saha, and Tavenas, 2018]

There exists a fixed constant ν and an explicit multilinear polynomial $Q_{n,d}$ (over $\mathrm{poly}(n, d)$ many variables and degree d) such that for all $d \in \left[\log^2 n, n^{\nu}\right]$ any *syntactically* multi-r-ic depth four circuit computing it must have size $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$.

We shall first define the explicit polynomial.

Iterated matrix multiplication polynomial

The iterated matrix multiplication polynomial is the $(1,1)$th entry of product of d many generic $n \times n$ matrices $X_1, X_2, \ldots, X_d$ over disjoint set of variables.

$$\mathsf{IMM}_{n,d}(X) = \sum_{i_1, i_2, \ldots, i_{d-1} \in [n]} x^{(1)}_{(1,i_1)} x^{(2)}_{(i_1,i_2)} \ldots x^{(d-1)}_{(i_{(d-2)}, i_{(d-1)})} x^{(d)}_{(i_{(d-1)},1)}$$

where $x^{(k)}_{(i,j)}$ is the variable in X_k indexed by $(i,j) \in [n] \times [n]$.

$$\mathsf{IMM}_{n,d}(X) = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x^{(1)}_{1,1} & \cdots & x^{(1)}_{1,n} \\ \vdots & \ddots & \vdots \\ x^{(1)}_{n,1} & \cdots & x^{(1)}_{n,n} \end{bmatrix} \cdots \begin{bmatrix} x^{(d)}_{1,1} & \cdots & x^{(d)}_{1,n} \\ \vdots & \ddots & \vdots \\ x^{(d)}_{n,1} & \cdots & x^{(d)}_{n,n} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Iterated matrix multiplication polynomial

The iterated matrix multiplication polynomial is the $(1,1)$th entry of product of d many generic $n \times n$ matrices $X_1, X_2, \ldots, X_d$ over disjoint set of variables.

$$\mathsf{IMM}_{n,d}(X) = \sum_{i_1, i_2, \ldots, i_{d-1} \in [n]} x^{(1)}_{(1,i_1)} x^{(2)}_{(i_1,i_2)} \ldots x^{(d-1)}_{(i_{(d-2)}, i_{(d-1)})} x^{(d)}_{(i_{(d-1)},1)}$$

where $x^{(k)}_{(i,j)}$ is the variable in X_k indexed by $(i,j) \in [n] \times [n]$.

$$\mathsf{IMM}_{n,d}(X) = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x^{(1)}_{1,1} & \cdots & x^{(1)}_{1,n} \\ \vdots & \ddots & \vdots \\ x^{(1)}_{n,1} & \cdots & x^{(1)}_{n,n} \end{bmatrix} \cdots \begin{bmatrix} x^{(d)}_{1,1} & \cdots & x^{(d)}_{1,n} \\ \vdots & \ddots & \vdots \\ x^{(d)}_{n,1} & \cdots & x^{(d)}_{n,n} \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Iterated Matrix Multiplication polynomial $\mathsf{IMM}_{n,d}$ can be expressed in terms of Det_{nd} and thus a lower bound for $\mathsf{IMM}_{n,d}$ implies a lower bound for Det_{nd}.

Previous work for multi-r-ic depth four circuits

Theorem [Kayal, Saha, and Tavenas, 2018]

There exists a fixed constant ν such that for all $d \in \left[\log^2 n, n^{\nu}\right]$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must have size $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$.

Previous work for multi-r-ic depth four circuits

Theorem [Kayal, Saha, and Tavenas, 2018]

There exists a fixed constant ν such that for all $d \in \left[\log^2 n, n^{\nu}\right]$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must have size $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$.

- With increasing r, the lower bound deteriorates.

Previous work for multi-r-ic depth four circuits

Theorem [Kayal, Saha, and Tavenas, 2018]

There exists a fixed constant ν such that for all $d \in \left[\log^2 n, n^{\nu}\right]$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must have size $\left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$.

- With increasing r, the lower bound deteriorates.
- Lower bound only holds for r that is $o(d)$.

Attempt 1

Theorem [Chillara, 2020a]

There exists a constant $\eta \in (0, 1)$ such that for all $r \leqslant n^{\eta}$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,\Theta(\log^2 n)}$ must have size $n^{\Omega(\log n)}$.

Attempt 1

Theorem [Chillara, 2020a]

There exists a constant $\eta \in (0, 1)$ such that for all $r \leqslant n^{\eta}$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,\Theta(\log^2 n)}$ must have size $n^{\Omega(\log n)}$.

- For the setting of $d = \Theta(\log^2 n)$, [Kayal, Saha, and Tavenas, 2018] gives a lower bound of $n^{\Omega\left(\frac{\log n}{\sqrt{r}}\right)}$ and this is super polynomial only when $r = o(\log^2 n)$.

Attempt 1

Theorem [Chillara, 2020a]

There exists a constant $\eta \in (0, 1)$ such that for all $r \leqslant n^{\eta}$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,\Theta(\log^2 n)}$ must have size $n^{\Omega(\log n)}$.

- For the setting of $d = \Theta(\log^2 n)$, [Kayal, Saha, and Tavenas, 2018] gives a lower bound of $n^{\Omega\left(\frac{\log n}{\sqrt{r}}\right)}$ and this is super polynomial only when $r = o(\log^2 n)$.
- Their lower bound deteriorates when r gets closer to $\log^2 n$.

Attempt 1

Theorem [Chillara, 2020a]

There exists a constant $\eta \in (0, 1)$ such that for all $r \leqslant n^{\eta}$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,\Theta(\log^2 n)}$ must have size $n^{\Omega(\log n)}$.

- For the setting of $d = \Theta(\log^2 n)$, [Kayal, Saha, and Tavenas, 2018] gives a lower bound of $n^{\Omega\left(\frac{\log n}{\sqrt{r}}\right)}$ and this is super polynomial only when $r = o(\log^2 n)$.
- Their lower bound deteriorates when r gets closer to $\log^2 n$.
- We give a bound that does not change with increasing r but holds only for degrees that are $\Theta(\log^2 n)$.

Attempt 1

Theorem [Chillara, 2020a]

There exists a constant $\eta \in (0, 1)$ such that for all $r \leqslant n^{\eta}$, any *syntactically* multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,\Theta(\log^2 n)}$ must have size $n^{\Omega(\log n)}$.

- For the setting of $d = \Theta(\log^2 n)$, [Kayal, Saha, and Tavenas, 2018] gives a lower bound of $n^{\Omega\left(\frac{\log n}{\sqrt{r}}\right)}$ and this is super polynomial only when $r = o(\log^2 n)$.
- Their lower bound deteriorates when r gets closer to $\log^2 n$.
- We give a bound that does not change with increasing r but holds only for degrees that are $\Theta(\log^2 n)$.
- Our bound holds for a value of r that is much larger than d.

Attempt 2

Theorem [Chillara, 2020b]

There exist constants $a \leqslant b \in (0, 1)$ such that for all $d \leqslant n^a$ and $r \leqslant n^b$, any syntactically multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must be of size $n^{\Omega(\sqrt{d})}$.

Attempt 2

Theorem [Chillara, 2020b]

There exist constants $a \leqslant b \in (0,1)$ such that for all $d \leqslant n^a$ and $r \leqslant n^b$, any syntactically multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must be of size $n^{\Omega(\sqrt{d})}$.

- Extends [Chillara, 2020a] to give lower bounds that do not deteriorate with increasing values of r, for a wider range of d.

Attempt 2

Theorem [Chillara, 2020b]

There exist constants $a \leqslant b \in (0, 1)$ such that for all $d \leqslant n^a$ and $r \leqslant n^b$, any syntactically multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must be of size $n^{\Omega(\sqrt{d})}$.

- Extends [Chillara, 2020a] to give lower bounds that do not deteriorate with increasing values of r, for a wider range of d.
- Though [Kayal, Saha, and Tavenas, 2018] give a lower bound that holds for a "slightly larger" range of d, our lower bound is quantitively better in comparable range of d.

Attempt 2

Theorem [Chillara, 2020b]

There exist constants $a \leqslant b \in (0,1)$ such that for all $d \leqslant n^a$ and $r \leqslant n^b$, any syntactically multi-r-ic depth four circuit computing $\mathsf{IMM}_{n,d}$ must be of size $n^{\Omega(\sqrt{d})}$.

- Extends [Chillara, 2020a] to give lower bounds that do not deteriorate with increasing values of r, for a wider range of d.
- Though [Kayal, Saha, and Tavenas, 2018] give a lower bound that holds for a "slightly larger" range of d, our lower bound is quantitively better in comparable range of d.
- As with [Chillara, 2020a], we give a bound for a range of $r \geqslant d$.

Shallow separation

Theorem (Implicit in [Chillara, 2020a])

There exists an explicit polynomial Q_n such that

- it can be computed by a depth five multi-r-ic circuit of size $\mathrm{poly}(n)$
- but any depth four multi-r-ic circuit computing it must have size $n^{\Omega(\log n)}$.

Lower bounds for determinant polynomial

Lemma [Valiant, 1979]

$\mathsf{IMM}_{n,d}$ can be expressed as a Determinant of a $nd \times nd$ matrix whose entries are either variables or constants.

Theorem (Implicit in [Kayal, Saha, and Tavenas, 2018; Chillara, 2020b])

There exist fixed constants $a, b \in (0, 1)$ such that for all $r \leqslant N^a$, any syntactically multi-r-ic depth four circuit computing the Determinant of a generic $N \times N$ matrix must have size $2^{\Omega(N^b)}$.

Tools & Techniques

Broad theme of the proofs

Define a suitable complexity measure $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ such that the following holds:

- If f is computed by a small depth four multi-r-ic circuit then $\Gamma(f)$ is *small*.
- For the hard polynomial P, $\Gamma(P)$ is *large*.

Broad theme of the proofs

Define a suitable complexity measure $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ such that the following holds:

- If f is computed by a small depth four multi-r-ic circuit then $\Gamma(f)$ is *small*.
- For the hard polynomial P, $\Gamma(P)$ is *large*.

We "define", and then use the dimension of Projected Shifted Skew Partial Derivatives as the complexity measure.

Broad theme of the proofs

Define a suitable complexity measure $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ such that the following holds:

- If f is computed by a small depth four multi-r-ic circuit then $\Gamma(f)$ is *small*.
- For the hard polynomial P, $\Gamma(P)$ is *large*.

We "define", and then use the dimension of Projected Shifted Skew Partial Derivatives as the complexity measure.

This measure is related to

- Shifted Partial Derivatives measure of [Kayal, 2012],
- Skew Shifted Partial Derivatives measure of [Kayal, Saha, and Tavenas, 2018], and
- Projected Shifted Partial Derivatives measure of [Kayal, Limaye, Saha, and Srinivasan, 2014].

Broad theme of proofs for depth four circuits

- Let a depth four circuit C be expressed as $T_1 + \ldots + T_s$ where $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$.

Broad theme of proofs for depth four circuits

- Let a depth four circuit C be expressed as $T_1 + \ldots + T_s$ where $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$.
- Let $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ be a measure defined to be the dimension of a suitable vector space.

Broad theme of proofs for depth four circuits

- Let a depth four circuit C be expressed as $T_1 + \ldots + T_s$ where $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$.
- Let $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ be a measure defined to be the dimension of a suitable vector space.
- By subadditivity, $\Gamma(C) \leqslant s \cdot \max_{i \in [s]} \{\Gamma(T_i)\}$.

Broad theme of proofs for depth four circuits

- Let a depth four circuit C be expressed as $T_1 + \ldots + T_s$ where $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$.
- Let $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ be a measure defined to be the dimension of a suitable vector space.
- By subadditivity, $\Gamma(C) \leqslant s \cdot \max_{i \in [s]} \{\Gamma(T_i)\}$.
- If C computes a polynomial f then $\Gamma(C) = \Gamma(f)$.

Broad theme of proofs for depth four circuits

- Let a depth four circuit C be expressed as $T_1 + \ldots + T_s$ where $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$.
- Let $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ be a measure defined to be the dimension of a suitable vector space.
- By subadditivity, $\Gamma(C) \leqslant s \cdot \max_{i \in [s]} \{\Gamma(T_i)\}$.
- If C computes a polynomial f then $\Gamma(C) = \Gamma(f)$.
- $$\Gamma(f) \leqslant s \cdot \max_{i \in [s]} \{\Gamma(T_i)\} \implies s \geqslant \frac{\Gamma(f)}{\max_{i \in [s]} \{\Gamma(T_i)\}}.$$

Broad theme of proofs for depth four circuits

- Let a depth four circuit C be expressed as $T_1 + \ldots + T_s$ where $T_i = Q_{i,1} \cdot \ldots \cdot Q_{i,D}$.
- Let $\Gamma : \mathbb{F}[X] \mapsto \mathbb{N}$ be a measure defined to be the dimension of a suitable vector space.
- By subadditivity, $\Gamma(C) \leqslant s \cdot \max_{i \in [s]} \{\Gamma(T_i)\}$.
- If C computes a polynomial f then $\Gamma(C) = \Gamma(f)$.
- $$\Gamma(f) \leqslant s \cdot \max_{i \in [s]} \{\Gamma(T_i)\} \implies s \geqslant \frac{\Gamma(f)}{\max_{i \in [s]} \{\Gamma(T_i)\}}.$$
- Show that for all $i \in [s]$, $\Gamma(T_i)$ is not too large, and $\Gamma(f) \gg \Gamma(T_i)$.

Shifted partial derivatives

Dimension of Shifted Partial Derivatives [Kayal, 2012]

For a polynomial $f \in \mathbb{F}[X]$,

$$\partial^{=k} f := \left\{ \partial_m^k f \mid m \text{ is a monomial of degree } k \right\},$$

$$\mathbf{x}^{\leqslant \ell} \cdot \partial^{=k} f := \left\{ m_2 \cdot \partial_{m_1}^k f \mid \deg(m_1) = k \text{ and } \deg(m_2) \leqslant \ell \right\},$$

$$\text{and} \quad \Gamma_{k,\ell}^{[\mathrm{SPD}]}(f) := \dim(\mathbb{F}\text{-span}\left(\mathbf{x}^{\leqslant \ell} \cdot \partial^{=k}(f)\right)).$$

Shifted partial derivatives

Dimension of Shifted Partial Derivatives [Kayal, 2012]

For a polynomial $f \in \mathbb{F}[X]$,

$$\partial^{=k} f := \left\{ \partial_m^k f \mid m \text{ is a monomial of degree } k \right\},$$
$$\mathbf{x}^{\leqslant \ell} \cdot \partial^{=k} f := \left\{ m_2 \cdot \partial_{m_1}^k f \mid \deg(m_1) = k \text{ and } \deg(m_2) \leqslant \ell \right\},$$
$$\text{and} \quad \Gamma_{k,\ell}^{[\mathrm{SPD}]}(f) := \dim(\mathbb{F}\text{-span}\left(\mathbf{x}^{\leqslant \ell} \cdot \partial^{=k}(f)\right)).$$

Theorem [Gupta, Kamath, Kayal, and Saptharishi, 2014]

Let $T = Q_1 \cdot \ldots \cdot Q_D$ where D is "small" and $Q_{i,j}$'s are polynomials of "bounded degree". Then,

- $\Gamma_{k,\ell}^{[\mathrm{SPD}]}(T)$ is not too large for some range of k and ℓ, and
- there exists a polynomial f and parameters k, ℓ such that $\Gamma_{k,\ell}^{[\mathrm{SPD}]}(f) \gg \Gamma_{k,\ell}^{[\mathrm{SPD}]}(T)$.

Multi-r-ic depth four circuits

Let $T = Q_1 \cdot \ldots \cdot Q_D$ be a syntactic multi-r-ic product of polynomials.

Observation 1

Since T is syntactically multi-r-ic , $D \leqslant N \cdot r$.

Multi-r-ic depth four circuits

Let $T = Q_1 \cdot \ldots \cdot Q_D$ be a syntactic multi-r-ic product of polynomials.

Observation 1

Since T is syntactically multi-r-ic , $D \leqslant N \cdot r$.

Observation 2

For a random restriction $\rho : X \mapsto \{0, *\}$, with a high probability, $\rho(Q_i)$ is a low degree polynomial. That is, $\rho(T) = Q'_1 \cdot \ldots \cdot Q'_D$ is a product of low degree polynomials.

Multi-r-ic depth four circuits

$$\rho(T) = Q'_1 \cdot \ldots \cdot Q'_D$$

Obstacle: D is still too large

When $D \leqslant N \cdot r$, we get that $\Gamma^{[SPD]}_{k,\ell}(\rho(T)) \gg \Gamma^{[SPD]}_{k,\ell}(\rho(\mathsf{IMM}_{n,d}))$ for all k and ℓ.

Multi-r-ic depth four circuits

$$\rho(T) = Q'_1 \cdot \ldots \cdot Q'_D$$

Obstacle: D is still too large

When $D \leqslant N \cdot r$, we get that $\Gamma^{[SPD]}_{k,\ell}(\rho(T)) \gg \Gamma^{[SPD]}_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))$ for all k and ℓ.

Fix 1: Skew partitions [Kayal, Saha, and Tavenas, 2018]

- Partition X into $Y \sqcup Z$ such that $|Y| \gg |Z|$.
- Under suitable renaming, let

$$\rho(T) = Q_1(Y, Z) \cdot \ldots \cdot Q_t(Y, Z) \cdot R(Y).$$

- Observation: $t \leqslant |Z| \cdot r$.

Shifted skew partial derivatives

Dimension of shifted skew partial derivatives [Kayal, Saha, and Tavenas, 2018]

- Partition X into $Y \sqcup Z$ such that $|Y| \gg |Z|$.
- Under suitable renaming, let

$$\rho(T) = Q_1(Y, Z) \cdot \ldots \cdot Q_t(Y, Z) \cdot R(Y) \quad \text{where} \quad t \leqslant |Z| \cdot r.$$

- $\sigma_Y : \mathbb{F}[Y \sqcup Z] \mapsto \mathbb{F}[Z]$ such that $\sigma_Y(f) \in \mathbb{F}[Z]$. That is, all Y variables are set to 0.

$$\Gamma_{k,\ell}^{[KST]}(f) = \dim \left(\mathbb{F}\text{-span} \left\{ \left(\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y \left(\partial_Y^{=k} f \right) \right) \right\} \right)$$

Shifted skew partial derivatives

Dimension of shifted skew partial derivatives [Kayal, Saha, and Tavenas, 2018]

- Partition X into $Y \sqcup Z$ such that $|Y| \gg |Z|$.
- Under suitable renaming, let

$$\rho(T) = Q_1(Y,Z) \cdot \ldots \cdot Q_t(Y,Z) \cdot R(Y) \quad \text{where} \quad t \leqslant |Z| \cdot r.$$

- $\sigma_Y : \mathbb{F}[Y \sqcup Z] \mapsto \mathbb{F}[Z]$ such that $\sigma_Y(f) \in \mathbb{F}[Z]$. That is, all Y variables are set to 0.

$$\Gamma^{[KST]}_{k,\ell}(f) = \dim\left(\mathbb{F}\text{-span}\left\{\left(\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y\left(\partial_Y^{=k} f\right)\right)\right\}\right)$$

Theorem [Kayal, Saha, and Tavenas, 2018]

For a suitable random restriction ρ and carefully chosen values of k and ℓ, $\Gamma^{[KST]}_{k,\ell}(\rho(\mathrm{IMM}_{n,d})) \gg \Gamma^{[KST]}_{k,\ell}(\rho(T))$ and $s \geqslant \left(\frac{n}{r^{1.1}}\right)^{\Omega\left(\sqrt{\frac{d}{r}}\right)}$.

Refined complexity measure [Chillara, 2020a]

Observation

- $\sigma_Y(\partial^{=k}(\rho(\mathsf{IMM}_{n,d})))$ is a multilinear polynomial in $\mathbb{F}[Z]$.
- $\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y\left(\partial_Y^{=k} f\right)$ could potentially lead to non-multilinear polynomials.

Refined complexity measure [Chillara, 2020a]

Observation

- $\sigma_Y(\partial^{=k}(\rho(\mathsf{IMM}_{n,d})))$ is a multilinear polynomial in $\mathbb{F}[Z]$.
- $\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y\left(\partial_Y^{=k} f\right)$ could potentially lead to non-multilinear polynomials.

Projected Shifted Skew Partial Derivatives

- Partition X into $Y \sqcup Z$ such that $|Y| \gg |Z|$.
- $\sigma_Y : \mathbb{F}[Y \sqcup Z] \mapsto \mathbb{F}[Z]$ such that $\sigma_Y(f) \in \mathbb{F}[Z]$. That is, all Y variables are set to 0.
- mult : $\mathbb{F}[Z] \mapsto \mathbb{F}[Z]$ sets coefficients of all non-multilinear monomials to 0.

$$\Gamma_{k,\ell}(f) = \dim\left(\mathbb{F}\text{-span}\left\{\mathrm{mult}\left(\mathbf{z}^{\leqslant \ell} \cdot \sigma_Y\left(\partial_Y^{=k} f\right)\right)\right\}\right)$$

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathsf{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathsf{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathsf{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.
  - [Chillara, 2020a]: Uniform and independent random restrictions.

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.
  - [Chillara, 2020a]: Uniform and independent random restrictions.
  - [Chillara, 2020b]: Extends the random restrictions of [Kumar and Saraf, 2017] for $\mathrm{IMM}_{n,d}$.

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.
  - [Chillara, 2020a]: Uniform and independent random restrictions.
  - [Chillara, 2020b]: Extends the random restrictions of [Kumar and Saraf, 2017] for $\mathrm{IMM}_{n,d}$.
- [Nuanced but not hard] Show that $\Gamma_{k,\ell}(\rho(T))$ is not too large.

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathsf{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.
  - [Chillara, 2020a]: Uniform and independent random restrictions.
  - [Chillara, 2020b]: Extends the random restrictions of [Kumar and Saraf, 2017] for $\mathsf{IMM}_{n,d}$.
- [Nuanced but not hard] Show that $\Gamma_{k,\ell}(\rho(T))$ is not too large.
- [Harder part] Show that $\Gamma_{k,\ell}(\rho(\mathsf{IMM}_{n,d}))$ is large.

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.
  - [Chillara, 2020a]: Uniform and independent random restrictions.
  - [Chillara, 2020b]: Extends the random restrictions of [Kumar and Saraf, 2017] for $\mathrm{IMM}_{n,d}$.
- [Nuanced but not hard] Show that $\Gamma_{k,\ell}(\rho(T))$ is not too large.
- [Harder part] Show that $\Gamma_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))$ is large.
  - [Chillara, 2020a]: Uses leading monomial distance property (cf. [Chillara and Mukhopadhyay, 2019]), adapted to this new measure.

Devil lies in the details!

$$\frac{\Gamma_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))}{\Gamma_{k,\ell}(\rho(T_i))} \geqslant n^{\Omega(\sqrt{d})}.$$

- [Nuanced but not hard] Carefully design random restrictions ρ.
  - [Chillara, 2020a]: Uniform and independent random restrictions.
  - [Chillara, 2020b]: Extends the random restrictions of [Kumar and Saraf, 2017] for $\mathrm{IMM}_{n,d}$.
- [Nuanced but not hard] Show that $\Gamma_{k,\ell}(\rho(T))$ is not too large.
- [Harder part] Show that $\Gamma_{k,\ell}(\rho(\mathrm{IMM}_{n,d}))$ is large.
  - [Chillara, 2020a]: Uses leading monomial distance property (cf. [Chillara and Mukhopadhyay, 2019]), adapted to this new measure.
  - [Chillara, 2020b]: Uses a refined and careful counting through Leading Monomial approach (cf. [Kumar and Saraf, 2017]), adapted to this new measure.

Future work

- Prove better bounds against multi-r-ic depth four circuits.
- Combination of syntactic multi-r-ic and homogenity restrictions for formulas computing multi-r-ic polynomials is somewhat like monotone computation (cf. [Jerrum and Snir, 1982; Hrubeš and Yehudayoff, 2011]). Can we
  - weaken the restrictions or
  - prove bounds for multi-$(r-1)$-ic polynomials.
- Polynomial identity testing of depth three and depth four multi-r-ic circuits.

Thank you!