Electronic Colloquium on Computational Complexity, 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 [31] even for computing functions in depth-3 AC^0. Yet, no separation among the two depth-2 threshold circuit classes were known. In fact, it is not clear a priori that they ought to be different. In particular, Goldmann, Håstad and Razborov [14] showed that the class $\mathsf{MAJ} \circ \mathsf{MAJ}$ is identical to the class $\mathsf{MAJ} \circ \mathsf{THR}$.

In this work, we provide an exponential separation between $\mathsf{THR} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$. We achieve this by showing a function f that is computed by linear size $\mathsf{THR}\circ\mathsf{THR}$ circuits and yet has exponentially large *sign rank*. This, by a well-known result, implies that f requires exponentially large $\mathsf{THR} \circ \mathsf{MAJ}$ circuits to be computed. Our result suggests that the sign rank method alone is unlikely to prove strong lower bounds against $\mathsf{THR} \circ \mathsf{THR}$ circuits.

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

†Supported by a DAE 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, 16, 20, 14, 30, 5, 23, 24, 12, 13, 31, 17, 18, 22, 10], 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 $x \in \{-1, 1\}^n$,

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

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 [26] showed that the 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 [27] 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 lower bound was shown only in the nineties by Håstad [19]. 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, 14, 34, 17, 18, 30, 16, 21, 15]. 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 of 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 [8, 29, 34], Goldmann, Håstad and Razborov [14] 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 [35] 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. [16] 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. [14] 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 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 [12] 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 [17] 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^0_ℓ: 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:

$$\mathrm{OMB}^0_\ell(x) = -1 \iff \sum_{i=1}^{\ell} (-1)^{i+1} 2^i \left(1 + x_i\right) \geq 0.5$$

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. Then, we show the following:

Theorem 1.1. *Let F_n be defined on $n = 2\ell^{4/3}$ bits as $\mathsf{OMB}^0_\ell \circ \mathsf{OR}_{\ell^{1/3} - \log l} \circ \mathsf{XOR}_2$. 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 computed by an ETHR gate. Hence, F_n is in $\mathsf{THR} \circ \mathsf{ETHR}$, a class that Hansen and Podolskii [17] 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 $\mathsf{THR} \circ \mathsf{THR}$.*

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, [12, 31, 7]) 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 $x \in \{-1, 1\}^{|X|}, y \in \{-1, 1\}^{|Y|}$. Any real matrix sign represents M_f if each if 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 [12] deep result (see Theorem 2.7) 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 [31] 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 on devising an f that is in $\mathsf{THR} \circ \mathsf{THR}$ that plausibly has high sign rank. On this, we were guided by another interpretation of sign rank, due to Paturi and Simon [28]. Paturi and Simon introduced a model of 2-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, by far, the strongest 2-party known model against which we know how to prove lower bounds. [28] 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: assume that the sum of the magnitude of the weights of the top THR gate is 1. 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 as small as the smallest weight of the top gate. 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 our function 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 [31] 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 [33]. 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 the Razborov-Sherstov work overcomes is the analysis of the dual of this LP using and building appropriate tools of approximation theory. We take cue from this work and follow its general framework of analysing 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 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 $1/n^2$.

The specific new technical challenge that one faces is the following: instead of low degree orthogonality, one now needs a distribution μ w.r.t which the outer function has low correlation with *all parities* (see LP1). Just dealing with high degree parity constraints, though non-trivial, was done in the recent work of the authors [9]. However, unlike there, here one needs the additional constraint of the distribution being (approximately) smooth enough. Analysing this combination of high degree parity constraints and the smoothness constraints, is the main new technical challenge that our work addresses. We do this by a novel combination of ideas that differs entirely from the Razborov-Sherstov analysis.

Analyzing the dual of our LP (LP2) involves arguing against the existence of a certain kind of (possibly high degree) polynomial representation. We require several ideas to deal with it. First, the dual polynomial P has unit weight. While it does not necessarily sign represent $f = \mathsf{THR}_\ell \circ \mathsf{OR}_m$, it is constrained to not stray too far away from zero on the wrong side on each point of its domain $\{-1, 1\}^n$. Moreover, over a set X, where we want the distribution to be smooth in in the primal LP, roughly speaking, P's margin in representing f on average has to be good. Since the set X has to be large (to get good approximate smoothness), we are essentially forced to include in X all inputs that are mapped to 1^l by the bottom ORs of f. In particular, we set X to be precisely the set of such points. With this setting, our bound on the sign rank becomes roughly δ/OPT, where OPT is the optimal value of the LP.

The first idea we use is an averaging argument that appeared in the work of Krause and Pudlák [24]. What this does is that for each possible input $y \in \{-1, 1\}^\ell$ to THR_ℓ, it takes the average of all values of P under the uniform distribution over all points x such that $\mathsf{OR}_m(x) = y$. This achieves the following as described in Lemma 3.1: the polynomial P over x transforms to *an* OR *polynomial* Q, over y's, of the same weight as P, plus an error term whose magnitude is exponentially small in the fan-in of the bottom OR gates of f. Here, an OR polynomial is a linear combination of ORs of subsets of variables. Assuming, for the sake of contradiction, OPT to be large enough, we can safely ignore the error term. This gives us a passage to an OR polynomial of unit weight representing our top THR function g, with the same worst-case guarantee that held for P. Additionally, we get the guarantee that at $y = -1^l$, Q's margin is better by the average margin of P on the set X. The intuition is that when OPT is large, this average margin is also large compared to Δ, the worst case margin.

Now we want to argue that such a Q does not exist if we select our top threshold g judiciously. We select the ODD-MAX-BIT function, denoted by OMB, for this purpose. We then observe that if we randomly restrict each variable to -1, then the expected weight of OR monomials of degree at least d that do not get fixed is as small as $1/2^d$. Ignoring this high degree monomials, therefore does not decrease our margin by too much. Further, with high probability, the restriction induces an OMB of sufficiently large number of free variables. This now gives us a polynomial of Q' of degree less than d that has worst case margin not too small, but does somewhat better on -1^l. While margin bounds against sign representing polynomials of sufficiently small degree have been obtained several times before, our setting is different. Q' is not sign-representing OMB. It is here that our choice of the ODD-MAX-BIT function comes in very handy. We show that a standard approximation theoretic lemma of Ehlich and Zeller [11], Rivlin and Cheney [32] can be used to argue against the existence of such a Q' for OMB.

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 [17] showed the following.

Theorem 2.3 (Hansen and Podolskii [17])**.** *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 f be an exact threshold function with the representation $\sum_{i=1}^n w_i x_i = t$. There exists an $\varepsilon_f > 0$ such that $\sum_{i=1}^n w_i x_i > t \implies \sum_{i=1}^n w_i x_i > t + \varepsilon_f$. Consider a $\mathsf{THR} \circ \mathsf{ETHR}$ circuit of size s, say it computes $\mathrm{sgn}(c_0 + \sum_{i=1}^s f_i)$, where f_is 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 $\mathrm{sgn}\left(\sum_{i=1}^{s} c_i(g_{i,2} - g_{i,1} + 1)\right)$, where g_is 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 [17] 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 to be $\sum_{S \subseteq [n]} |a_S|$, and its degree to be $\max_{S \subseteq [n]}\{|S| : a_S \neq 0\}$.*

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

Theorem 2.7 (Forster [12])**.** *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 [31].

Theorem 2.8 (Razborov and Sherstov [31])**.** *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 from Forster et al. [13] tells us that functions that have efficient $\mathsf{THR} \circ \mathsf{MAJ}$ representations have low sign rank.

Lemma 2.9 (Forster et al. [13])**.** *Let $f : \{-1,1\}^n \times \{-1,1\}^n \to \{-1,1\}$ be a boolean function computed by a $\mathsf{THR} \circ \mathsf{MAJ}$ 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.

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 they 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 on the worst case input and randomness. 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 the notation used in [25], for example. Unbounded error communication complexity was introduced by Paturi and Simon [28], and is defined as follows.

$$\mathsf{UPP}(f) = \min_{\varepsilon}(R_\varepsilon(f)).$$

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

Definition 2.10.

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

Paturi and Simon [28] showed an equivalence between $\mathsf{UPP}(f)$ and the sign rank of M_f, where M_f denotes the communication matrix of f.

Theorem 2.11 (Paturi and Simon [28])**.** *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 2.12 (Folklore)**.** *Let $f : \{-1,1\}^n \times \{-1,1\}^n \to \mathbb{R}$ 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 [26].

Lemma 2.13 (Minsky and Papert [26])**.** *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 Hardness of approximating $\mathsf{OMB}_l^0 \circ \mathsf{OR}_m$

For notational convenience, denote $g = \mathsf{OMB}_l^0, f = g \circ \bigvee_m$. Let $n = lm$. We first use an idea from Krause and Pudlák [24] which enables us work with g, rather than $g \circ \bigvee_m$.

Lemma 3.1. *Let $f = g_l \circ \bigvee_m : \{-1,1\}^{ml} \to \{-1,1\}, \Delta \in \mathbb{R}, e_x \geq 0\ \forall x \in X$, 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*

$$\begin{aligned} \forall x \in \{-1,1\}^{ml}, \qquad & f(x)p(x) \geq \Delta, \\ \forall x \in \{-1,1\}^{ml} \text{ such that } \bigvee_m(x) = -1^l, \qquad & f(x)p(x) \geq \Delta + e_x. \end{aligned}$$

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

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

Proof. For any $y \in \{-1,1\}^l$, denote by $\mathbb{E}_y[f(x)]$ the expected value of $f(x)$ with respect to the uniform distribution over all $x \in \{-1,1\}^{ml}$ such that $\bigvee_m(x) = y$. For each $I_k \subseteq [l]\times[m]$, define $J_k \subseteq [l]$ to be the projection of I_k on $[l]$. Formally,

$$i \in J_k \iff \exists j,\ x_{i,j} \in I_k.$$

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

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

and

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

Next, we approximate each monomial by an OR function. The following argument appears in the proof of Lemma 2.3 in [24]. However, we reproduce the proof below for clarity and completeness.

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

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

Let $A = \{j \in [l] : y_j = -1\}$. If $A \cap J_k = \emptyset$, then

$$\mathop{\mathbb{E}}_{y}\left[\bigoplus_{(i,j)\in I_k} x_{i,j}\right] = \bigvee_{j\in J_k} y_j = 1$$

Else, let $B = A \cap J_k$. In this case, $\bigvee_{j\in J_k} y_j = -1$. Also,

$$\mathop{\mathbb{E}}_{y}\left[\bigoplus_{(i,j)\in I_k} x_{i,j}\right] = \mathop{\mathbb{E}}_{x\in\{-1,1\}^{A\cap J_k} : \bigvee(x) = -1^{|A\cap J_k|}}\left[\bigoplus_{(i,j)\in I_k, y_i=-1} x_{i,j}\right] \tag{1}$$

Note that

$$\mathop{\mathbb{E}}_{x\in\{-1,1\}^{A\cap J_k}}\left[\bigoplus_{(i,j)\in I_k, y_i=-1} x_{i,j}\right] = 0 \tag{2}$$

Denote $|A \cap J_k| = q$. Using Equation 2 and a simple counting argument, the absolute value of the RHS (and thus the LHS) of Equation 1 can be upper bounded as follows.

$$\begin{aligned}\left|\mathbb{E}_y\left[\bigoplus_{(i,j)\in I_k} x_{i,j}\right]\right| &\le \frac{2^{mq} - (2^m-1)^q}{(2^m-1)^q} \\ &\le \frac{q \cdot 2^{mq-m}}{2^{mq}} \le l2^{-m}\end{aligned}$$

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

$$\left|\mathop{\mathbb{E}}_{y}\left[\bigoplus_{(i,j)\in I_k} x_{i,j}\right] - \frac{1}{2} - \frac{1}{2}\bigvee_{j\in J_k} y_j\right| \le 2l2^{-m}. \tag{3}$$

Say $p = v_0 + \sum_k v_k p_k$, where $p_k(x) = \oplus_{(i,j)\in I_k} x_{i,j}$ is the unique multilinear expansion of p. Define

$$p' = v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2} \bigvee_{j\in J_k} y_j.$$

Note that

$$wt(p') = wt\left(v_0 - \frac{\sum_k v_k}{2} - \sum_k \frac{v_k}{2}\bigvee_{j\in J_k} y_j\right) = \left|v_0 - \frac{\sum_k v_k}{2}\right| + \sum_k \left|\frac{v_k}{2}\right| \le wt(p).$$

Thus, using linearity of expectation, we obtain that for all $y \in \{-1,1\}^l$,

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

and

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

□

Next, we use random restrictions which reduces the degree of the approximating OR polynomial, at the cost of a small error.

Lemma 3.2. *Let* $g_l = \mathsf{OMB}^0 : \{-1,1\}^l \to \{-1,1\}, f = g_l \circ \bigvee_m$*, and* $\Delta, \{e_x \geq 0 : x \in X\}$ *(where* X *is defined as in Lemma 3.1), and* p *be a real polynomial such that*

$$\begin{cases} \forall x \in \{-1,1\}^{ml}, & f(x)p(x) \geq \Delta \\ \forall x \in \{-1,1\}^{ml} \text{ such that } \bigvee_m(x) = -1^l, & p(x) \geq \Delta + e_x. \end{cases}$$

Then, for any integer $d > 0$*, there exists an* OR *polynomial* p''*, of degree* d *and weight at most* $wt(p)$*, such that*

$$\text{For all } y \in \{-1,1\}^{l/8}, \qquad p''(y)g_{l/8}(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m} + 2^{-(d-1)}\right)$$

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

Proof. Lemma 3.1 guarantees the existence of an OR polynomial p', of weight at most $wt(p)$, such that

$$\forall y \in \{-1,1\}^l, \qquad p'(y)g(y) \geq \Delta - wt(p)\left(2l \cdot 2^{-m}\right)$$

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

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 p' 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 p' 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 p' to obtain p'', 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,

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

□

3.1 Hardness of OMB^0

In this section, we show that approximating OMB^0 by a low weight polynomial p must imply that the degree of p is large.

We require the following result by Ehlich and Zeller [11] and Rivlin and Cheney [32].

Lemma 3.3 ([11, 32])**.** *The following holds true for any real valued $\alpha > 0$ and $k > 0$. Let p 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 next use the idea of 'doubling' for the OMB^0 function, as in [3, 6] to show that a low degree polynomial of bounded weight cannot represent OMB^0 well. This is our main approximation theoretic lemma.

Lemma 3.4. *Suppose p is a polynomial of degree $d < \sqrt{n/4}$ and $a > 0, b \in \mathbb{R}$ be reals such that $\mathsf{OMB}^0(-1^n) \geq a$ and $\mathsf{OMB}^0(x)p(x) \geq b$ for all $x \in \{-1,1\}^n$. Then, for all $i \in \{0, 1, \ldots, \lfloor n/10d^2 \rfloor\}$, there exists an $x_i \in \{-1,1\}^n$ (not necessarily distinct) such that $|p(x_i)| \geq 2^i a + \left(3 \cdot 2^i - 3\right) b$.*

The argument will be an iterative one, inspired by the arguments of Beigel and Buhrman et al. [3, 6].

Claim 3.5. *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 $i \geq 0$, then $2^j a + \left(3 \cdot 2^j - 3\right) b < 0$ for all $j > i$.*

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 $2^j a + \left(3 \cdot 2^j - 3\right) b = 2^{j-i} \left(2^i a + \left(3 \cdot 2^i - 2\right) b\right) + 3 \cdot (2^{j-i+1} - 3) b < 0$. □

Proof of Lemma 3.4. We will assume, for the rest of the proof, that

$$2^i a + \left(3 \cdot 2^i - 2\right) b \geq 0 \;\forall i \in \left[\lfloor n/10d^2 \rfloor\right]. \tag{4}$$

If not, the lemma is trivially true by Claim 3.5.

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

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

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

We now prove the induction hypothesis.

- **Base case:** Say $i = 1$. We know from our assumption that $\mathsf{OMB}^0(-1^n) \geq a$ and $\mathsf{OMB}^0(x)p(x) \geq b$ for all $x \in \{-1, 1\}^n$. Set the variables corresponding to the even indices in the first block to -1, and all variables to the right of the first block to -1. Denote the free variables by $y_1, \ldots, y_{5d^2}$. Define a polynomial $p_1 : \{-1, 1\}^{5d^2} \to \mathbb{R}$ by $p_1(y) = \mathbb{E}_{\sigma \in S_{5d^2}} \tilde{p}(\sigma(y))$, 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_1 is a symmetric polynomial of degree at most d, and satisfies
$$p_1(-1^{5d^2}) \geq a, \qquad p_1(y) \leq -b \ \forall y \neq -1^{5d^2}.$$
By Lemma 2.13, there exists a univariate polynomial p_1' such that for all $i \in \{0\} \cup [5d^2]$,
$$p_1'(i) = p_1(y) \ \forall y \text{ such that } \#1(y) = i$$
Thus,
$$p_1'(0) \geq a, \qquad p_1'(j) \leq -b \ \forall j \in [5d^2].$$
Define $p_1'' = p_1' + b$. Thus, $p_1''(0) \geq a + b \geq 0$, and $p_1''(j) \leq 0 \ \forall j \in [5d^2]$.

  By Lemma 3.3, there exists a $j \in [5d^2]$ such that $p_1''(j) < -2a - 2b$. This means $p_1'(j) < -2a - 3b < 0$, because of Equation 4. This implies existence of an $x \in \{-1, 1\}^n$ (with all variables to the right of the first block still set to -1) such that $p(x) < -2a - 3b$.

- **Inductive step:** 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 integers 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
$$p_i(-1^{5d^2}) \geq 2^i a + \left(3 \cdot 2^i - 3\right) b, \qquad p_1(y) \leq -b \ \forall y \neq -1^{5d^2}.$$
By Lemma 2.13, 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,

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

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

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

By Lemma 3.3, there exists a $j \in [5d^2]$ such that $p_i''(j) \leq -2^{i+1}a - \left(3 \cdot 2^{i+1} - 2\right) b$, and hence $p_i'(j) \leq -2^{i+1}a - \left(3 \cdot 2^{i+1} - 3\right) b$, by Equation 4. This implies the existence of an x 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 $p(x) < -2^{i+1}a - \left(3 \cdot 2^{i+1} - 3\right) b$.

□

4 Lower bounds

In this section, we prove our lower bounds. We first use linear programming duality to give us a sufficient approximation theoretic condition f for showing that the sign rank of $f \circ \mathsf{XOR}$ is large. Let $\delta > 0$ be a parameter, and X be any subset of $\{-1,1\}^n$.

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

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

Standard manipulations (as in [9], for example) and strong linear programming duality reveal that the optimum of the above linear program equals the optimum of the following program.

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

The variable Δ represents the worst margin guaranteed to exist at all points. At each point x over the smooth set, the dual polynomial has to better the worst margin by at least ξ_x. If the OPT is large, then it means that on average the dual polynomial did significantly better than the worst margin. Below is our main technical result of this section, which says that no such dual polynomial exists, even when the smoothness parameter δ is as high as $1/4$.

Theorem 4.1. *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 the optimal value,* OPT, *of (LP2) is at most* $2^{-\frac{l^{1/3}}{81}}$.

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 3.2 (set $m = l^{1/3} - \log l$) shows the existence of an OR polynomial p' on $l/8$ variables, of weight 1, such that

$$\text{For all } y \in \{-1,1\}^{l/8}, \qquad p'(y)\mathsf{OMB}^0(y) \geq \Delta_{\mathsf{OPT}} - 2\cdot 2^{-l^{1/3}} - 2\cdot 2^{-l^{1/3}}$$

$$\text{and} \qquad p'(-1^{l/8}) \geq \Delta + \frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{|X|} - 2\cdot 2^{-l^{1/3}} - 2\cdot 2^{-l^{1/3}}.$$

Note that

$$\mathsf{OPT} \geq 2^{-\frac{l^{1/3}}{81}} \implies \Delta_{\mathsf{OPT}} \geq 2^{-\frac{l^{1/3}}{81}} - \delta\frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{2^n} \tag{5}$$

p' satisfies the assumptions of Lemma 3.4 with $d = \deg(p') = l^{1/3} < \sqrt{l/32}$ (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 $b = \Delta_{\mathsf{OPT}} - 4\cdot 2^{-l^{1/3}}$.

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

Let us denote $k = l^{1/3}/80$ for the remaining of this proof. Thus, by Lemma 3.4, there exists an $x \in \{-1,1\}^{l/8}$ such that

$$\begin{aligned} |p'(x)| &\geq 2^k a + \left(3\cdot 2^k - 3\right) b \\ &\geq \Delta_{\mathsf{OPT}}(4\cdot 2^k - 3) + 2^k\frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{|X|} - 4\cdot 2^{-80k}(4\cdot 2^k - 3) \\ &\geq \left(4\cdot 2^k - 3\right)\left(2^{-\frac{l^{1/3}}{81}} - \delta\frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{2^n}\right) + 2^k\frac{\sum_{x\in X}\xi_{x,\mathsf{OPT}}}{|X|} - 4\cdot 2^{-80k}(4\cdot 2^k - 3) \\ &\qquad\qquad \text{Using Equation 5.} \\ &\geq \left(4\cdot 2^k - 3\right)\left(2^{-80k/81} - 4\cdot 2^{-80k}\right) \qquad \text{Since } \delta = 1/4. \\ &> 1 \qquad \text{Assuming } k > 81. \end{aligned}$$

This yields a contradiction, since p' was a polynomial of weight at most 1. □

Theorem 4.2. *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,*

$$sr(f \circ \mathsf{XOR}) \geq \frac{2^{l^{1/3}-2\log l}}{16}$$

Proof. Let $n = l^{4/3} - l\log l$. Theorem 4.1 tells us that the optimum of (LP2) (and hence (LP1), by duality) is at most $2^{-\frac{l^{1/3}}{81}}$, when $f = \mathsf{OMB}^0 \circ \bigvee_{l^{1/3}-\log l}$. 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 $\frac{l^2}{2^{l^{1/3}}}$, and hence $|X^c| \leq 2^n \cdot \frac{l^2}{2^{l^{1/3}}}$. By Lemma 2.12 and Theorem 2.8,

$$\begin{aligned} sr(f \circ \mathsf{XOR}) \geq sr(f\mu \circ \mathsf{XOR}) &\geq \frac{\frac{\delta}{2^n}2^{2n}}{\mathsf{OPT}\cdot 2^n + \frac{\delta}{2^n}\cdot h} \\ &\geq \frac{1/4}{2^{-\frac{l^{1/3}}{81}} + \frac{1}{4}\frac{|X^c|}{2^n}} \\ &\geq \frac{1/4}{2^{-\frac{l^{1/3}}{81}} + \frac{1}{4}l^2 \cdot 2^{-l^{1/3}}} \\ &\geq \frac{2^{l^{1/3}-2\log l}}{8} \end{aligned}$$

This gives us a function f on n input variables such that for large enough n,

$$sr(f \circ \mathsf{XOR}) \geq \frac{2^{n^{1/4}-\frac{3}{2}\log n}}{8}$$

□

Corollary 4.3. *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*

$$\mathsf{UPP}(f \circ \mathsf{XOR}) \geq n^{1/4} - \frac{3}{2}\log n - 3.$$

Proof. It follows from Theorem 4.2 and Theorem 2.11. □

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

Proof of Theorem 1.1. Suppose $\mathsf{OMB}^0 \circ \bigvee_{l^{1/3}-\log l} \circ \mathsf{XOR}_2$ could be represented by a $\mathsf{THR} \circ \mathsf{MAJ}$ circuit of size s. Let $n = 2l^{4/3} - 2l\log l$. By Lemma 2.9 and Theorem 4.2,

$$s\left(2l^{4/3} - 2l\log l\right) \geq sr(f) \geq \frac{2^{l^{1/3}-2\log l}}{8}.$$

Thus,

$$s \geq \frac{2^{l^{1/3}-\frac{10}{3}\log l}}{16} = 2^{\Omega\left(n^{1/4}\right)}.$$

□

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 $n = 2l^{4/3} - 2l \log l$. By Lemma 2.5, $f = \mathsf{OMB}^0 \circ \bigvee_{l^{1/3}-\log l} \circ \mathsf{XOR}_2$ can be computed by a $\mathsf{THR} \circ \mathsf{AND} \circ \mathsf{XOR}_2$ circuit of size n. 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 $2^{\Omega\left(n^{1/4}\right)}$. □

5 Conclusions

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

$$\widehat{LT}_1 \subsetneq LT_1 \subsetneq \widehat{LT}_2 = \mathsf{MAJ} \circ \mathsf{THR} \subsetneq \mathsf{THR} \circ \mathsf{MAJ} \subsetneq LT_2 \subseteq \widehat{LT}_3 \subseteq \mathrm{NP/poly}$$

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 [17] 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 [22] 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.

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.

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

[6] 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.

[7] Mark Bun and Justin Thaler. Improved bounds on the sign-rank of acˆ0. In *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, pages 37:1–37:14, 2016.

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

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

[10] 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.

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

[12] 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.

[13] 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.

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

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

[16] 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.

[17] 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.

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

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

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

[21] 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.

[22] 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.

[23] 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.

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

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

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

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

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

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

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

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

[32] 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.

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

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

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