Non-Commutative Circuits and the Sum of Squares problem

Hrubes-Wigderson-Yehudayoff

TIFR, Mumbai

December 9th, 2017

1 Introduction

The sum of squares problem can be stated as follows.

Suppose we are working over the field $\mathbb{F}$. We want to find the complexity of n in terms of k for which an identity of the following kind exists:

$$\left(x_1^2 + x_2^2 + \cdots x_k^2\right) \cdot \left(y_1^2 + y_2^2 + \cdots y_k^2\right) = \left(f_1^2 + f_2^2 + \cdots f_n^2\right)$$

where each f_i is a bilinear form in $\{x_1, x_2, \ldots, x_k\}$, $\{y_1, y_2, \ldots, y_k\}$ over $\mathbb{F}$.

The main result in the paper then, is as follows.

If $\mathbb{F} = \mathbb{C}$, then showing $n = \Omega(k^{1+\varepsilon})$ with $\varepsilon > 0$ is enough to show that any non-commutative circuit computing the $n \times n$ permanent requires $\exp(n)$ size.

1 *The model we are working with: Non-Commutative Circuits*

Non-commutative circuits are like normal algebraic circuits with the only difference being that each multiplication gate has a specified left child and a right child. Note that this makes a big difference since in particular,

$$x^2 - y^2 \neq (x+y)(x-y)$$

in the the non-commutative world.

Previous Works

We note that there is no better lower-bound known for general non-commutative circuits than in the commutative setting. However, there have been some non-trivial work in this area as well. A few important results known are as follows:

1. **Nisan**: Any non-commutative formula computing the $n \times n$ determinant or permanent must have size $\Omega(2^n)$.
2. **Nisan**: There exists an explicit polynomial over n variables that has an $O(n)$ sized non-commutative circuit, but any non-commutative formula computing it requires size $2^{\Omega(n)}$.

3. **Chien Sinclair et al.**: The permanent can be approximated well efficiently if the determinant of some corresponding matrix can be computed efficiently.

4. **Arvind-Srinivasan**: In the non-commutative world, computing the determinant is as hard as computing the permanent.

These are however not directly related to the result we will present here. So before going any further, let us look at the Sum-of-Squares more carefully.

2 *The Sum of Squares Problem*

Consider the polynomial

$$SOS_k = \left(x_1^2 + x_2^2 + \cdots x_k^2\right) \cdot \left(y_1^2 + y_2^2 + \cdots y_k^2\right).$$

Let $S_{\mathbb{F}}(k)$ denote the minimum value of n for which

$$SOS_k = z_1^2 + z_2^2 + \cdots + z_n^2,$$

where each z_i is a bilinear form in $\{x_1, x_2, \ldots, x_k\}, \{y_1, y_2, \ldots, y_k\}$ over $\mathbb{F}$. The Sum of Squares problem is to find $S_{\mathbb{F}}(k)$.

Note: If $\mathbb{F}$ has charcteristic 2, then $n = 1$ and for any other field, the trivial bounds are: $k \leq S_{\mathbb{F}}(k) \leq k^2$.

The Sum of Squares problem over Reals

The sum of squares problem over reals has been studied for a long time by mathematicians. Let us first look at some non-trivial cases for which $S_{\mathbb{R}}(k) = k$.

1. For $k = 1$, $x_1^2 y_1^2 = (x_1 y_1)^2$.
   Note that this is the same as saying $|\alpha|^2 |\beta|^2 = |\alpha\beta|^2$ for $\alpha = x_1$ and $\beta = y_1$.

2. For $k = 2$, $(x_1^2 + y_1^2) = (x_1 y_1 - x_2 y_2)^2 + (x_1 y_2 + x_2 y_1)^2$.
   Note that this is the same as saying $|z_1|^2 |z_2|^2 = |z_1 z_2|^2$ when we view $z_1 = (x_1, x_2)$ and $z_2 = (y_1, y_2)$ as complex numbers.

3. For $k = 4$, Euler showed that $S_{\mathbb{R}}(4) = 4$.
   A similar interpretation can be made as before if we view $z_1 = (x_1, x_2, x_3, x_4)$ and $z_2 = (y_1, y_2, y_3, y_4)$ as quarternions — defined by Hamilton after Euler's proof.

4. For $k = 8$ again, a similar interpretation is possible by viewing $z_1 = (x_1, x_2, \ldots, x_8)$ and $z_2 = (y_1, y_2, \ldots, y_8)$ as octonions

After this, people tried to show that $S_{\mathbb{R}}(16) = 16$. However, in 1898, Hurwitz showed that $S_{\mathbb{R}}(k) > k$ for every $k \notin \{1, 2, 4, 8\}$. Using topological and algebraic tools, it was shown that

$$S_{\mathbb{R}}(k) \geq (2 - o(1))k$$

which is the current best lower-bound. The current best upperbound was given by Radon-Hurwitz. They showed that

$$S_{\mathbb{R}}(k) \leq O\left(\frac{k^2}{\log k}\right).$$

Their proof also works over $\mathbb{Z}$. Thus,

$$S_{\mathbb{Z}}(k) \leq O\left(\frac{k^2}{\log k}\right).$$

In this paper, Hrubes-Wigderson-Yehudayoff show that

$$S_{\mathbb{R}}(k) \geq \Omega\left(k^{6/5}\right).$$

We will however, not be seeing the proof here.

3 *The connection between Non-commutative Circuits and SOS*

As noted before the main result in the paper shows that a sufficiently strong super-linear lower-bound for $S_{\mathbb{C}}(k)$ implies an exponential lowerbound for Non-commutative circuits computing the $n \times n$ permanent.

Now in the non-commutative setting, one can define the permanent in many ways. We define it in a row-by-row manner as follows:

$$\mathsf{Perm}_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i\sigma(i)}.$$

Formally, the main theorem in the paper is as follows:

Theorem 1.1. *Let $\mathbb{F}$ be a field which contain $\sqrt{-1}$. Assume that $S_{\mathbb{F}}(k) \geq \Omega(k^{1+\varepsilon})$ for some constant $\varepsilon > 0$. Then, any non-commutative circuit computing Perm_n requires size $2^{\Omega(n)}$.*

Sum of Squares Complexity and Bilinear Complexity

We now define a few operators which we will use in the proof of Theorem 1.1.

Definition 1.2. *Let f be a commutative polynomial of degree 4 over a field $\mathbb{F}$. f is said to be bi-quadratic in $X = \{x_1, x_2, \ldots, x_k\}$ and $Y = \{y_1, y_2, \ldots, y_k\}$, if every monomial in f has the form $x_{i_1} x_{i_2} y_{j_1} y_{j_2}$.* ◊

Definition 1.3. *For a commutative bi-quadratic polynomial over $X = \{x_1, x_2, \ldots, x_k\}$ and $Y = \{y_1, y_2, \ldots, y_k\}$, define:*

- ***Sum of Squares Complexity**: $S_{\mathbb{F}}(f)$*
  Smallest n (possibly infinite) so that f can be written as $f = z_1^2 + \cdots + z_n^2$
- ***Bilinear Complexity**: $B_{\mathbb{F}}(f)$*
  Smallest n (possibly infinite) so that f can be written as $f = z_1 z_1' + \cdots + z_n z_n'$

where each z_i, z_i' are bilinear forms in X, Y. ◊

Note: $S_{\mathbb{F}}(SOS_k) = S_{\mathbb{F}}(k)$.

Relation between $S_{\mathbb{F}}(k)$ and $B_{\mathbb{F}}(k)$

We want to prove Theorem 1.1 by using $B_{\mathbb{F}}$ instead of $S_{\mathbb{F}}$. For that we need to see how the two relate to each other. Clearly,

$$B_{\mathbb{F}}(k) \leq S_{\mathbb{F}}(k).$$

Now, assume $\sqrt{-1} \in \mathbb{F}$. Then

$$2zz' = (z + z')^2 + (\sqrt{-1}z)^2 + (\sqrt{-1}z')^2$$

and so $S_{\mathbb{F}}(f) \leq 3B_{\mathbb{F}}(f)$. Thus if $\sqrt{-1} \in \mathbb{F}$, then $B_{\mathbb{F}}(k) = \Theta(S_{\mathbb{F}}(k))$.

using the above observation, the following theorem is clearly enough to show Theorem 1.1.

Theorem 1.4. *For any field $\mathbb{F}$, assume $B_{\mathbb{F}}(k) \geq \Omega(k^{1+\varepsilon})$ for some constant $\varepsilon > 0$. Then, any non-commutative circuit computing Perm_n requires size $2^{\Omega(n)}$.*

2 The Proof Strategy

Theorem 1.4 will be proved in broadly three parts:

Part I: This part will consist of two steps.

1. Homogenise the circuit for Perm_n: Note that the usual homogenisation respects non-commutativity. Thus if there exists a non-commutative circuit computing Perm_n of size s, there is a corresponding homogeneous non-commutative circuit computing Perm_n of size $s' = O(n^2 s)$.
2. Define "width" of a non-commutative polynomial and show that

$$\mathsf{width}(\mathsf{Perm}_n) = O(ns').$$

Thus at this point, it is enough to show the following statement:

$$B_{\mathbb{F}}(SOS_k) = \Omega(k^{1+\varepsilon}) \Rightarrow \mathsf{width}(\mathsf{Perm}_n) = 2^{\Omega(n)}.$$

Part II: If $\mathsf{ID}_k = \sum_{i,j\in[k]} x_i y_j x_i y_j$, then show that

$$\mathsf{width}(\mathsf{ID}_k) = \Theta(B_{\mathbb{F}}(SOS_k)).$$

Thus at this point, it is enough to show the following statement:

$$\mathsf{width}(\mathsf{ID}_k) = \Omega(k^{1+\varepsilon}) \Rightarrow \mathsf{width}(\mathsf{Perm}_n) = 2^{\Omega(n)}.$$

Part III: This part will consist of three steps.

1. Show that $\mathsf{width}(\mathsf{ID}_k) = \mathsf{width}(\mathsf{ID}_k')$ for $\mathsf{ID}_k' = \sum_{i,j\in[k]} x_i x_j x_i x_j$.

2. Show that $\mathsf{width}(\mathsf{LID}_r) = \Omega(2^{-r}\,\mathsf{width}(\mathsf{ID}'_k))$ for $\mathsf{LID}_r = \sum_{e\in\{0,1\}^{2r}} z_e z_e$ if $k = 2^r$ and $z_e = \prod_{j=1}^{2r} z_{e_j}$ where $e = (e_1, e_2, \ldots, e_{2r}) \in \{0,1\}$.
3. Show that $\mathsf{LID}_r = \mathsf{Perm}_{4r}$.

Thus,

$$\begin{aligned}\mathsf{width}(\mathsf{Perm}_{4r}) &= \mathsf{width}(\mathsf{LID}_r) = \Omega(2^{-r}\,\mathsf{width}(ID'_k)) \\ &= \Omega(2^{-r}\,\mathsf{width}(ID_k)) = \Omega(2^{-r}\cdot 2^{r(1+\varepsilon)}) \\ &= \Omega(2^{r\varepsilon}) = 2^{\Omega(r)}.\end{aligned}$$

We will now look at the proof of each part separately. From now on, the term "polynomial" will be used to mean non-commutative polynomial, unless mentioned otherwise.

3 A Sufficient condition for Proving Non-commutative circuit Lower-bounds

Intuitively, we want to say that if a homogeneous polynomial has a small circuit computing it, then its monomials can be grouped into not too many groups where each group share a common central part.

More formally, let us call a homogeneous polynomial f central, if $\exists m, d_0, d_1, d_2$ such that

$$f = \sum_{i=1}^{m} h_i g h'_i$$

where $d = \deg(f)$, $\frac{d}{3} \le d_0 \le \frac{2d}{3}$, $d_0 + d_1 + d_2 = d$ and

- g is a homogeneous polynomial of degree d_0
- $\forall i$, h_i is a homogeneous polynomial of degree d_1
- $\forall i$, h'_i is a homogeneous polynomial of degree d_2

As there is no bound on $'M'$ as such, we can assume that h_i is a scalar times a monomial for every i and that h'_i is a monomial for every i.

Further, a homogeneous polynomial f is said to have "width" n, denoted by

$$\mathsf{width}(f) = n$$

if n is the smallest number for which

$$f = \sum_{i=1}^{n} f_i$$

and each f_i is a central polynomial.

Clearly, the following is be enough to to show what was required to be shown in the part.

If s is the size of a homogeneous circuit computing a polynomial f, then

$$\mathsf{width}(f) = O(ds)$$

where $d = \deg(f)$.

Let f be a homogeneous polynomial of degree d, and let s be the size of a homogeneous circuit $\mathcal{C}$ computing it. We want to show that $\mathsf{width}(f) \leq ds$. We will do so by showing the following claim.

Claim 3.1. *Let $\{g_1, g_2, \ldots, g_t\}$ be the set of polynomials being computed at the various gates in $\mathcal{C}$ of degree in the range $[\frac{d}{3}, \frac{2d}{3}]$. Then any polynomial g, that is computed by any of the gates in $\mathcal{C}$ must have the form*

$$g = \sum_{i \in [t]} \left(\sum_{j \in [m]} h_{ij} g_i h'_{ij} \right)$$

if $\deg(g) \geq \frac{d}{3}$.

It is not too hard to see why this is enough. Taking $g = f$, we get

$$\begin{aligned} f &= \sum_{i \in [t], j \in [m]} h_{ij} g_i h'_{ij} \\ &= \sum_{i \in [t], j \in [m]} \sum_{k=0}^{d-\deg(g_i)} h_{ij}^{(k)} g_i {h'}_{ij}^{(d-k-\deg(g_i))} \\ &= \sum_{i \in [t]} \sum_{k=0}^{d-\deg(g_i)} \left(\sum_{j \in [m]} h_{ij}^{(k)} g_i {h'}_{ij}^{(d-k-\deg(g_i))} \right) \end{aligned}$$

where $\sum_{j \in [m]} h_{ij}^{(k)} g_i {h'}_{ij}^{(d-k-\deg(g_i))}$ is a central polynomial and $k \leq d$.

Just to clarify notation, for any polynomial p and any integer $r \leq \deg(p)$, $p^{(r)}$ denotes the homogeneous degree r part of p.

Let us now look at the proof of Claim 3.1.

Proof of Claim 3.1. Let g be any polynomial with $\deg(g) \geq \frac{d}{3}$ that is computed by some gate in $\mathcal{C}$. Then,

Case 1: $\deg(g) \leq \frac{2d}{3}$

Set
$$m = 1 \text{ and } h_{i1}, h'_{i1} = \begin{cases} 1 & \text{if} \quad g_i = g \\ 0 & \text{otherwise} \end{cases}$$

Case 2: $\deg(g) > \frac{2d}{3}$

We prove this case by induction on the depth at which g is calculated.

Let g be calculated at a vertex which is a + gate. Then, $g = g' + g''$ where

$$\deg(g'), \deg(g'') > \frac{2d}{3}.$$

By induction, $\exists m$ and

$$\{h_{ij}\}_{i\in[t],j\in[m]}, \{h'_{ij}\}_{i\in[t],j\in[m]}, \{\bar{h}_{ij}\}_{i\in[t],j\in[m]}, \{\bar{h}'_{ij}\}_{i\in[t],j\in[m]}$$

such that

$$g' = \sum_{i\in[t]}\left(\sum_{j\in[m]} h_{ij} g_i h'_{ij}\right)$$

and

$$g'' = \sum_{i\in[t]}\left(\sum_{j\in[m]} \bar{h}_{ij} g_i \bar{h}'_{ij}\right).$$

Thus,

$$g = \sum_{i\in[t]}\left(\sum_{j\in[m]} \left(h_{ij} g_i h'_{ij} + \bar{h}_{ij} g_i \bar{h}'_{ij}\right)\right).$$

Next, let g be calculated at a vertex which is a $\times$ gate. Then, $g = g' \times g''$ where

$$\deg(g') > \frac{d}{3} \text{ or } \deg(g'') > \frac{d}{3}.$$

Without loss, assume it is g'. By induction, $\exists m, \{h_{ij}\}_{i\in[t],j\in[m]}, \{h'_{ij}\}_{i\in[t],j\in[m]}$ such that

$$g' = \sum_{i\in[t]}\left(\sum_{j\in[m]} h_{ij} g_i h'_{ij}\right).$$

Thus,

$$g = g'g'' = \sum_{i\in[t]}\left(\sum_{j\in[m]} (h_{ij}g'')g_i(h'_{ij}g'')\right).$$

This completes the proof of Claim 3.1. □

Hence, for any size s homogeneous circuit computing a polynomial f,

$$\mathsf{width}(f) = O(ds)$$

where $d = \deg(f)$. In particular this proves that if s is the size of any homogeneous non-commutative circuit computing Perm_n, then

$$\mathsf{width}(\mathsf{Perm}_n) = O(ns).$$

Thus, finding a lower-bound for non-commutative circuit size computing Perm_n is reduced to finding a lower-bound for $\mathsf{width}(\mathsf{Perm}_n)$.

Now note that SOS_k is a commutative polynomial. In the next section we will define a non-commutative analogue of SOS_k and show that the "width" of that polynomial is the same as $B_{\mathbb{F}}(SOS_k)$.

4 Width of degree-four Non-commutative polynomials and Bilinear Complexity of the Commutative counterparts

We begin by defining a non-commutative analogue of the SOS_k polynomial, namely

$$\mathsf{ID}_k = \sum_{i,j\in[k]} x_i y_j x_i y_j.$$

We want to show that $\mathsf{width}(\mathsf{ID}_k) = B_{\mathbb{F}}(SOS_k)$. However before we go into that, let us fix some notations.

1 *Some Notations and Observations*

Let $X = \{x_1, x_2, \ldots, x_n\}$ be the variables on which the polynomials of our interest depends, and let $X_1, X_2, \ldots, X_r$ be (not necessarily disjoint) subsets of X. For a polynomial f, let $f[X_1, X_2, \ldots, X_r]$ be a homogeneous degree r polynomial of the following type:

$$\mathsf{coeff}_\alpha(f[X_1, X_2, \ldots, X_r]) = \begin{cases} \mathsf{coeff}_\alpha(f) & \text{if } \alpha = x_1x_2\ldots x_r \text{ with } x_i \in X_i \text{ for every } i \\ 0 & \text{otherwise} \end{cases}$$

With the above definition, the following observation is not too hard.

Observation 4.1. *If f is a central polynomial such that*

$$f = f[X_1X_2X_3X_4],$$

then either $f = g[X_1, X_2]h[X_3, X_4]$ or $f = \sum_{i\in[m]} h_i[X_1]g[X_2, X_3]h_i'[X_4]$. Here g, h, h_i, h_i' are some appropriate polynomials.

Sketch of Proof. It is not hard to show that $f = g[X_1, X_2]h[X_3, X_4]$ when $d_1 = 0$ or $d_2 = 0$. Similarly, $f = \sum_{i\in[m]} h_i[X_1]g[X_2, X_3]h_i'[X_4]$ when $d_1 = 1 = d_2$. □

The above observation immediately proves the following lemma.

Lemma 4.2. *If $f = F[X_1, X_2, X_3, X_4]$, then $\mathsf{width}(f)$ is the smallest n such that f can be written as $f = f_1 + f_2 + \cdots + f_n$, where for every $t \in [n]$ one of the following two is true:*

- $f_t = g_t[X_1, X_2]h_t[X_3, X_4]$
- $f_t = \sum_{i\in[m]} h_{t_i}[X_1]g_t[X_2, X_3]h_{t_i}'[X_4]$

Here g_t, h_t, h_{t_i}, h_{t_i}' are some appropriate polynomials.

2 *Connection between degree-4 non-commutative polynomials and Bilinear Complexity*

Let f be a polynomial on variables $X = \{x_1, x_2, \ldots, x_k\}$ and $Y = \{y_1, y_2, \ldots, y_k\}$ such that $f = F[X, Y, X, Y]$. Then, f will look like

$$f = \sum_{i_1, i_2, j_1, j_2 \in [k]} a_{i_1 j_1 i_2 j_2} x_{i_1} y_{j_1} x_{i_2} y_{j_2}.$$

f is said to be (X, Y)-symmetric if for every $(i_1, j_1, i_2, j_2) \in [k]^4$,

$$a_{i_1 j_1 i_2 j_2} = a_{i_2 j_1 i_1 j_2} = a_{i_1 j_2 i_2 j_1} = a_{i_2 j_2 i_1 j_1}.$$

The following theorem relates the "width" of a non-commutative polynomial and the bilinear complexity of its commutative counterpart.

Notation 4.3. *For a non-commutative polynomial g, let $g^{(c)}$ denote its commutative counterpart.* ◊

Theorem 4.4. *Let f be a homogeneous non-commutative polynomial of degree 4 such that $f = f[X, Y, X, Y]$. Then,*

1. $B\left(f^{(c)}\right) \leq \mathsf{width}(f)$
2. *If the characteristic of $\mathbb{F}$ is not 2 and f is (X, Y)-symmetric, then*

$$\mathsf{width}(f) \leq 4 B_{\mathbb{F}}(f^{(c)}).$$

Proof. The first part is easy to see. The second part is slightly non-trivial.

1. As $f = F[X, Y, X, Y]$, by Lemma 4.2 if $\mathsf{width}(f) = n$ then $f = f_1 + f_2 + \cdots + f_n$ where each f_t looks like

$$f_t = g_t[X_1, X_2] h_t[X_3, X_4] \text{ or } f_t = \sum_{i \in [m]} h_{t_i}[X_1] g_t[X_2, X_3] h'_{t_i}[X_4].$$

Thus, $f^{(c)} = f_1^{(c)} + f_2^{(c)} + \cdots + f_n^{(c)}$ where each $f_t^{(c)}$ looks like

$$f_t^{(c)} = g_t^{(c)}[X_1, X_2] h_t^{(c)}[X_3, X_4]$$

or

$$f_t^{(c)} = g_t^{(c)}[X_2, X_3] \sum_{i \in [m]} h_{t_i}^{(c)}[X_1] {h'}_{t_i}^{(c)}[X_4].$$

Viewing $\sum_{i \in [m]} h_{t_i}^{(c)}[X_1] {h'}_{t_i}^{(c)}[X_4]$ as $h_t^{(c)}$, we have that if $\mathsf{width}(f) = n$ then

$$f^{(c)} = \sum_{i \in [n]} f_t^{(c)}$$

where each $f_t^{(c)}$ is a product of two bilinear forms in X and Y. Thus,

$$B_{\mathbb{F}}\left(f^{(c)}\right) \leq \mathsf{width}(f).$$

2. To see the opposite direction, let $B\left(f^{(c)}\right) = n$. Then,

$$f^{(c)} = z_1 z_1' + \cdots + z_n z_n'$$

where each z_i, z_i' is a bilinear form in X, Y. Thus,

$$z_i = \sum_{j=1}^{n} x_j g_{ij}(Y) \text{ and } z_i' = \sum_{j=1}^{n} x_j g_{ij}'(Y)$$

where each g_{ij}, g_{ij}' are homogeneous degree-one polynomials in Y. Define

$$\begin{aligned} f_i = &\left(\sum_{j=1}^{n} x_j g_{ij}(Y)\right)\left(\sum_{j=1}^{n} x_j g_{ij}'(Y)\right) + \left(\sum_{j=1}^{n} x_j g_{ij}'(Y)\right)\left(\sum_{j=1}^{n} x_j g_{ij}(Y)\right) \\ &+ \sum_{j=1}^{n} x_j \left(\sum_{k=1}^{n} g_{ik}(Y) x_k\right) g_{ij}'(Y) + \sum_{j=1}^{n} x_j \left(\sum_{k=1}^{n} g_{ik}'(Y) x_k\right) g_{ij}(Y). \end{aligned}$$

Clearly, every f_i is the sum of four central polynomials. Thus, to show that $\mathsf{width}(f) \le 4n$, it is enough to show that

$$f = \frac{1}{4}\sum_{i=1}^{n} f_i.$$

Firstly, it is easy to see that $f_i^{(c)} = z_i z_i'$ and hence $f = \frac{1}{4}\sum_{i=1}^{n} f_i$. Also, as f is (X,Y)-symmetric, for any monomial $\alpha = x_{i_1} y_{j_1} x_{i_2} y_{j_2}$,

$$\mathsf{coeff}_\alpha(f) = \begin{cases} \mathsf{coeff}_\alpha\left(f^{(c)}\right) & \text{if} \quad i_1 = i_2 \text{ and } j_1 = j_2 \\ 2\,\mathsf{coeff}_\alpha\left(f^{(c)}\right) & \text{if} \quad i_1 = i_2 \;\&\; j_1 \ne j_2 \text{ or } i_1 \ne i_2 \;\&\; j_1 = j_2 \\ 4\,\mathsf{coeff}_\alpha\left(f^{(c)}\right) & \text{if} \quad i_1 \ne i_2 \text{ and } j_1 \ne j_2 \end{cases}$$

Further, we have constructed the f_i in such a way that they are (X,Y)-symmetric and thus a similar relation will hold between $\mathsf{coeff}_\alpha(f_i)$ and $\mathsf{coeff}_{\alpha^{(c)}}(f_i^{(c)})$. This shows that

$$f = \frac{1}{4}\sum_{i=1}^{n} f_i$$

which is what we wanted. □

Clearly for $f = \mathsf{ID}_k$, the above theorem shows that

$$\mathsf{width}(\mathsf{ID}_k) = \Theta(B_{\mathbb{F}}(SOS_k)).$$

Thus at this point, it is enough to show that a sufficiently strong lower-bound on $\mathsf{width}(\mathsf{ID}_k)$ will imply an exponential lower-bound on $\mathsf{width}(\mathsf{Perm}_n)$ and hence on non-commutative circuits computing Perm_n.

We will now proceed to show that a sufficiently strong super-linear lower-bound for the width of degree four polynomials imply exponential lower-bounds for the width of a related high degree polynomial. It will turn out that for ID_k, the related high degree polynomial is Perm_{4r} if $k = 2^r$.

5 Super-linear lower-bound for the width of degree four polynomials imply Exponential lower-bound for the width of a related high degree polynomial

Firstly, we note that there is a natural way to go from a homogeneous degree $4r$ polynomial in 2 variables to a homogeneous degree 4 polynomial in 2^r variables and vice-versa (upto renaming of variables).

So let f be a homogeneous polynomial of degree $4r$ in two variables, say z_0, z_1. For every monomial α of degree r in z_0, z_1, define a new variable x_α and define a homogeneous polynomial g of degree 4 over variables $\{x_\alpha\}_\alpha$ as follows:

$$\mathsf{coeff}_{x_{\alpha_1} x_{\alpha_2} x_{\alpha_3} x_{\alpha_4}}(g) = \mathsf{coeff}_{\alpha_1\alpha_2\alpha_3\alpha_4}(f).$$

Conversely, for a homogeneous degree 4 polynomial g over 2^r variables, define f to be a homogeneous polynomial of degree $4r$ over two variables as follows:

$$\mathsf{coeff}_{z_{(i_1)} z_{(i_2)} z_{(i_3)} z_{(i_4)}}(f) = \mathsf{coeff}_{x_{i_1} x_{i_2} x_{i_3} x_{i_4}}(g).$$

Here, (i) is the binary representation of i and

$$z_{(i)} = \prod_{j \in [r]} z_{i_j}$$

where $(i) = (i_1, i_2, \ldots, i_r) \in \{0,1\}^r$.

For a polynomial f of degree $4r$ in two variables, let $f^{(\lambda)}$ denote the corresponding polynomial of degree 4 in 2^r variables. We want to relate the width of f and $f^{(\lambda)}$. The reason is as follows.

Even though ID_k is a polynomial of degree 4 in $2k$ variables, since we are in the non-commutative setting, the position of a variable is more important than its name. Thus, we can define another polynomial ID'_k, which is a degree 4 polynomial in only k variables but has the property that

$$\mathsf{width}(\mathsf{ID}_k) = \mathsf{width}(\mathsf{ID}'_k).$$

Formally, we define ID'_k as follows:

$$\mathsf{ID}'_k = \sum_{i,j \in [k]} x_i x_j x_i x_j.$$

Thus if f is the polynomial for which $f^{(\lambda)} = \mathsf{ID}'_k$, then the problem is now reduced to showing the following things:

1. $\mathsf{width}(\mathsf{ID}_k) = \mathsf{width}(\mathsf{ID}'_k)$
2. $\mathsf{width}(f) = \Omega(2^{-r}\,\mathsf{width}(f^{(\lambda)}))$ if $k = 2^r$
3. $f = \mathsf{Perm}(M)$ for some suitable matrix.

First, let us see what the polynomial f looks like. Using the process described,

$$f = \sum_{i,j \in [k]} z_{(i)} z_{(j)} z_{(i)} z_{(j)}$$

where $z_{(i)} = \prod_{k=1}^{r} z_{i_k}$ if $(i) \in \{0,1\}^r$ and $(i) = (i_1, i_2, \dots, i_r) \in \{0,1\}^r$.

Clearly for $k = 2^r$, $f^{(\lambda)} = \mathsf{ID}'_k$. We will call the polynomial f as LID_r. Formally,

$$\mathsf{LID}_r = \sum_{e \in \{0,1\}^{2r}} z_e z_e.$$

We will now proceed to prove the three statements noted above.

Proof of 1. Clearly $\mathsf{width}(\mathsf{ID}'_k) \le \mathsf{width}(\mathsf{ID}_k)$. To see the opposite inequality, let

$$\mathsf{ID}'_k = f'_1 + \cdots + f'_n$$

and let

$$\mathsf{ID}_k = \sum_{i,j \in [k]} x_i y_j x_i y_j = \sum_{i,j \in [k]} x_{i,0} x_{j,1} x_{i,0} x_{j,1}.$$

We know that each f_i is a homogeneous degree 4 polynomial. Thus for any i,

$$f'_i = g'h' \text{ or } f'_i = \sum_{j \in [m]} h'_{ij} g'_i \bar{h}'_{ij}$$

where g', h', g'_i are homogeneous polynomials of of degree 2 and $h'_{ij}, \bar{h}'_{ij}$ are homogeneous degree 1 polynomials.

Now if

$$g' = \sum \alpha_k x_{k_1} x_{k_2}, \quad h' = \sum \beta_k x_{k_1} x_{k_2}, \quad g'_i = \sum \gamma_k x_{k_1} x_{k_2}$$

and

$$h'_{ij} = \sum \delta_k x_k, \quad \bar{h}'_{ij} = \sum \rho_k x_k,$$

let us define

$$g = \sum \alpha_k x_{k_1} x_{k_2}, \quad h = \sum \beta_k x_{k_1} x_{k_2}, \quad g_i = \sum \gamma_k x_{k_1} x_{k_2}$$

and

$$h_{ij} = \sum \delta_k x_k, \quad \bar{h}_{ij} = \sum \rho_k x_k.$$

With these definitions, let us define

$$f_i = \begin{cases} gh & \text{if} \quad f'_i = g'h' \\ \sum_{j \in [m]} h_{ij} g_i \bar{h}_{ij} & \text{if} \quad f'_i = \sum_{j \in [m]} h'_{ij} g'_i \bar{h}'_{ij} \end{cases}$$

Then, $\mathsf{ID}_k = \sum_{i \in [n]} f_i$ where each f_i is central and thus

$$\mathsf{width}(\mathsf{ID}_k) \le \mathsf{width}(\mathsf{ID}'_k).$$

□

Proof of 2. We want to show that for any polynomial f of degree $4r$ over two variables, $\mathsf{width}(f) = \Omega(2^{-r}\,\mathsf{width}(f^{(\lambda)}))$.

Note that $f^{(\lambda)}$ is a degree 4 polynomial over 2^r variables and f is connected to $f^{(\lambda)}$ in the following way. f has 4 blocks of homogeneous polynomials of degree r, one block each for the 4 variables in a monomial of $f^{(\lambda)}$.

It was easy to work with degree 4 polynomials beacuse central polynomials of degree 4 have a nice structure. It is natural to try and work in a similar way on $f^{(\lambda)}$ because of its connection to f.

To do so, we define block-central polynomials:

Definition 5.1. *A homogeneous polynomial f of degree $4r$ is said to be block-central if either of the following is true:*

- *$f = gh$ where g, h are homogeneous polynomials with $\deg(g) = 2r = \deg(h)$.*
- *$f = \sum_{i\in[m]} h_i g \bar{h}_i$ where for every i, h_i, g, $\bar{h}_i$ are homogeneous polynomials with degrees r, $2r$ and r respectively.* ◊

Clearly, every block-central polynomial is central. We will show that every central polynomial can be written as a sum of 2^r block-central polynomials. This will allow us to prove the required result. Let us first see why this is the case.

Let f be a homogeneous polynomial of degree $4r$ over two variables. Further, let $f = f_1 + f_2 + \cdots + f_n$ where each f_i is a central polynomial. Then

$$\begin{aligned}
& f = f_1 + f_2 + \cdots + f_{n'} \text{ where } n' \le 2^r n \text{ and} \\
& \qquad\qquad\qquad\qquad\qquad\qquad \text{each } f_i \text{ is a block central polynomial} \\
\Rightarrow\ & f = f_1 + f_2 + \cdots + f_{n'} \text{ where each } f_i^{(\lambda)} \text{ is a central polynomial} \\
& \qquad\qquad\qquad\qquad \text{by making the same natural transition in } g, g, g_i, \bar{h}_i \\
\Rightarrow\ & \mathsf{width}(f^{(\lambda)}) = O(2^r\,\mathsf{width}(f))
\end{aligned}$$

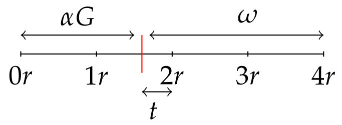
So now, the only thing left to prove is to show that every central polynomial can be written as the sum of 2^r block-central polynomials. Let f be a homogeneous central polynomial of degree $4r$. Then,

$$f = \sum_{\alpha\in M(d_1),\omega\in M(d_2)} c(\alpha,\omega)\alpha G\omega$$

for some fixed d_0, d_1, d_2 such that $\frac{4r}{3} \le d_0 \le \frac{8r}{3}$ and $d_0 + d_1 + d_2 = d$. Here G is a homogeneous polynomial of degree d_0, αs are monomials of degree d_1 and ωs are monomials of degree d_2. Further, $c(\alpha,\omega)$ is a scalar depending on α, ω and $M(k)$ is the set of all monomials of degree k over z_0, z_1.

We want to write f as the sum of at most 2^r block central polynomials. To do so, we basically want to write a similar expression for f, but this time with each of d_0, d_1, d_2 being multiples of r.

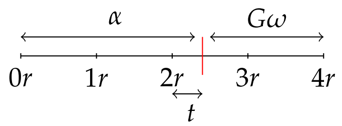
Case 1: $d_0 + d_1 \le 2r$



$$\begin{aligned}
f &= \sum_{\substack{\alpha \in M(d_1) \\ \omega_1 \in M(t) \\ \omega_2 \in M(d_2 - t)}} c(\alpha, \omega_1 \omega_2) \alpha G \omega_1 \omega_2 \\
&= \sum_{\substack{\alpha \in M(d_1) \\ \omega_1 \in M(t)}} \left((\alpha G \omega_1) \left(\sum_{\omega_2 \in M(d_2 - t)} c(\alpha, \omega_1 \omega_2) \omega_2 \right) \right) \\
&= \sum_{\substack{\alpha \in M(d_1) \\ \omega_1 \in M(t)}} \left(g_{\alpha, \omega_1} h_{\alpha, \omega_1} \right)
\end{aligned}$$

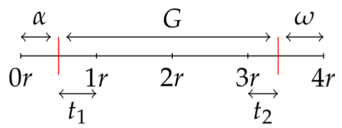
Thus, f can be written as a sum of $2^{d_1 + t} \le 2^{\frac{2r}{3}}$ block central polynomials of the type gh.

Case 2: $d_0 + d_2 \le 2r$



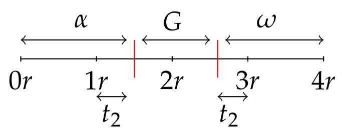
Similar to last case and f can be written as a sum of $2^{d_2 + t} \le 2^{\frac{2r}{3}}$ block central polynomials of the type gh.

Case 3: $d_1, d_2 \le r$



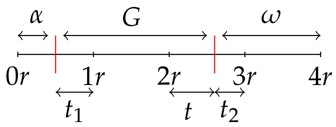
A similar natural way of grouping terms to make blocks of length that is a multiple of r will allow us to write f as the sum of $2^{t_1 + t_2} \le 2^{\frac{2r}{3}}$ block central polynomials of the type $\sum h_i g h_i'$.

Case 4: $d_1, d_2 \ge r$



Similar to the last case and f can be written as the sum of $2^{t_1 + t_2} \le 2^{\frac{2r}{3}}$ block central polynomials of the type $\sum h_i g h_i'$.

Case 5: $d_1 \leq r, d_2 \geq r$



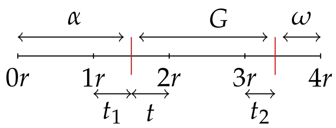
A similar natural way of grouping terms to make blocks of length that is a multiple of r will allow us to write f as the sum of

$$\begin{cases} 2^{d_1+t} & \text{if} \quad d_0 + 2d_1 \leq 3r \\ 2^{t_1+t_2} & \text{if} \quad d_0 + 2d_1 \geq 3r \end{cases}$$

block central polynomials of the type $\begin{cases} gh & \text{if} \quad d_0 + 2d_1 \leq 3r \\ \sum h_i g_i \bar{h} i & \text{if} \quad d_0 + 2d_1 \geq 3r \end{cases}$

In the case when $d_0 + 2d_1 = 3r$, f is written as a sum of exactly 2^r block central polynomials.

Case 6: $d_1 \geq r, d_2 \leq r$



Similar to last case and f can be written as the sum of

$$\begin{cases} 2^{d_2+t} & \text{if} \quad d_0 + 2d_2 \leq 3r \\ 2^{t_1+t_2} & \text{if} \quad d_0 + 2d_2 \geq 3r \end{cases}$$

block central polynomials of the type $\begin{cases} gh & \text{if} \quad d_0 + 2d_2 \leq 3r \\ \sum h_i g_i \bar{h} i & \text{if} \quad d_0 + 2d_2 \geq 3r \end{cases}$

Similar to last time, f is written as a sum of exactly 2^r block central polynomials when $d_0 + 2d_2 = 3r$.

This completes the proof. □

Proof of 3. We want to show that $\mathsf{LID}_r = \mathsf{Perm}(M)$ where M is a matrix of dimension $4r \times 4r$ whose non-zero entries are variables z_0, z_1.

For $j \in \{0, 1\}$, let D_j be a $2r \times 2r$ matrix with z_j on the diagonal and zero everywhere else. The matrix M is defined as:

$$M = \begin{bmatrix} D_0 & D_1 \\ D_1 & D_0 \end{bmatrix}.$$

Then

$$\mathsf{Perm}(M) = \sum_{\sigma} M_{1\sigma(1)} \ldots M_{4r\sigma(4r)}.$$

Further, the σs for which $M_{1,\sigma(1)} \ldots M_{4r,\sigma(4r)} \neq 0$ have the following property:

$$\sigma(i) = i \Rightarrow \sigma(2r+i) = 2r+i$$

and

$$\sigma(i) = 2r+i \Rightarrow \sigma(2r+i) = i.$$

Thus by the structure of M, for every $i \in [2r]$ $M_{i,\sigma(i)} = M_{2r+i,\sigma(2r+i)}$ and as we go over all possible values of σ, every value of $\{z_e\}_{e \in \{0,1\}^{2r}}$ is covered. This gives the required result: $\mathsf{LID}_r = \mathsf{Perm}(M)$. $\square$