Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 83 (2017)

Weights at the Bottom Matter When the Top is Heavy

Arkadev Chattopadhyay*[1] and Nikhil S. Mande†[1]

[1] *School of Technology and Computer Science, TIFR, Mumbai*

Abstract

Proving super-polynomial lower bounds against depth-2 threshold circuits of the form $\mathsf{THR} \circ \mathsf{THR}$ is a well-known open problem that represents a frontier of our understanding in boolean circuit complexity. By contrast, exponential lower bounds on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits were shown by Razborov and Sherstov [36] even for computing functions in depth-3 AC^0. Yet, no separation among the two depth-2 threshold circuit classes was known.

In this work, we provide the first exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$ answering an open problem explicitly posed by Hansen and Podolskii [21]. We achieve this by showing a simple function f on n bits, which is a linear-size decision list of 'Equalities', has *sign rank* $2^{\Omega(n^{1/4})}$. It follows, by a well-known result, that $\mathsf{THR} \circ \mathsf{MAJ}$ circuits need size $2^{\Omega(n^{1/4})}$ to compute f, while it is not difficult to observe that f can be computed by $\mathsf{THR} \circ \mathsf{THR}$ circuits of only linear size. Our result, thus, suggests that the sign rank method alone is unlikely to prove strong lower bounds against $\mathsf{THR} \circ \mathsf{THR}$ circuits.

Additionally, our function f yields new communication complexity class separations. In particular, f lies in the class P^{MA}. As f has large sign-rank, this shows that $\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}$, resolving a recent open problem of Göös, Pitassi and Watson [19].

The main technical ingredient of our work is to prove a strong sign rank lower bound for an XOR function. This requires novel use of approximation theoretic tools.

*Partially supported by a Ramanujan fellowship of the DST. arkadev.c@tifr.res.in

†Partially supported by a TCS fellowship. nikhil.mande@tifr.res.in

ISSN 1433-8092

1 Introduction

Understanding the computational power of constant-depth, unbounded fan-in threshold circuits is one of the most fundamental open problems in theoretical computer science. Despite several years of intensive research [1, 20, 24, 15, 35, 6, 28, 29, 13, 14, 36, 21, 22, 27, 11], we still do not have strong lower bounds against depth-3 or depth-2 threshold circuits, depending on how we define threshold gates. The most natural definition of such a gate, denoted by $\mathsf{THR}_{\mathbf{w}}$, is just a linear halfspace induced by the real weight vector $\mathbf{w} = (w_0, w_1, \ldots, w_n) \in \mathbb{R}^{n+1}$. In other words, on an input $x \in \{-1, 1\}^n$,[1]

$$\mathsf{THR}_{\mathbf{w}}(x) = \operatorname{sgn}\left(w_0 + \sum_{i=1}^{n} w_i x_i\right).$$

The class of all boolean functions that can be computed by circuits of depth d and polynomial size, comprising such gates, is denoted by LT_d. The seminal work of Minsky and Papert [31] showed that a simple function, Parity, is not in LT_1. While it is not hard to verify that Parity is in LT_2, an outstanding problem is to exhibit an explicit function that is not in LT_2. This problem is now a well-identified frontier for research in circuit complexity.

A natural question is how large the individual weights in the weight vector $\mathbf{w}$ need to be if we allow just integer weights. It was well-known [32] that for every threshold gate with n inputs, there exists a threshold representation for it that uses only integer weights of magnitude at most $2^{O(n \log n)}$. While proving a $2^{\Omega(n)}$ lower bound is not very difficult, a matching $2^{\Omega(n \log n)}$ lower bound was shown only in the nineties by Håstad [23]. Understanding the power of large weights vs. small weights in the more general context of small-depth circuits has attracted attention by several works [1, 15, 40, 21, 22, 35, 20, 25, 16]. More precisely, let $\widehat{LT}_d$ denote the class of boolean functions that can be computed by polynomial size and depth d circuits comprising only threshold gates each of whose integer weights are polynomially bounded in n, the number of input bits to the circuit. Interestingly, improving upon an earlier line of work [9, 34, 40], Goldmann, Håstad and Razborov [15] showed, among other things, that $LT_d \subseteq \widehat{LT}_{d+1}$. It also remains open to exhibit an explicit function that is not in $\widehat{LT}_3$. This is a very important frontier, as the work of Yao [41] and Beigel and Tarui [4] show that the entire class ACC is contained in the class of functions computable by quasi-polynomial size threshold circuits of small weight and depth three. By contrast, the relatively early work of Hajnal et al. [20] established the fact that the Inner-Product modulo 2 function (denoted by IP), that is easily seen to be in $\widehat{LT}_3$, is not in $\widehat{LT}_2$. Summarizing, we have $\widehat{LT}_2 \subseteq LT_2 \subseteq \widehat{LT}_3$. Where precisely between $\widehat{LT}_2$ and $\widehat{LT}_3$ do current techniques for lower bounds stop working?

In search of the answer to the above question, researchers have investigated the finer structure of depth-2 threshold circuits, and this has generated many new techniques that are interesting in their own right. Recall the Majority function, denoted by MAJ, that outputs 1 precisely when the majority of its n input bits are set to 1. It is simple to verify that $\widehat{LT}_2 = \mathsf{MAJ} \circ \mathsf{MAJ}$. Goldmann et al. [15] proved two very interesting results. First, they showed that the class $\mathsf{MAJ} \circ \mathsf{MAJ}$ and $\mathsf{MAJ} \circ \mathsf{THR}$ are identical, i.e. weights of the bottom gates do not matter if the top gate is allowed only polynomial weight. Second, they

[1] Throughout this paper, we consider the input domain to be $\{-1, 1\}^n$, rather than $\{0, 1\}^n$.

showed that $\mathsf{MAJ} \circ \mathsf{MAJ}$ is strictly contained in the class $\mathsf{THR} \circ \mathsf{MAJ}$, i.e. the weight at the top does matter if the bottom weights are restricted to be polynomially bounded in the input length. This revealed the following structure:

$$\widehat{LT}_2 = \mathsf{MAJ} \circ \mathsf{THR} \subsetneq \mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2 \subseteq \widehat{LT}_3.$$

This raised the following two questions: how powerful is the class $\mathsf{THR} \circ \mathsf{MAJ}$ and how does one prove lower bounds on the size of such circuits?

In a breakthrough work, Forster [13] showed that IP requires size $2^{\Omega(n)}$ to be computed by $\mathsf{THR} \circ \mathsf{MAJ}$ circuits. This yielded an exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\widehat{LT}_3$. This also meant that at least one of the two containments $\mathsf{THR} \circ \mathsf{MAJ} \subseteq LT_2$ and $LT_2 \subseteq \widehat{LT}_3$ is strict. While it is quite possible that both of them are strict, until now no progress on this question was made. In particular, Amano and Maruoka [1] and Hansen and Podolskii [21] state that separating $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR} = LT_2$ would be an important step for shedding more light on the structure of depth-2 boolean circuits. However, as far as we know, there was no clear target function identified for the purpose of separating the two classes.

In this work, we exhibit such a function and prove that it achieves the desired separation. To state our result formally, consider the following function that is a simple adaptation of a well-known function called $\mathsf{ODD\text{-}MAX\text{-}BIT}$, which we denote by OMB_l^0: it outputs -1 precisely if the rightmost bit that is set to 1 occurs at an odd index. It is simple to observe that it is a linear threshold function:

$$\mathsf{OMB}_l^0(x) = -1 \iff \sum_{i=1}^{l} (-1)^{i+1} 2^i (1 + x_i) \geq 0.5.$$

Let $f_m \circ g_n : \{-1,1\}^{m \times n} \to \{-1,1\}$ be the composed function on mn input bits, where each of the m input bits to the outer function f is obtained by applying the inner function g to a block of n bits. We show the following:

Theorem 1.1. *Let F_n be defined on $n = 2l^{4/3} + 2l \log l$ bits as $\mathsf{OMB}_l^0 \circ \mathsf{OR}_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$. Every $\mathsf{THR} \circ \mathsf{MAJ}$ circuit computing F_n needs size $2^{\Omega\left(n^{1/4}\right)}$.*

To show that the above suffices to provide us with the separation of threshold circuit classes, we first observe the following: for each $x \in \{-1,1\}^n$, let $\mathsf{ETHR}_{\mathbf{w}}(x) = -1 \iff w_0 + w_1 x_1 + \cdots + w_n x_n = 0$. Thus, ETHR gates are also called exact threshold gates. By first observing that every function computed by a circuit of the form $\mathsf{THR} \circ \mathsf{OR}$ can also be computed by a circuit of the form $\mathsf{THR} \circ \mathsf{AND}$ with a linear blow-up in size, it follows that F_n can be computed by linear size circuits of the form $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$. Observe that each $\mathsf{AND} \circ \mathsf{XOR}_2$ is computable by an ETHR gate. Hence, F_n is in $\mathsf{THR} \circ \mathsf{ETHR}$, a class that Hansen and Podolskii [21] showed is identical to the class $\mathsf{THR} \circ \mathsf{THR}$. Thus, Theorem 1.1 yields the following fact:

Corollary 1.2. *The function F_n (exponentially) separates the class $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR}$.*

Göös [17] pointed out that the function F_n has another interesting application in delineating the reach of current lower-bound techniques against communication complexity classes. These classes were first introduced in the seminal work of Babai, Frankl and Simon [2] as an analogue to the standard Turing complexity classes. While *unconditionally* understanding the relative power of (non)determinism and randomness in the context of Turing machines seem well beyond current techniques, Babai et al. hoped that making progress in the mini-world of communication protocols would be less difficult. Indeed, we understand a lot more in the latter world. For instance, the class P^{cc} is *strictly* contained in both BPP^{cc} and NP^{cc}, while BPP^{cc} and NP^{cc} are provably different. A major goal, set by Babai et al., is to prove lower bounds against the polynomial hierarchy for which the simple function of Inner-Product has long been identified as a target. Unfortunately, we cannot even exhibit a function that is not in the second level of the hierarchy, which remains a long-standing open problem.

Henceforth, we often drop cc from the superscript for convenience since we deal exclusively with communication complexity classes rather than Turing machine classes. The strongest lower bound technique currently known in communication complexity is the *sign-rank* method, discussed before. Functions whose communication matrix of dimension $2^n \times 2^n$ have sign rank upper bounded by a quasi-polynomial in n correspond exactly to the complexity class UPP. The lower bound on the sign rank by Razborov and Sherstov [36] implied that PH (in fact, $\Pi_2\mathsf{P}$) contains functions with large sign rank, rendering the sign-rank technique essentially useless to prove lower bounds against the second level. A natural question is to understand until where, between the first and second level, does the sign-rank method suffice to prove lower bounds.

Indeed, there is a rich landscape of communication complexity classes below the second level as discussed in a recent, almost exhaustive survey by Göös, Pitassi and Watson [19]. Göös et al. conjectured the two classes, $\mathsf{AM} \cap \mathsf{coAM}$ and $\mathsf{S_2P}$ to be two potentially incomparable frontier classes for the sign-rank method to fail. In a very recent work, Bouland et al. [5] showed that there is a *partial function* in $\mathsf{AM} \cap \mathsf{coAM}$ which has large sign rank, confirming one part of the conjecture. Our result confirms the second part by exhibiting the first *total function* in a complexity class contained, plausibly strictly, in Π_2P that has large sign rank. More precisely, it is not difficult to show that our function F_n is in P^{MA}, a class below $\mathsf{S_2P}$. This yields the following new result:

Theorem 1.3. *The total function* $\mathsf{OMB}_l^0 \circ \mathsf{OR}_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$ *witnesses the following separation.*

$$\mathsf{P}^{\mathsf{MA}} \not\subseteq \mathsf{UPP}.$$

It is interesting to recall that, on the other hand, $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$. This fact combined with Theorem 1.3 shows that P^{MA} is right on the frontier between what we understand and what we do not. Thus, efforts to prove lower bounds against P^{MA} protocols should be a natural program for advancing the set of currently known techniques.

1.1 Our Techniques and Related Work

The starting point for our lower bound is the same as for all known lower bounds (see, for example, [13, 36, 8]) on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits. We strive to prove a lower bound on a quantity called the *sign rank* of our target function F. Given a partition of the input bits of F into two parts X, Y, consider the real matrix M_F, given by $M_F[x,y] = F(x,y)$ for each $x \in \{-1,1\}^{|X|}, y \in \{-1,1\}^{|Y|}$. A real matrix sign represents M_F if each of its entries agrees in sign with the corresponding entry of M_F. The sign rank of M_F (also informally called sign rank of F, when the input partition is clear from the context) is the rank of a minimal rank matrix that sign represents it. It is not hard to see that the sign rank of a function F computed by $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size s is at most $O(n \cdot s)$. This sets a target of proving a strong lower bound on the sign rank of F for showing that it is hard for $\mathsf{THR} \circ \mathsf{MAJ}$.

Sign rank has a matrix-rigidity flavor to it, and therefore is quite non-trivial to bound. Forster's Theorem [13] (see Theorem 2.8) shows that the sign rank of a matrix can be lower bounded by appropriately upper-bounding its spectral norm. This is enough to lower bound the sign rank of functions like IP as the corresponding matrices are orthogonal and therefore have relatively small spectral norm. However for other functions F, the spectral norm of the sign matrix M_F can be large. This is true, for example, for many functions in AC^0. In a beautiful work, Razborov and Sherstov [36] showed that Forster's basic method can be adapted to prove exponentially strong lower bounds on the sign rank of such a function F. However, our first problem is devising an F that is in $\mathsf{THR} \circ \mathsf{THR}$ *and* plausibly has high sign rank. On this, we are guided by another interpretation of sign rank, due to Paturi and Simon [33]. Paturi and Simon introduced a model of two-party randomized communication, called the unbounded error model. In this model, Alice and Bob have to give the right answer with probability just greater than $1/2$ on every input. This is the strongest known two-party model against which we know how to prove lower bounds. [33] showed that the sign rank of the communication matrix of F essentially characterizes its unbounded error complexity.

Why should some function $F \in \mathsf{THR} \circ \mathsf{THR}$ have large unbounded-error complexity? The natural protocol one is tempted to use is the following. Sample a sub-circuit of the top gate with a probability proportional to its weight. Then, use the best protocol for the sampled bottom THR gate. Note that for any given input x, with probability $1/2 + \varepsilon$, one samples a bottom gate that agrees with the value of $F(x)$. Here, ε can be inverse exponentially small in the input size. Thus, if we had a small cost randomized protocol for the bottom THR gate that errs with probability significantly less than ε we would have a small cost unbounded-error protocol for F. Fortunately for us (the lower bound prover), there does not seem to exist any such efficient randomized protocol for THR, when $\varepsilon = 1/2^{n^{\Omega(1)}}$.

Taking this a step further, one could hope that the bottom gates could be any function that is hard to compute with such tiny error ε. The simplest such canonical function is Equality (denoted by EQ). Therefore, a plausible target is $\mathsf{THR} \circ \mathsf{EQ}$. This still turns out to be in $\mathsf{THR} \circ \mathsf{THR}$ as $\mathsf{EQ} \in \mathsf{ETHR}$. Moreover, EQ has a nice composed structure. It is just $\mathsf{AND} \circ \mathsf{XOR}$, which lets us re-express our target as $F = \mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}$, for some top THR that is 'suitably' hard; hard so that the sign rank of F becomes large! At this point, we view F as an XOR function whose outer function, g, needs to have sufficiently good analytic properties for us to prove that $g \circ \mathsf{XOR}$ has high sign rank.

We are naturally drawn to the work of Razborov and Sherstov [36] for inspiration as they bound the sign rank of a three-level composed function as well. They showed that $\mathsf{AND} \circ \mathsf{OR} \circ \mathsf{AND}_2$, an AND *function*, has high sign rank. They exploited the fact that AND functions embed inside them *pattern matrices*, which have nice convenient spectral properties as observed in [38]. These spectral properties dictate them to look for an *approximately smooth orthogonalizing* distribution w.r.t which the outer function $f = \mathsf{AND} \circ \mathsf{OR}$ has zero correlation with small degree parities. This gives rise naturally to an LP that seeks to maximize the *smoothness* of the distribution under the constraints of low-degree orthogonality. The main technical challenge that Razborov and Sherstov overcome is the analysis of the dual of this LP using and building appropriate approximation theoretic tools. We take cue from this work and follow its general framework of analyzing the dual of a suitable LP. However, as we are forced to work with an XOR function, there are new challenges that crop up. This is expected, for if we take the same outer function of $\mathsf{AND} \circ \mathsf{OR}$, then the resulting XOR function has small sign rank. Indeed, this remains true even if one were to harden the outer function to $\mathsf{MAJ} \circ \mathsf{OR}$. This is simply because $\mathsf{OR} \circ \mathsf{XOR}$ is non-equality. A simple efficient UPP protocol for $\mathsf{MAJ} \circ \mathsf{EQ}$ exists: pick a random EQ and then execute a protocol of cost $O(\log n)$ that solves this EQ with error less than $1/n^2$.

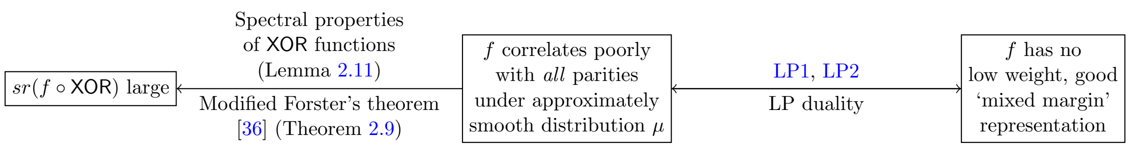

Figure 1: Approximation theoretic hardness of f implies large sign rank of $f \circ \mathsf{XOR}$ (Theorem 3.1).

Figure 1 describes a general passage from the problem of lower bounding the sign rank of a function $f \circ \mathsf{XOR}$ to a sufficient problem of proving an approximation theoretic hardness property of f: namely f has no good 'mixed margin' representation by low weight polynomials. Theorem 3.1 states the precise connection between the approximation theoretic problem for f and the sign rank of $f \circ \mathsf{XOR}$. This passage is made possible by using well known spectral properties of XOR functions and LP duality. This is similar to the works of Razborov and Sherstov [36] and Sherstov [39] where the authors used spectral properties of pattern matrices. The key difference between our work and theirs is in the nature of the approximation theoretic problem that we end up with. While both previous works had to rule out good *low degree* representations, our Theorem 3.1 stipulates us to rule out good *low weight* representations of otherwise *unrestricted degree*.

A similar flavored but simpler problem had been tackled in a recent work of the authors [10], which characterized the *discrepancy* of XOR functions. Roughly speaking, in that work, the primal program constrained the distribution μ such that f correlates poorly with all parities w.r.t μ. However, there was no smoothness constraint imposed on μ in [10], which is what we are constrained to have in this work. Analyzing this combination of high degree parity constraints and the smoothness constraints is the main new technical challenge that our work addresses.

Our main technical contribution is Theorem 5.1 which shows that the function $\mathsf{OMB}^0 \circ \mathsf{OR}$ has no low weight, good 'mixed margin' polynomial representation. We prove this by a novel combination of ideas, sketched in Figure 2, that differs entirely from the Razborov-Sherstov analysis. We believe this result to be of independent interest in the area of analysis of Boolean functions.

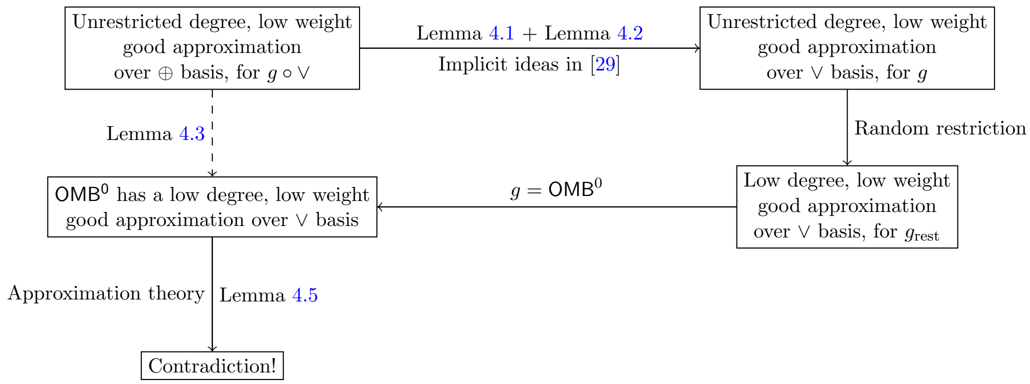

Figure 2: Approximation theoretic analysis (Theorem 5.1)

The first step in our method is to borrow an averaging idea from Krause and Pudlák [29] to show the following: a low weight good approximation of $g \circ \mathsf{OR}_m$ by a polynomial p over the Parity (Fourier) basis implies that there exists a low weight polynomial q over the OR basis which approximates g as well as p approximates $g \circ \mathsf{OR}_m$, save an additive loss of at most 2^{-m}. This transformation to q is very useful because although it is still unrestricted in degree, it is over the OR basis, that is vulnerable to random restrictions. Indeed, in the next step, we hit q with random restrictions to reduce its degree. At this point, we extract a low weight *and* low degree polynomial r that still approximates g_{rest}, the restriction of g. We now appeal to interesting properties of the $\mathsf{ODD\text{-}MAX\text{-}BIT}$ function by setting $g = \mathsf{OMB}^0$. First, we observe that OMB^0 on l bits, under random restrictions, retains its hardness as it contains OMB^0 on $l/8$ bits with high probability. Next, we show that OMB^0 does not have low degree good approximations by appealing to classical approximation theoretic tools, suitably modifying the argument of Buhrman et al. [7] and Beigel [3]. This provides us with the required contradiction.

2 Preliminaries

In this section, we provide the necessary preliminaries.

Definition 2.1 (Threshold functions)**.** *A function $f : \{-1,1\}^n \to \{-1,1\}$ is called a* linear threshold function *if there exist integer weights $a_0, a_1, \ldots, a_n$ such that for all inputs $x \in \{-1,1\}^n$, $f(x) = \mathrm{sgn}(a_0 + \sum_{i=1}^n a_i x_i)$. Let THR denote the class of all such functions.*

Definition 2.2 (Exact threshold functions)**.** *A function $f : \{-1,1\}^n \to \{-1,1\}$ is called an* exact threshold function *if there exist reals $w_1, \ldots, w_n, t$ such that*

$$f(x) = -1 \iff \sum_{i=1}^{n} w_i x_i = t.$$

Let ETHR denote the class of exact threshold functions.

Hansen and Podolskii [21] showed the following.

Theorem 2.3 (Hansen and Podolskii [21])**.** *If a function $f : \{-1,1\}^n \to \{-1,1\}$ can be represented by a* $\mathsf{THR} \circ \mathsf{ETHR}$ *circuit of size s, then it can be represented by a* $\mathsf{THR} \circ \mathsf{THR}$ *circuit of size $2s$.*

For the sake of completeness and clarity, we provide the proof below.

Proof. Let h be an exact threshold function with the representation $\sum_{i=1}^n w_i x_i = t$. There exists an $\varepsilon_h > 0$ such that $\sum_{i=1}^n w_i x_i > t \implies \sum_{i=1}^n w_i x_i > t + \varepsilon_h$. Consider a $\mathsf{THR} \circ \mathsf{ETHR}$ circuit of size s which computes f. Say it computes $\mathrm{sgn}(c_0 + \sum_{i=1}^s c_i f_i)$, where f_i's have exact threshold representations $\sum_{j=1}^n w_{i,j} x_j = t_i$, respectively. Consider the $\mathsf{THR} \circ \mathsf{THR}$ circuit of size $2s$, given by $\mathrm{sgn}\left(\sum_{i=1}^s c_i(g_{i,1} - g_{i,2} + 1)\right)$, where g_i's are threshold functions with representations as follows.

$$g_{i,1} = 1 \iff \sum_{j=1}^n w_{i,j} x_j \geq t_i,$$

$$g_{i,2} = 1 \iff \sum_{j=1}^n w_{i,j} x_j \geq t_i + \varepsilon_{f_i}.$$

It is easy to verify that this circuit computes f. □

Remark 2.4. *In fact, Hansen and Podolskii [21] showed that the circuit class* $\mathsf{THR} \circ \mathsf{THR}$ *is identical to the circuit class* $\mathsf{THR} \circ \mathsf{ETHR}$*. However, we do not require the full generality of their result.*

We now note that any function computable by a $\mathsf{THR} \circ \mathsf{OR}$ circuit can be computed by a $\mathsf{THR} \circ \mathsf{AND}$ circuit without a blowup in the size.

Lemma 2.5. *Suppose $f : \{-1,1\}^n \to \{-1,1\}$ can be computed by a* $\mathsf{THR} \circ \mathsf{OR}$ *circuit of size s. Then, f can be computed by a* $\mathsf{THR} \circ \mathsf{AND}$ *circuit of size s.*

Proof. Consider a $\mathsf{THR} \circ \mathsf{OR}$ circuit of size s, computing f, say

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^s w_i \bigvee_{j \in S_i} x_j\right).$$

Note that

$$\sum_{i=1}^s w_i \bigvee_{j \in S_i} x_j = \sum_{i=1}^s -w_i \bigwedge_{j \in S_i} x_j^c.$$

Thus, $\mathrm{sgn}\left(\sum_{i=1}^s -w_i \bigwedge_{j \in S_i} x_j^c\right)$ is a $\mathsf{THR} \circ \mathsf{AND}$ representation of f, of size s. □

Definition 2.6 (OR polynomials)**.** *Define a function $p : \{-1,1\}^n \to \mathbb{R}$ of the form $p(x) = \sum_{S \subseteq [n]} a_S \bigvee_{i \in S} x_i$ to be an* OR polynomial*. Define the weight of p (in the* OR *basis) to be $\sum_{S \subseteq [n]} |a_S|$, and its degree to be $\max_{S \subseteq [n]}\{|S| : a_S \neq 0\}$.*

Remark 2.7. *In the above definition, '*OR *monomials' are defined as follows.*

$$\bigvee_{i\in S} x_i = \begin{cases} 1 & x_i = 1\ \forall i \in S \\ -1 & \textit{otherwise} \end{cases}$$

Unless mentioned otherwise, all polynomials we consider will be over the parity *basis.*

Define the *sign rank* of a real matrix $A = [A_{ij}]$, denoted by $sr(A)$ to be the least rank of a matrix $B = [B_{ij}]$ such that $A_{ij}B_{ij} > 0$ for all (i, j) such that $A_{ij} \neq 0$.

Forster [13] proved the following relation between the sign rank of a $\{\pm 1\}$ valued matrix and its spectral norm.

Theorem 2.8 (Forster [13])**.** *Let $A = [A_{xy}]_{x\in X, y\in Y}$ be a $\{\pm 1\}$ valued matrix. Then,*

$$sr(A) \geq \frac{\sqrt{|X||Y|}}{||A||}.$$

We require the following generalization of Forster's theorem by Razborov and Sherstov [36].

Theorem 2.9 (Razborov and Sherstov [36])**.** *Let $A = [A_{xy}]_{x\in X, y\in Y}$ be a real valued matrix with $s = |X||Y|$ entries, such that $A \neq 0$. For arbitrary parameters $h, \gamma > 0$, if all but h of the entries of A satisfy $|A_{xy}| \geq \gamma$, then*

$$sr(A) \geq \frac{\gamma s}{||A||\sqrt{s} + \gamma h}.$$

The following lemma by Forster et al. [14] tells us that functions with efficient THR∘MAJ representations have small sign rank.

Lemma 2.10 (Forster et al. [14])**.** *Let $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$ be a boolean function computed by a* THR ∘ MAJ *circuit of size s. Then,*

$$sr(M_F) \leq sn,$$

where M_F denotes the communication matrix of F.

For the purpose of this paper, we abuse notation, and use $sr(F)$ and $sr(M_F)$ interchangeably, to denote the sign rank of M_F.

Consider the vector space of functions from $\{-1, 1\}^n$ to $\mathbb{R}$, equipped with the following inner product.

$$\langle f, g\rangle = \mathop{\mathbb{E}}_{x\in\{-1,1\}^n} [f(x)g(x)] = \frac{1}{2^n} \sum_{x\in\{-1,1\}^n} f(x)g(x).$$

Define 'characters' χ_S for every $S \subseteq [n]$ by $\chi_S(x) = \prod_{i\in S} x_i$. The set $\{\chi_S : S \subseteq [n]\}$ forms an orthonormal basis for this vector space. Thus, every $f : \{-1, 1\}^n \to \mathbb{R}$ can be uniquely written as $f = \sum\limits_{S\subseteq[n]} \widehat{f}(S)\chi_S$, where

$$\widehat{f}(S) = \langle f, \chi_S\rangle = \mathop{\mathbb{E}}_{x\in\{-1,1\}^n} [f(x)\chi_S(x)]\,. \tag{1}$$

Define $\mathrm{mon}(f) = \Big|S \subseteq [n] : \widehat{f}(S) \neq 0\Big|$.

Lemma 2.11 (Folklore)**.** *For any function $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{x \in \{-1,1\}^n}[|f(x)|] \geq \max_{S \subseteq [n]} \left| \widehat{f}(S) \right|.$$

Fact 2.12 (Plancherel's identity)**.** *For any functions $f, g : \{-1,1\}^n \to \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{x \in \{-1,1\}^n}[f(x)g(x)] = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{g}(S).$$

Definition 2.13 (Signed monomial complexity)**.** *The* signed monomial complexity *of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted by $\mathrm{mon}_{\pm}(f)$, is the minimum number of monomials required by a polynomial p to sign represent f on all inputs.*

Remark 2.14. *Note that the signed monomial complexity of a function f exactly corresponds to the minimum size Threshold of Parity circuit computing f.*

In the model of communication we consider, two players, say Alice and Bob, are given inputs $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ for some finite input sets $\mathcal{X}, \mathcal{Y}$. They are given access to *private* randomness and wish to compute a given function $F : \mathcal{X} \times \mathcal{Y} \to \{-1,1\}$. We will use $\mathcal{X} = \mathcal{Y} = \{-1,1\}^n$ for the purposes of this paper. Alice and Bob communicate using a randomized protocol which has been agreed upon in advance. The cost of the protocol is the maximum number of bits communicated in the worst case input and coin toss outcomes. A protocol Π computes F with advantage ε if the probability of F agreeing with Π is at least $1/2 + \varepsilon$ for all inputs. We denote the cost of the best such protocol to be $R_\varepsilon(F)$. Note here that we deviate from standard notation (used in [30], for example). Unbounded error communication complexity was introduced by Paturi and Simon [33], and is defined as follows.

$$\mathsf{UPP}(F) = \inf_{\varepsilon > 0}(R_\varepsilon(F)).$$

This measure gives rise to the following natural communication complexity class, as introduced by Babai et al. [2].

Definition 2.15.

$$\mathsf{UPP}^{cc}(F) \equiv \{F : \mathsf{UPP}(F) = \mathrm{polylog}(n)\}.$$

Paturi and Simon [33] showed an equivalence between $\mathsf{UPP}(F)$ and the sign rank of M_F, where M_F denotes the communication matrix of F.

Theorem 2.16 (Paturi and Simon [33])**.** *For any function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$,*

$$\mathsf{UPP}(F) = \log sr(M_F) \pm O(1).$$

The following lemma characterizes the spectral norm of the communication matrix of XOR functions.

Lemma 2.17 (Folklore)**.** *Let $f : \{-1,1\}^n \to \mathbb{R}$ be any real valued function and let M denote the communication matrix of $f \circ \mathsf{XOR}$. Then,*

$$||M|| = 2^n \cdot \max_{S \subseteq [n]} \left| \widehat{f}(S) \right|.$$

Finally, we require the following well-known lemma by Minsky and Papert [31].

Lemma 2.18 (Minsky and Papert [31]). *Let $p : \{-1,1\}^n \to \mathbb{R}$ be any symmetric real polynomial of degree d. Then, there exists a univariate polynomial q of degree at most d, such that for all $x \in \{-1,1\}^n$,*

$$p(x) = q(\#1(x))$$

where $\#1(x) = |\{i \in [n] : x_i = 1\}|$.

3 Sign rank to polynomial approximation

In this section, we prove how a certain approximation theoretic hardness property of f implies that the sign rank of $f \circ \mathsf{XOR}$ is large, as outlined in Figure 1.

Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function, $\delta > 0$ be a parameter, and X be any subset of $\{-1,1\}^n$. Consider the following linear program.

$$\text{(LP1)}\quad \boxed{\begin{array}{llll} \text{Variables} & \varepsilon, \{\mu_x : x \in \{-1,1\}^n\} & & \\ \text{Minimize} & \varepsilon & & \\ \text{s.t.} & \left|\sum_x \mu(x) f(x) \chi_S(x)\right| & \leq \varepsilon & \forall S \subseteq [n] \\ & \sum_x \mu(x) & = 1 & \\ & \varepsilon \geq 0 & & \\ & \mu(x) \geq \frac{\delta}{2^n} & & \forall x \in X \\ & \mu(x) \geq 0 & & \forall x \in \{-1,1\}^n \end{array}}$$

The first constraint above specifies that correlation of f against *all parities* need to be small w.r.t a distribution μ. The second last constraint enforces the fact that μ is 'δ-smooth' over the set X. As we had indicated earlier in Section 1.1, these constraints make analyzing the LP challenging.

Standard manipulations (as in [10], for example) and strong linear programming duality reveal that the optimum of the above linear program equals the optimum of the following program. Let OPT denote the optima of these programs.

$$\text{(LP2)}\quad \boxed{\begin{array}{llll} \text{Variables} & \Delta, \{\alpha_S : S \subseteq [n]\}, \{\xi_x : x \in X\} & & \\ \text{Maximize} & \Delta + \frac{\delta}{2^n} \sum_{x \in X} \xi_x & & \\ \text{s.t.} & f(x) \sum_{S \subseteq [n]} \alpha_S \chi_S(x) & \geq \Delta & \forall x \in \{-1,1\}^n \\ & f(x) \sum_{S \subseteq [n]} \alpha_S \chi_S(x) & \geq \Delta + \xi_x & \forall x \in X \\ & \sum_{S \subseteq [n]} |\alpha_S| & \leq 1 & \\ & \Delta \in \mathbb{R} & & \\ & \alpha_S \in \mathbb{R} & & \forall S \subseteq [n] \\ & \xi_x \geq 0 & & \forall x \in X \end{array}}$$

The first constraint of the above program indicates that the variable Δ represents the worst margin guaranteed to exist at all points. The second constraint says that at each point x over the smooth set X, the dual polynomial has to better the worst margin by at least ξ_x. If OPT is large, then it means that on average, the dual polynomial did significantly better than the worst margin. It is for this reason we call the optimum the 'mixed margin' as mentioned in Section 1.1.

We now show that upper bounding OPT for any function f yields sign rank lower bounds against $f \circ \mathsf{XOR}$. The proof idea is depicted in Figure 1.

Theorem 3.1. *Let $f : \{-1,1\}^n \to \mathbb{R}$ be any function. For any $\delta > 0$ and $X \subseteq \{-1,1\}^n$, suppose the value of the optimum of (LP2) (and hence (LP1)) is at most OPT. Then,*

$$sr(f \circ \mathsf{XOR}) \geq \frac{\delta}{OPT + \delta \cdot \frac{|X^c|}{2^n}}.$$

Proof. By (LP1), there exists a distribution μ on $\{-1,1\}^n$ such that $\mu(x) \geq \frac{\delta}{2^n}$ for all $x \in X$, and $\max_{S \subseteq [n]} |\widehat{f\mu}(S)| \leq \frac{OPT}{2^n}$. By Lemma 2.17,

$$||M_{f\mu \circ \mathsf{XOR}}|| = 2^n \cdot \max_{S \subseteq [n]} \left|\widehat{f\mu}(S)\right| \leq OPT.$$

Each $x \in X$ contributes to 2^n entries of $M_{f\mu \circ \mathsf{XOR}}$ whose absolute value is at least δ. Plugging values in Theorem 2.9, we obtain

$$sr(f \circ \mathsf{XOR}) \geq sr(f\mu \circ \mathsf{XOR}) \geq \frac{\frac{\delta}{2^n} \cdot 2^{2n}}{OPT \cdot 2^n + \frac{\delta}{2^n} \cdot 2^n \cdot |X^c|} = \frac{\delta}{OPT + \delta \cdot \frac{|X^c|}{2^n}},$$

which proves the desired sign rank lower bound.

□

Theorem 3.1 provides us with a technique for proving sign rank lower bounds against XOR functions. In Section 4, we show an upper bound on OPT when $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3} + \log l}$. In Section 5, we use this along with Theorem 3.1 to prove sign rank lower bounds against a function $f \in \mathsf{THR} \circ \mathsf{THR}$, which yields an exponential separation between the circuit classes $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$, since it is known that sign rank lower bounds against f yields lower bounds on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits computing f.

4 Hardness of approximating $\mathsf{OMB}_l^0 \circ \mathsf{OR}_m$

In this section, we show that $\mathsf{OMB}_l^0 \circ \mathsf{OR}_m$ is hard to approximate in a certain sense for specific choices of l and m, by following the steps as depicted in Figure 2.

We first use an idea from Krause and Pudlák [29] which enables us to work with polynomial approximations for g, given a polynomial approximation for $g \circ \bigvee_m$.

We use the following notation for the following two lemmas. For any set $I \subseteq [l] \times [m]$, define $J \subseteq [l]$ to be the projection of I on $[l]$; $i \in J \iff \exists j,\ x_{i,j} \in I$. For any $y \in \{-1,1\}^l$ and $h : \{-1,1\}^{ml} \to \{-1,1\}$, denote by $\mathbb{E}_y[h(x)]$ the expected value of $h(x)$ with respect to the uniform distribution over all $x \in \{-1,1\}^{ml}$ such that $\bigvee_m(x) = y$.

Lemma 4.1 and Lemma 4.2 represent the first implication in Figure 2. The first tool we use approximates monomials (in the parity basis) by OR functions, with a small error.

Lemma 4.1. *Let l, m be positive integers such that $m > \log l$. For any set $I \subseteq [l] \times [m], y \in \{-1,1\}^l$,*

$$\left| \mathop{\mathbb{E}}_y \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] - \frac{1}{2} - \frac{1}{2} \bigvee_{i \in J} y_i \right| \le 2l2^{-m}.$$

The proof of the above lemma appears in the proof of Lemma 2.3 in [29]. However, we reproduce the proof below for clarity and completeness.

Proof. First observe that for all $y \in \{-1,1\}^l$, and for all x satisfying $\bigvee_m(x) = y$, the monomial corresponding to I equals

$$\bigoplus_{(i,j)\in I} x_{i,j} = \bigoplus_{(i,j)\in I, y_i=-1} x_{i,j}.$$

Let $A = \{i \in [l] : y_i = -1\}$. If $A \cap J = \emptyset$, then

$$\mathop{\mathbb{E}}_y \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] = \bigvee_{i\in J} y_i = 1$$

Else, $\bigvee_{i\in J} y_i = -1$. Also,

$$\mathop{\mathbb{E}}_y \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] = \mathop{\mathbb{E}}_{x\in\{-1,1\}^{(A\cap J)\times[m]} : \bigvee(x) = -1^{|A\cap J|}} \left[\bigoplus_{(i,j)\in I, y_i=-1} x_{i,j} \right] \tag{2}$$

Note that

$$\mathop{\mathbb{E}}_{x\in\{-1,1\}^{(A\cap J)\times[m]}} \left[\bigoplus_{(i,j)\in I, y_i=-1} x_{i,j} \right] = 0 \tag{3}$$

Denote $|A \cap J| = t$. Using Equation (3) and a simple counting argument, the absolute value of the RHS (and thus the LHS) of Equation (2) can be upper bounded as follows (note that we require $1 \le t \le l$ in the following computations).

$$\begin{aligned}
\left| \mathbb{E}_y \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] \right| &\le \frac{2^{mt} - (2^m - 1)^t}{(2^m - 1)^t} \\
&\le \frac{2^{mt} - (2^{mt} - t2^{m(t-1)})}{(2^m - 1)^t}
\end{aligned}$$

(Sum of remaining terms in binomial expansion of $(2^m - 1)^t$ is positive since $m > \log l$)

$$\begin{aligned}
&\le \frac{t \cdot 2^{mt-m}}{2^{mt}/2} && \text{(since } m > \log l\text{)} \\
&\le 2l2^{-m}.
\end{aligned}$$

Hence, for all $y \in \{-1,1\}^l$, we have

$$\left| \mathop{\mathbb{E}}_y \left[\bigoplus_{(i,j)\in I} x_{i,j} \right] - \frac{1}{2} - \frac{1}{2} \bigvee_{i \in J} y_i \right| \le 2l2^{-m}. \tag{4}$$

□

Lemma 4.2. *Let l, m be positive integers such that $m > \log l$, and $g : \{-1,1\}^l \to \{-1,1\}$ be any function. Define $f = g \circ \bigvee_m : \{-1,1\}^{ml} \to \{-1,1\}, \Delta \in \mathbb{R}, e_x \geq 0\ \forall x \in X$, where X denotes the set of all inputs x in $\{-1,1\}^{ml}$ such that $\bigvee_m(x) = -1^l$, and let p be a real polynomial such that*

$$\begin{aligned} \forall x \in \{-1,1\}^{ml}, \qquad & f(x)p(x) \geq \Delta, \\ \forall x \in X, \qquad & f(x)p(x) \geq \Delta + e_x. \end{aligned}$$

Then, there exists an OR *polynomial q, of weight at most $wt(p)$, such that*

$$\begin{aligned} \forall y \in \{-1,1\}^l, \qquad & q(y)g(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m}\right), \\ & q(-1^l)g(-1^l) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\left(2l \cdot 2^{-m}\right). \end{aligned}$$

Proof. Note that for any $y \in \{-1,1\}^l$,

$$\mathop{\mathbb{E}}_{y}[f(x)p(x)] = g(y) \cdot \mathop{\mathbb{E}}_{y}[p(x)] \geq \Delta \tag{5}$$

and

$$\mathop{\mathbb{E}}_{-1^l}[f(x)p(x)] = g(-1^l) \cdot \mathop{\mathbb{E}}_{-1^l}[p(x)] \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|}. \tag{6}$$

Denote the unique multilinear expansion of p by $p = v_0 + \sum_k v_k p_k$, where $p_k(x) = \oplus_{(i,j) \in I_k} x_{i,j}$. Let J_k denote the projection of I_k on $[l]$. Define

$$q = v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{i \in J_k} y_i.$$

Note that

$$wt(q) = wt\left(v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{i \in J_k} y_i\right) = \left|v_0 - \frac{\sum_k v_k}{2}\right| + \sum_k \left|\frac{v_k}{2}\right| \leq wt(p).$$

Thus, using linearity of expectation and Lemma 4.1, Equation (5) and Equation (6) yield that for all $y \in \{-1,1\}^l$,

$$q(y) \cdot g(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m}\right)$$

and

$$q(-1^l) \cdot g(-1^l) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\left(2l \cdot 2^{-m}\right)$$

□

Next, we use random restrictions which reduces the degree of the approximating OR polynomial, at the cost of a small error. In particular, we consider the case when $g = \mathsf{OMB}_l^0$. This represents the dashed implication in Figure 2.

Lemma 4.3. *Let l, m be any positive integers such that $m > \log l$. Let $g_l = \mathsf{OMB}_l^0 : \{-1,1\}^l \to \{-1,1\}, f = g_l \circ \bigvee_m$, and $\Delta, \{e_x \geq 0 : x \in X\}$ (where X is defined as in Lemma 4.2), and p be a real polynomial such that*

$$\begin{cases} \forall x \in \{-1,1\}^{ml}, & f(x)p(x) \geq \Delta \\ \forall x \in X, & p(x) \geq \Delta + e_x \end{cases}$$

Then, for any integer $d > 0$, there exists an OR polynomial $r : \{-1,1\}^{l/8} \to \mathbb{R}$, of degree d and weight at most $wt(p)$, such that

$$\text{For all } y \in \{-1,1\}^{l/8}, \qquad r(y)g_{l/8}(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m} + 2^{-(d-1)}\right)$$
$$\text{and} \qquad r(-1^{l/8}) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\left(2l \cdot 2^{-m} + 2^{-(d-1)}\right).$$

Proof. Lemma 4.2 guarantees the existence of an OR polynomial q, of weight at most $wt(p)$, such that

$$\forall y \in \{-1,1\}^l, \qquad q(y)g_l(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m}\right) \tag{7}$$
$$q(-1^l)g(-1^l) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\left(2l \cdot 2^{-m}\right).$$

Now, set each of the l variables to -1 with probability $1/2$, and leave it unset with probability $1/2$. Call this random restriction r. Any OR monomial of degree at least d gets fixed to -1 with probability $1 - 2^{-d}$. Thus, by linearity of expectation, the expected weight of surviving monomials of degree at least d in q is at most $wt(p) \cdot 2^{-d}$. Let $M|_r$ denote the value of a monomial M after the restriction r. By Markov's inequality,

$$\Pr_r\left[\sum_{M : \deg(M|_r) \geq d} wt(M|_r) > wt(p) \cdot 2^{-d+1}\right] < 1/2.$$

Consider $l/2$ pairs of variables, $\{(x_i, x_{i+1}) : i \in [l/2]\}$ (assume w.l.o.g that l is even). For any pair, the probability that both of its variables remain unset is $1/4$. This probability is independent over pairs. Thus, by a Chernoff bound, the probability that at most $l/16$ pairs remain unset is at most $2^{-\frac{l}{64}}$.

By a union bound, there exists a setting of variables such that at least $l/16$ pairs of variables are left free, and the weight of degree $\geq d$ monomials in q is at most $wt(p) \cdot 2^{-d+1}$. Set the remaining $7l/8$ variables to the value -1. After the restriction, drop the monomials of degree $\geq d$ from q to obtain r, which is now an OR polynomial of degree less than d and weight at most $wt(p)$. Note that the function g_l hit with this restriction just becomes $g_{l/8}$.

Thus, Equation (7) yields the following.

$$\text{For all } y \in \{-1,1\}^{l/8}, \qquad r(y)g_{l/8}(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m} + 2^{-(d-1)}\right)$$
$$\text{and} \qquad r(-1^{l/8}) \geq \Delta + \frac{\sum_{x \in X} e_x}{|X|} - wt(p)\left(2l \cdot 2^{-m} + 2^{-(d-1)}\right).$$

□

4.1 Hardness of OMB^0

In this section, we show that approximating OMB^0 well by a low weight polynomial p is not possible unless the degree of p is large.

We require the following result by Ehlich and Zeller [12] and Rivlin and Cheney [37].

Lemma 4.4 ([12, 37])**.** *The following holds true for any real valued $\alpha > 0$ and integer $k > 0$. Let $p : \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree $d < \sqrt{k/4}$, such that $p(0) \geq \alpha$, and $p(i) \leq 0$ for all $i \in [k]$. Then, there exists $i \in [k]$ such that $p(i) < -2\alpha$.*

We now show that a low degree (multivariate) polynomial of bounded weight cannot represent OMB^0 well. This is our main approximation theoretic lemma.

Lemma 4.5. *Suppose $p : \{-1,1\}^n \to \mathbb{R}$ is a polynomial of degree $d < \sqrt{n/4}$ and $a > 0, b \in \mathbb{R}$ be reals such that $p(-1^n) \geq a$ and $\mathsf{OMB}^0_n(x)p(x) \geq b$ for all $x \in \{-1,1\}^n$. Define*

$$p_{max} = \max_{i \in \{0,\ldots,\lfloor n/10d^2 \rfloor\}} \{2^i a + \left(3 \cdot 2^i - 3\right) b\}.$$

Then, there exists $x \in \{-1,1\}^n$ such that $|p(x)| \geq p_{max}$.

We remark here that a simple consequence of the above lemma is that the weight of a polynomial p (in either the OR basis, or the parity basis) satisfying the assumptions of Lemma 4.5 is at least p_{max}. This property of p suffices for our need.

The proof of Lemma 4.5 will be an iterative argument, inspired by the arguments of Beigel [3] and Buhrman et al. [7]. This captures the last implication in Figure 2.

Claim 4.6. *If a and b are reals such that $a > 0, b \in \mathbb{R}$ and $2^i a + \left(3 \cdot 2^i - 2\right) b < 0$ for some integer $i \geq 0$, then $2^j a + \left(3 \cdot 2^j - 3\right) b < 0$ for all integers $j > i$.*

Proof. Note that since $a > 0$ and $2^i a + \left(3 \cdot 2^i - 2\right) b < 0$, b must be negative. For any $j > i$, write $2^j a + \left(3 \cdot 2^j - 3\right) b = 2^{j-i} \left(2^i a + \left(3 \cdot 2^i - 2\right) b\right) + (2^{j-i+1} - 3)b < 0$. □

Proof of Lemma 4.5. Divide the n variables into $\lfloor n/10d^2 \rfloor$ contiguous blocks of size $10d^2$ each.

Induction hypothesis: For each $i \in \{0, \ldots, \lfloor n/10d^2 \rfloor\}$, there exists an input $x^i \in \{-1,1\}^n$ such that

- $x^i_j = -1$ for all indices j to the right of the ith block (thus, $x^0 = (-1)^n$).
- The values of x^i_j for indices j to the left of the ith block are set as dictated by the previous step. That is, $x^i_j = x^{i-1}_j$ for all indices j to the left of the ith block.
- $\left|p(x^i)\right| \geq 2^i a + \left(3 \cdot 2^i - 3\right) b$.
- The value of $p(x^i)$ is negative if i is odd, and positive if i is even.

Clearly, proving this hypothesis proves Lemma 4.5. We now prove the induction hypothesis.

- **Base case:** Say $i = 0$. By assumption, $p(-1^n) \geq a$.

- **Inductive step:** Say the hypothesis is true for all $0 \leq j \leq i-1$ for some $i \geq 1$. In the ith block, set the variables corresponding to the even indices to -1 if i is odd, and set the odd indexed variables to -1 if i is even. Set the variables outside the ith block as dictated by the previous step. Assume that i is odd (the argument for even i follows in a similar fashion, with suitable sign changes). Denote the free variables by $y_1, \ldots, y_{5d^2}$. Define a polynomial $p_i : \{-1,1\}^{5d^2} \to \mathbb{R}$ by $p_i(y) = \mathbb{E}_{\sigma \in S_{5d^2}} \tilde{p}(\sigma(y))$, where $\tilde{p}(y)$ denotes the value of p on input $y_1, \ldots, y_{5d^2}$, and the remaining variables are set as described earlier. The expectation is over the uniform distribution. Note that p_i is a symmetric polynomial of degree at most d, and satisfies

$$p_i(-1^{5d^2}) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 3\right) b, \qquad p_i(y) \leq -b \ \forall y \neq -1^{5d^2}.$$

By Lemma 2.18, there exists a univariate polynomial p_i' such that for all $j \in \{0\} \cup [5d^2]$,

$$p_i'(j) = p_i(y) \ \forall y \text{ such that } \#1(y) = j.$$

Thus,

$$p_i'(0) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 3\right) b, \qquad p_1'(j) \leq -b \ \forall j \in [5d^2].$$

Define $p_i'' = p_i' + b$. Thus,

$$p_i''(0) \geq 2^{i-1}a + \left(3 \cdot 2^{i-1} - 2\right) b \qquad p_i''(j) \leq 0 \ \forall j \in [5d^2].$$

If $2^{i-1}a + \left(3 \cdot 2^{i-1} - 2\right) b < 0$, then by Claim 4.6, the inductive hypothesis is true for all integers $j \geq i$. Thus, assume $2^{i-1}a + \left(3 \cdot 2^{i-1} - 2\right) b \geq 0$.

By Lemma 4.4, there exists a $j \in [5d^2]$ such that $p_i''(j) \leq -2^i a - \left(3 \cdot 2^i - 4\right) b$, and hence $p_i'(j) \leq -2^i a - \left(3 \cdot 2^i - 3\right) b$. This implies the existence of an x^i in $\{-1,1\}^n$ (with all variables to the right of the ith block still set to -1, and variables to the left of the ith block as dictated by the previous step) such that $p(x^i) < -2^i a - \left(3 \cdot 2^i - 3\right) b$.

□

5 A separation of depth-2 threshold circuit classes

We now see how to use Theorem 3.1 and the tools from Section 4 to prove sign rank lower bounds against $f \circ \mathsf{XOR}$ when $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3} + \log l}$.

Below is our main technical result of this section, which says that no dual polynomial exists with a large optimum value for (LP2) when $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} : \{-1,1\}^{l^{4/3} + l \log l} \to \{-1,1\}$, even when the smoothness parameter δ is as high as $1/4$.

The following theorem captures the essence of Figure 2.

Theorem 5.1. *Let $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} : \{-1,1\}^{l^{4/3}+l\log l} \to \{-1,1\}, \delta = 1/4$ and $X = \{x \in \{-1,1\}^{l^{4/3}+l\log l} : \bigvee(x) = -1^l\}$. Then for sufficiently large values of l, the optimal value,* OPT*, of (LP2) is less than $2^{-\frac{l^{1/3}}{81}}$.*

Proof. Let p be a polynomial of weight 1, for which (LP2) attains its optimum. Denote the values taken by the variables at the optimum by $\Delta_{\mathsf{OPT}}, \{\xi_{x,\mathsf{OPT}} : x \in X\}$. Towards a contradiction, assume $\mathsf{OPT} \geq 2^{-\frac{l^{1/3}}{81}}$.

Lemma 4.3 (set $m = l^{1/3} + \log l$) shows the existence of an OR polynomial r, on $l/8$ variables, of degree $l^{1/3}$ and weight 1, such that

$$\text{For all } y \in \{-1,1\}^{l/8}, \qquad r(y)\mathsf{OMB}_l^0(y) \geq \Delta_{\mathsf{OPT}} - 2 \cdot 2^{-l^{1/3}} - 2 \cdot 2^{-l^{1/3}}$$
$$\text{and} \qquad r(-1^{l/8}) \geq \Delta + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 2 \cdot 2^{-l^{1/3}} - 2 \cdot 2^{-l^{1/3}}.$$

Note that

$$\mathsf{OPT} \geq 2^{-\frac{l^{1/3}}{81}} \implies \Delta_{\mathsf{OPT}} \geq 2^{-\frac{l^{1/3}}{81}} - \delta \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{2^n} \tag{8}$$

r satisfies the assumptions of Lemma 4.5 with $d = \deg(r) = l^{1/3} < \sqrt{l/32}$ (since any OR polynomial of degree d can be represented by a polynomial of degree at most d), $a = \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-l^{1/3}}$, and $b = \Delta_{\mathsf{OPT}} - 4 \cdot 2^{-l^{1/3}}$. a is non-negative because of the following.

$$\begin{aligned} a &= \Delta_{\mathsf{OPT}} + \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-l^{1/3}} \\ &\geq 2^{-\frac{l^{1/3}}{81}} - 4 \cdot 2^{-l^{1/3}} \geq 0. \end{aligned}$$

Let us denote $k = l^{1/3}/80$ for the remaining of this proof. By Lemma 4.5, there exists an $x \in \{-1,1\}^{l/8}$ such that

$$\begin{aligned} |r(x)| &\geq 2^k a + \left(3 \cdot 2^k - 3\right) b \\ &\geq \Delta_{\mathsf{OPT}}(4 \cdot 2^k - 3) + 2^k \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-80k}(4 \cdot 2^k - 3) \\ &\geq \left(4 \cdot 2^k - 3\right) \left(2^{-\frac{l^{1/3}}{81}} - \delta \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{2^n}\right) + 2^k \frac{\sum_{x \in X} \xi_{x,\mathsf{OPT}}}{|X|} - 4 \cdot 2^{-80k}(4 \cdot 2^k - 3) \\ &\qquad \text{Using Equation 8.} \\ &\geq \left(4 \cdot 2^k - 3\right) \left(2^{-80k/81} - 4 \cdot 2^{-80k}\right) \qquad \text{Since } \delta = 1/4. \\ &> 1 \qquad \text{Assuming } k \geq 1. \end{aligned}$$

This yields a contradiction, since r was a polynomial of weight (in the OR basis) at most 1. ☐

We are now ready to prove our sign rank lower bound.

Theorem 5.2. *Let $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} : \{-1,1\}^{l^{4/3}+l\log l} \to \{-1,1\}$. Then, for sufficiently large values of l,*

$$sr(f \circ \mathsf{XOR}) \geq 2^{\frac{l^{1/3}}{81} - 3}.$$

Proof. Let $n = l^{4/3} + l \log l$. Theorem 5.1 tells us that the optimum of (LP2) (and hence (LP1), by duality) is at most $2^{-\frac{l^{1/3}}{81}}$, when $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l}$, $\delta = 1/4$, and $X = \{x \in \{-1,1\}^{l^{4/3}+l\log l} : \bigvee(x) = -1^l\}$. We first estimate the size of X^c. The probability (over the uniform distribution on the inputs) of a particular OR gate firing a 1 is $\frac{1}{2^{l^{1/3}+\log l}}$. By a union bound, the probability of any OR gate firing a 1 is at most $\frac{1}{2^{l^{1/3}}}$, and hence $|X^c| \leq 2^n \cdot \frac{1}{2^{l^{1/3}}}$.

Plugging these values in Theorem 3.1, we obtain

$$sr(f \circ \mathsf{XOR}) \geq \frac{1/4}{2^{-\frac{l^{1/3}}{81}} + 2^{-l^{1/3}-2}} \geq 2^{\frac{l^{1/3}}{81}-3}.$$

□

Corollary 5.3. *Let $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} : \{-1,1\}^{l^{4/3}+l\log l} \to \{-1,1\}$, and let $n = l^{4/3} + l \log l$ denote the number of input variables. Then, for sufficiently large values of n,*

$$\mathsf{UPP}(f \circ \mathsf{XOR}) = \Omega\left(n^{1/4}\right).$$

Proof. It follows from Theorem 5.2 and Theorem 2.16. □

We now prove Theorem 1.1, which gives us a lower bound on the size of $\mathsf{THR} \circ \mathsf{MAJ}$ circuits computing $\mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$.

Proof of Theorem 1.1. Suppose $\mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$ could be represented by a $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size s. Let $n = 2l^{4/3} + 2l \log l$. By Lemma 2.10 and Theorem 5.2,

$$s\left(2l^{4/3} + 2l \log l\right) \geq sr(f) \geq 2^{\frac{l^{1/3}}{81}-3}.$$

Thus, $s = 2^{\Omega\left(n^{1/4}\right)}$.

□

Finally, we prove Corollary 1.2, which separates $\mathsf{THR} \circ \mathsf{MAJ}$ from $\mathsf{THR} \circ \mathsf{THR}$.

Proof of Corollary 1.2. Let $n = 2l^{4/3} + 2l \log l$ denote the number of input bits to $F = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$. By Lemma 2.5, F can be computed by a $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$ circuit of size $2l^{4/3} + 2l \log l$. Hence $F \in \mathsf{THR} \circ \mathsf{ETHR} = \mathsf{THR} \circ \mathsf{THR}$, by Theorem 2.3. By Theorem 1.1, $\mathsf{THR} \circ \mathsf{MAJ}$ circuits computing F require size $2^{\Omega\left(n^{1/4}\right)}$. □

6 Communication complexity class separations

In this section, we show explicit separations between certain communication complexity classes, resolving an open question posed in [19]. This application of our main result was brought to our attention by Göös [17].

6.1 Definitions

We first define a few communication complexity classes of interest. For any communication model $\mathcal{C}$, and function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$, denote $\mathcal{C}(F)$ to be the minimum cost of a correct protocol for F in the model $\mathcal{C}$. We denote by $\mathcal{C}^{cc}$ the class of all functions F with $\mathcal{C}(F)$ at most polylogarithmic in n.

Definition 6.1 (NP)**.** *An* NP *protocol* Π *outputs* -1 *or* 1 *indicating whether the input is in* $\bigcup_{w\in\{-1,1\}^k} R_w$*, where* $\{R_w : w \in \{-1,1\}^k\}$ *are rectangles. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if for all* $(x,y) \in \{-1,1\}^n \times \{-1,1\}^n$, $\Pi(x,y) = F(x,y)$*. The cost of the protocol is* k*.*

In an MA protocol Π, Merlin is an all-powerful prover who has access to Alice's and Bob's inputs. He sends a proof string to both Alice and Bob, who then run a randomized protocol, Π. The protocol is correct for a function $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ if for all inputs x, y to Alice and Bob, the probability of the output of Π agreeing with F is at least $2/3$. The cost of Π is the sum of the maximum cost of any constituent deterministic protocol and the length of Merlin's proof string. Formally,

Definition 6.2 (MA)**.** *An* MA *protocol is a distribution over deterministic protocols* Π*, that take an additional input* $w \in \{-1,1\}^k$ *(Merlin's proof string), visible to both Alice and Bob. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if it satisfies the following properties.*

$$\begin{aligned} &\textit{Completeness:} \quad \textit{If } F(x,y) = -1, \ \textit{then } \exists w : \Pr[\Pi(x,y,w) = -1] \geq 2/3, \\ &\textit{Soundness:} \quad \textit{If } F(x,y) = 1, \ \textit{then } \forall w : \Pr[\Pi(x,y,w) = -1] \leq 1/3. \end{aligned}$$

The cost of the protocol is the sum of the maximum cost of any constituent deterministic protocol, and k*.*

Definition 6.3 (S2P)**.** *An* S2P *protocol can be viewed as a matrix* Π*, with rows indexed by* $r \in \{-1,1\}^k$*, columns indexed by* $c \in \{-1,1\}^k$*, where each entry contains a deterministic protocol. The protocol correctly computes* $F : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ *if the matrix satisfies the following properties.*

$$\begin{aligned} &\textit{If } F(x,y) = -1, \ \textit{then } \exists c \ \forall r : \ \Pi_{r,c}(x,y) = -1, \\ &\textit{if } F(x,y) = 1, \ \textit{then } \exists r \ \forall c : \ \Pi_{r,c}(x,y) = 1. \end{aligned}$$

The cost of the protocol is the sum of the maximum cost of any constituent deterministic protocol, and k*.*

We now define protocols where Alice and Bob have access to certain oracles.

Definition 6.4 (P^{NP})**.** *A* P^{NP} *protocol* Π*, is a protocol in which at each step, one of the following actions occur.*

- *For cost 1, Alice sends a bit to Bob.*

- *For cost 1, Bob sends a bit to Alice.*
- *For cost k, Alice and Bob compute the value of $g(x, y)$, where g has an* NP *protocol of cost k.*

The protocol correctly computes $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$ if $\Pi(x, y) = F(x, y)$ for all $x, y \in \{-1, 1\}^n$.

Definition 6.5 (P^{MA})**.** *A P^{MA} protocol Π, is a protocol in which at each step, one of the following actions occur.*

- *For cost 1, Alice sends a bit to Bob.*
- *For cost 1, Bob sends a bit to Alice.*
- *For cost k, Alice and Bob compute the value of $g(x, y)$, where g has an* MA *protocol of cost k.*

The protocol correctly computes $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$ if $\Pi(x, y) = F(x, y)$ for all $x, y \in \{-1, 1\}^n$.

Definition 6.6 ($\mathsf{BPP}^{\mathsf{NP}}$)**.** *A $\mathsf{BPP}^{\mathsf{NP}}$ protocol is a distribution over P^{NP} protocols Π. The protocol correctly computes $F : \{-1, 1\}^n \times \{-1, 1\}^n \to \{-1, 1\}$ if $\Pr[\Pi(x, y) = F(x, y)] \geq 2/3$ for all $x, y \in \{-1, 1\}^n$.*

6.2 Class separations

The function we use for the class separations is $F = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2$. Note that $F = \mathsf{OMB}_l \circ \mathsf{EQ}_{l^{1/3}+\log l}$, where OMB_l outputs -1 iff the rightmost bit of the input set to -1 occurs at an odd index.

It is not hard to see that there is an MA protocol for $\bigvee_l \circ \mathsf{EQ}_{l^{1/3}+\log l}$ of cost polylogarithmic in l. Using this, and a binary search, we exhibit a P^{MA} upper bound for F in Protocol 1.

Protocol 1 P^{MA} protocol for $\mathsf{OMB}(\mathsf{EQ}_1, \ldots, \mathsf{EQ}_{l^{1/3}+\log l})$

---

if $\bigvee_{i=1}^{l^{1/3}+\log l} (\mathsf{EQ}_i) = 1$ **then** Output 1.
end if
$start = 1$
$end = l^{1/3} + \log l$
$mid = \lceil \frac{start+end}{2} \rceil$
while $start \neq end$ **do**
  if $\bigvee_{i=mid}^{end} (\mathsf{EQ}_i) = -1$ **then** $start \leftarrow mid$
  else if $\bigvee_{i=mid}^{end} (\mathsf{EQ}_i) = 1$ **then** $end \leftarrow mid - 1$
  end if
end while
Output -1 iff $start$ is odd.

Hence, we obtain

$$\mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} \circ \mathsf{XOR}_2 \in \mathsf{P}^{\mathsf{MA}cc}.$$

Along with Corollary 5.3, this yields the following result.

Theorem 6.7.

$$\mathsf{P}^{\mathsf{MA}cc} \nsubseteq \mathsf{UPP}^{cc}.$$

It is known that $\mathsf{P}^{\mathsf{MA}cc} \subseteq \mathsf{S2P}^{cc}$, and $\mathsf{P}^{\mathsf{MA}cc} \subseteq \mathsf{BPP}^{\mathsf{NP}cc}$ (see e.g. [19] for references for such containments, and an excellent overview on the landscape of communication complexity classes).

Thus, Theorem 6.7 yields

$$\mathsf{S2P}^{cc} \nsubseteq \mathsf{UPP}^{cc},$$

resolving an open question posed in [19]. We also obtain

$$\mathsf{BPP}^{\mathsf{NP}cc} \nsubseteq \mathsf{UPP}^{cc}.$$

To the best of our knowledge, ours is the first explicit total function to witness the above separation. We remark here that Bouland et al. [5] used a partial function to witness the same separation.

7 Conclusions

This work refines our understanding of depth-2 threshold circuits by providing the following summary:

$$\widehat{LT}_1 \subsetneq LT_1 \subsetneq \widehat{LT}_2 = \mathsf{MAJ} \circ \mathsf{THR} \subsetneq \underbrace{\mathsf{THR} \circ \mathsf{MAJ} \subsetneq LT_2}_{\text{This work}} \subseteq \widehat{LT}_3 \subseteq \mathrm{NP/poly}.$$

While we cannot rule out that SAT has efficient $\mathsf{THR} \circ \mathsf{THR}$ circuits, we do not even know whether IP is in LT_2. On the other hand, the most powerful method used to prove lower bounds on the size of depth-2 threshold circuits for computing an explicit function f exploits the fact that f has large sign rank. Before our work, it was not known if LT_2 contained any function of large sign rank. Our main result shows that indeed there are such functions, answering a question explicitly raised by Hansen and Podolskii [21] and Amano and Maruoka [1].

The central open question in the area is to prove super-polynomial lower bounds on the size of $\mathsf{THR} \circ \mathsf{THR}$ circuits. The best known explicit lower bounds due to Kane and Williams [27] is roughly $n^{3/2}$. We feel that there is a dire need of discovering new techniques for proving strong lower bounds against $\mathsf{THR} \circ \mathsf{THR}$ circuits.

On a second front, our main result shows that the communication complexity class P^{MA} has functions with large sign rank, i.e. is not contained in UPP, strongly resolving an open problem by Göös et al. [19]. This is in contrast to the known containment of $\mathsf{P}^{\mathsf{NP}} \subsetneq \mathsf{UPP}$. As the sign rank lower bound technique against UPP remains the strongest known technique for proving lower bounds against communication protocols (including quantum protocols), it suggests that new techniques need to be developed for proving bounds against P^{MA}. Indeed, there are specialized techniques for proving lower bounds against the class P^{NP} (see [26, 18]). How can they be generalized to P^{MA}?

Acknowledgements

We are grateful to Kristoffer Hansen for bringing to our attention the question of separating the classes $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$ at the Summer School on Lower Bounds, held in Prague in 2015. We also thank Michal Koucký for organizing and inviting us to the workshop. We thank Mika Göös for pointing out to us the consequences of our results in communication complexity class separations (Section 6), and Justin Thaler and Ryan Williams for invaluable feedback on an earlier version of this manuscript.

References

[1] Kazuyuki Amano and Akira Maruoka. Complexity of depth-2 circuits with threshold gates. In *, 30th International Symposium Mathematical Foundations of Computer Science MFCS*, pages 107–118, 2005.

[2] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science, Toronto, Canada, 27-29 October 1986*, pages 337–347, 1986.

[3] Richard Beigel. Perceptrons, PP, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[4] Richard Beigel and Jun Tarui. On ACC. *Computational Complexity*, 4:340–366, 1994.

[5] Adam Bouland, Lijie Chen, Dhiraj Holden, Justin Thaler, and Prashant Nalini Vasudevan. On SZK and PP. *CoRR*, abs/1609.02888, 2016. To appear in FOCS, 2017.

[6] Jehoshua Bruck. Harmonic analysis of polynomial threshold functions. *SIAM J. Discrete Math.*, 3(2):168–177, 1990.

[7] Harry Buhrman, Nikolay Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07, pages 24–32. IEEE Computer Society, 2007.

[8] Mark Bun and Justin Thaler. Improved bounds on the sign-rank of AC^0. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 37:1–37:14, 2016.

[9] A. K. Chandra, L. Stockmeyer, and U. Vishkin. Constant depth reducibility. *SIAM J. Computing*, 13:423–439, 1984.

[10] Arkadev Chattopadhyay and Nikhil S. Mande. Dual polynomials and communication complexity of XOR functions. *Arxiv*, 2017.

[11] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. Average-case lower bounds and satisfiability algorithms for small threshold circuits. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 1:1–1:35, 2016.

[12] Hartmut Ehlich and Karl Zeller. Schwankung von polynomen zwischen gitterpunkten. *Mathematische Zeitschrift*, 86(1):41–44, 1964.

[13] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001*, pages 100–106, 2001.

[14] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *FST TCS 2001: Foundations of Software Technology and Theoretical Computer Science, 21st Conference, Bangalore, India, December 13-15, 2001, Proceedings*, pages 171–182, 2001.

[15] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates VS. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[16] Mikael Goldmann and Marek Karpinski. Simulating threshold circuits by majority circuits. *SIAM J. Comput.*, 27(1):230–246, 1998.

[17] Mika Göös. Private Communication, 2017.

[18] Mika Göös, Pritish Kamath, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for PˆNP. In *32nd Computational Complexity Conference, CCC 2017, July 6-9, 2017, Riga, Latvia*, pages 12:1–12:16, 2017.

[19] Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 86:1–86:15, 2016.

[20] A. Hajnal, W. Maas, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[21] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Exact threshold circuits. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, June 9-12, 2010*, pages 270–279, 2010.

[22] Kristoffer Arnsfelt Hansen and Vladimir V. Podolskii. Polynomial threshold functions and boolean threshold circuits. *Inf. Comput.*, 240:56–73, 2015.

[23] Johan Håstad. On the size of weights for threshold gates. *SIAM J. Discrete Math*, 7(3):484–492, 1994.

[24] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

[25] Thomas Hofmeister. A note on the simulation of exponential threshold weights. In *Computing and Combinatorics, Second Annual International Conference, COCOON '96, Hong Kong, June 17-19, 1996, Proceedings*, pages 136–141, 1996.

[26] Russell Impagliazzo and Ryan Williams. Communication complexity with synchronized clocks. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, June 9-12, 2010*, pages 259–269, 2010.

[27] Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 633–643, 2016.

[28] Matthias Krause and Pavel Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theor. Comput. Sci.*, 174(1-2):137–156, 1997.

[29] Matthias Krause and Pavel Pudlák. Computing boolean functions by polynomials and threshold circuits. *Computational Complexity*, 7(4):346–370, 1998.

[30] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[31] Marvin Minsky and Seymour Papert. *Perceptrons - an introduction to computational geometry*. MIT Press, 1987.

[32] S. Muroga. *Threshold Logic and its Applications*. Wiley-Interscience, 1971.

[33] Ramamohan Paturi and Janos Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[34] N. Pippenger. The complexity of computations by networks. *IBM J.Res.Develop.*, 31:235–243, 1987.

[35] Alexander A. Razborov. On small depth threshold circuits. In *Third Scandinavian Workshop on Algorithm Theory (SWAT)*, pages 42–52, 1992.

[36] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[37] Theodore J Rivlin and Elliott W Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on numerical Analysis*, 3(2):311–320, 1966.

[38] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[39] Alexander A. Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011.

[40] K. I. Siu and J. Bruck. On the power of threshold circuits with small weights. *SIAM J. Discrete Math.*, 4(3):423–435, 1991.

[41] Andrew Chi-Chih Yao. On ACC and threshold circuits. In *Proceedings of the 31st IEEE FOCS*, pages 619–627, 1990.

A Signed monomial complexity lower bounds

In this section, we show how upper bounding the optimum of LP1 (and LP2) w.r.t a function f yields signed monomial complexity lower bounds for representing it. This is already implied by Theorem 3.1, as a sign rank lower bound on $f \circ \mathsf{XOR}$ directly implies a signed monomial complexity lower bound on f. The use of Theorem 3.1, whose proof makes use of the deep result of Forster [13], seems an overkill to just lower bound signed monomial complexity. In this section, we give a much more direct proof of this fact, entirely avoiding the use of Forster's theorem. This also allows us to generalize a classical result of Bruck [6] that gave a sufficient condition for lower bounding signed monomial complexity. One may note that our generalization is analogous to Razborov and Sherstov's [36] generalization of Forster's Theorem. Further, our generalized result, Theorem A.2, along with Theorem 5.1, will directly imply that there are functions in poly-size $\mathsf{THR} \circ \mathsf{OR}$ circuits that cannot be computed in sub-exponential size by $\mathsf{THR} \circ \mathsf{XOR}$ circuits. Such a result was first proved by Krause and Pudlák [29], using a different technique. Interestingly, Krause and Pudlák expressed the belief that such a separation cannot be done based on a spectral technique like that of Bruck's Theorem [6]. Our argument here shows that this belief was false.

We recall Bruck's Theorem below.

Theorem A.1 ([6]). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function. If $\max_{S\subseteq[n]} \left|\hat{f}(S)\right| \leq \varepsilon$, then*

$$\mathrm{mon}_{\pm}(f) \geq \frac{1}{\varepsilon}.$$

The following is our generalization of Theorem A.1.

Theorem A.2. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function, and X any subset of $\{-1,1\}^n$. Suppose there exists a distribution μ on $\{-1,1\}^n$ such that $\max_{S\subseteq[n]} \left|\widehat{f\mu}(S)\right| \leq \varepsilon$ and $\min_{x\in X} \mu(x) \geq \delta$. Then,*

$$\mathrm{mon}_{\pm}(f) \geq \frac{\delta}{\varepsilon + \delta \cdot \frac{|X^c|}{2^n}}.$$

Proof. Let $p : \{-1,1\}^n \to \mathbb{R}$ be any polynomial which sign represents f. By Fact 2.12,

$$\begin{aligned}\mathop{\mathbb{E}}_x[f(x)\mu(x)p(x)] = \sum_{S\subseteq[n]} \widehat{f\mu}(S)\widehat{p}(S) &\leq \max_{S\subseteq[n]}\left|\widehat{f\mu}(S)\right| \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathrm{mon}(p) &(9)\\ &\leq \varepsilon \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathrm{mon}(p). &(10)\end{aligned}$$

Note that

$$\begin{aligned}\mathop{\mathbb{E}}_x[f(x)\mu(x)p(x)] &= \frac{1}{2^n}\sum_{x\in X} f(x)\mu(x)p(x) + \frac{1}{2^n}\sum_{x\in X^c} f(x)\mu(x)p(x)\\ &\geq \frac{\min_{x\in X}\mu(x)}{2^n}\left[\sum_{x\in\{-1,1\}^n}|p(x)| - |X^c| \cdot \max_{x\in X^c}|p(x)|\right]\\ &\qquad\qquad\qquad\qquad\text{Since } p \text{ sign represents } f\\ &\geq \delta \cdot \max_{S\subseteq[n]}|\widehat{p}(S)| - \frac{\delta}{2^n}\cdot|X^c|\cdot\max_{x\in\{-1,1\}^n}|p(x)|. \qquad \text{Using Lemma 2.11}\end{aligned}$$

Combining the above and Equation 9, we obtain

$$\begin{aligned}
&\varepsilon \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| \cdot \mathrm{mon}(p) \geq \delta \cdot \max_{S\subseteq[n]} |\widehat{p}(S)| - \frac{\delta}{2^n} \cdot |X^c| \cdot \max_{x\in\{-1,1\}^n} |p(x)| \\
\Longrightarrow\ &\varepsilon \cdot \mathrm{mon}(p) \geq \delta - \frac{\delta}{2^n} \cdot |X^c| \cdot \frac{\max_{x\in\{-1,1\}^n} |p(x)|}{\max_{S\subseteq[n]} |\widehat{p}(S)|} \geq \delta - \frac{\delta}{2^n} \cdot |X^c| \cdot \mathrm{mon}(p) \\
\Longrightarrow\ &\mathrm{mon}(p) \geq \frac{\delta}{\varepsilon + \delta \cdot \frac{|X^c|}{2^n}}.
\end{aligned}$$

□

The following theorem provides a signed monomial complexity lower bound against a function in $\mathsf{THR} \circ \mathsf{OR}$.

Theorem A.3. *Let* $f = \mathsf{OMB}_l^0 \circ \bigvee_{l^{1/3}+\log l} : \{-1,1\}^{l^{4/3}+l\log l} \to \{-1,1\}$. *Then,*

$$\mathrm{mon}_{\pm}(f) \geq 2^{\frac{l^{1/3}}{81}-3}.$$

Proof. The proof follows from Theorem A.2 and Theorem 5.1 in the same way as the proof of Theorem 5.2 follows from Theorem 3.1 and Theorem 5.1. □

This gives us a function f on n input variables, computable by linear sized $\mathsf{THR} \circ \mathsf{AND}$ circuits, such that for large enough n,

$$\mathrm{mon}_{\pm}(f) \geq 2^{\Omega\left(n^{1/4}\right)}.$$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il