CHICAGO JOURNAL OF THEORETICAL COMPUTER SCIENCE 2020, Article 01, pages 1–15
http://cjtcs.cs.uchicago.edu/

Lower bounds for linear decision lists

Arkadev Chattopadhyay Meena Mahajan Nikhil Mande*
Nitin Saurabh†

Received April 20, 2019; Revised May 05, 2020; Published May 30, 2020

Abstract: We demonstrate a lower bound technique for linear decision lists, which are decision lists where the queries are arbitrary linear threshold functions. We use this technique to prove an explicit lower bound by showing that any linear decision list computing the function $\mathsf{MAJ} \circ \mathsf{XOR}$ requires size $2^{0.18n}$. This completely answers an open question of Turán and Vatan [19]. We also show that the spectral classes $\mathsf{PL}_1, \mathsf{PL}_\infty$, and the polynomial threshold function classes $\widehat{\mathsf{PT}}_1, \mathsf{PT}_1$, are incomparable to linear decision lists.

Key words and phrases: linear threshold functions, decision lists, threshold circuits

1 Introduction

Decision lists are a widely studied model of computation, first introduced by Rivest [18]. A decision list L of size ℓ computing a Boolean function $f \in B_n$ is a sequence of $\ell - 1$ instructions of the form **if** $f_i(x) = a_i$ **then output** b_i **and stop**, followed by the instruction **output** $\neg b_{\ell-1}$ **and stop**. Here B_n denotes the set of all Boolean functions in n variables, each $f_i \in B_n$ is called a *query function*, and a_i and b_i are Boolean constants. If the functions f_i all belong to a function class $S \subseteq B_n$, then L is said to be an S-decision list.

Krause [15] showed that there are functions with small representation as AND-decision lists, but requiring exponential size when computed by depth-two circuits with a linear threshold gate at the top and XOR gats at the bottom. On the other hand, Impagliazzo and Williams [14] showed that a certain condition is sufficient to prove lower bounds against a related computation model that can be termed rectangle-decision lists. Linear decision lists are decision lists where the query functions are linear threshold functions. Lower bounds against linear decision lists (and even against bounded-rank linear decision trees, a natural generalisation) for the Inner Product modulo 2 function were proved by Gröger,

*This work was done while the author was a graduate student at TIFR, Mumbai.
†This work was done while the author was at Max Planck Institut für Informatik, Saarbrücken, Germany.

 DOI: 10.4086/cjtcs.2020.001

Turán and Vatan, in [10, 19]. Subsequently, Uchizawa and Takimoto [20, 21] showed lower bounds against the class of linear decision lists and linear decision trees when the weights of the linear threshold queries are bounded by a polynomial in the input length. In fact, the lower bounds of [20, 21] apply to any function with large *unbounded-error communication complexity*.

We observe that the lower bound argument in [19] shows that functions efficiently computable by linear decision lists (with no restrictions on the weights of the queried linear threshold functions) must have large monochromatic rectangles. In fact, we build on their argument to establish a more general result (Lemma 3.2). Informally, we show that if a function has no "large" weight monochromatic rectangles under some product distribution then it cannot be expressed by "small" linear decision lists. We then use this fact to establish a lower bound for a seemingly simple function, $\mathsf{MAJ} \circ \mathsf{XOR}$ (see Definition 4.1). Our main theorem is as follows.

Theorem 1.1. *Any linear decision list computing* $\mathsf{MAJ}_n \circ \mathsf{XOR}$ *must have size* $2^{\Omega(n)}$.

It is not hard to see that $\mathsf{MAJ} \circ \mathsf{XOR}$ can be simulated by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits with only a *linear* blow-up in size. This immediately yields the following corollary, resolving an open question posed by Turán and Vatan in [19].

Corollary 1.2. *There exists a function that can be computed by polynomial sized* $\mathsf{MAJ} \circ \mathsf{MAJ}$ *circuits, but any linear decision list computing it requires exponential size.*

Impagliazzo and Williams [14] demonstrated a function, implicitly computable by polynomial sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits, which cannot be computed by polynomial sized rectangle-decision lists. We observe that our lower bound technique against linear decision lists (Lemma 3.2) coincides with the sufficient condition considered in [14] to prove lower bounds against rectangle-decision lists. Thus, their function also separates linear decision lists from $\mathsf{MAJ} \circ \mathsf{MAJ}$. However, we obtain a $2^{\Omega(n)}$ lower bound on the length of linear decision lists in Theorem 1.1, improving upon the bound implicit in the work of Impagliazzo and Williams, which is worse in the exponent by at least a quadratic factor. Very recently, Chattopadhyay, Mande and Sherif [5] showed several properties of the function $\mathsf{SINK} \circ \mathsf{XOR}$. We observe that as a consequence, our lower bound technique against linear decision lists (Lemma 3.2) also applies to this function. We elaborate more on these remarks in Section 5.

2 Preliminaries

Definition 2.1 (Sign function)**.** The function $\mathsf{sign} : \mathbb{R} \to \{0,1\}$ is defined as follows.

$$\mathsf{sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

Definition 2.2 (Linear Threshold Functions)**.** A function $f : \{0,1\}^n \to \{0,1\}$ is said to be a linear threshold function (LTF) if there exist real numbers $w_0, w_1, \ldots, w_n$ such that $f(x) = \mathsf{sign}\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)$.

For strings $x, y \in \mathbb{R}^n$, we denote their inner product by $\langle x, y \rangle \triangleq \sum_i x_i y_i$. With this notation, f is an LTF if for some $w_0 \in \mathbb{R}$, $\tilde{w} \in \mathbb{R}^n$, $f(x) = \mathsf{sign}(w_0 + \langle \tilde{w}, x \rangle)$.

Definition 2.3 (Majority)**.** The function $\mathsf{MAJ}_n : \{0,1\}^n \to \{0,1\}$ is the linear threshold function defined by $\mathsf{MAJ}_n(x) = \mathsf{sign}\,(x_1 + x_2 + \cdots + x_n - n/2)$.

Definition 2.4 (Function composition)**.** For functions $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, the function $f \circ g : \{0,1\}^{nm} \to \{0,1\}$ is defined as follows:

$$f \circ g(x_{11}, \dots, x_{1m}, \dots, x_{n1}, \dots, x_{nm}) = f(g(x_{11}, \dots, x_{1m}), \cdots, g(x_{n1}, \dots, x_{nm})).$$

We now formally define the model of computation that is of interest in this paper.

Definition 2.5 (Linear Decision Lists)**.** A linear decision list (LDL) of size k is a sequence $(L_1, a_1), (L_2, a_2), \dots, (L_k, a_k)$, where each $a_i \in \{0,1\}$, and each L_i is an LTF with L_k being the constant function 1. The decision list computes a function $f : \{0,1\}^n \to \{0,1\}$ as follows : If $L_1(x) = 1$, then $f(x) = a_1$; elseif $L_2(x) = 1$, then $f(x) = a_2$; elseif ..., elseif $L_k(x) = 1$, then $f(x) = a_k$. That is,

$$f(x) = \bigvee_{i=1}^{k} \left(a_i \wedge L_i(x) \wedge \bigwedge_{j<i} \neg L_j(x) \right).$$

Definition 2.6 (Communication matrix)**.** For a function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, its communication matrix M_F is the $2^n \times 2^n$ matrix with entries $M_F[x,y] := F(x,y)$.

Definition 2.7 (Monochromatic rectangles/squares)**.** Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be any function. For $b \in \{0,1\}$, a monochromatic b-rectangle is a tuple (X,Y), where $X, Y \subseteq \{0,1\}^n$ and $F(x,y) = b$ for every $(x,y) \in X \times Y$. We say that (X,Y) is a monochromatic square of size s if it is a monochromatic 0-rectangle or 1-rectangle and, furthermore, $|X| = |Y| = s$.

Definition 2.8 (Product distributions and weights)**.** A probability distribution η over $\{0,1\}^n \times \{0,1\}^n$ is said to be a *product distribution* if there are probability distributions μ, ν over $\{0,1\}^n$ such that for every $(x,y) \in \{0,1\}^n \times \{0,1\}^n$, $\eta(x,y) = \mu(x) \times \nu(y)$. We say that η is the product distribution $\mu \times \nu$.

Given a probability distribution μ over $\{0,1\}^n$ and $X \subseteq \{0,1\}^n$, $\mu(X)$ is defined to be the sum $\sum_{x \in X} \mu(x)$. For a rectangle (X,Y), its *weight* under a product distribution $\mu \times \nu$ is $(\mu \times \nu)(X \times Y) = \mu(X) \times \nu(Y)$.

We will denote the number of 1's in a string $x \in \{0,1\}^n$ by $|x|$.

Definition 2.9 (Hamming distance)**.** The (Hamming) distance between any two strings $x, y \in \{0,1\}^n$, denoted $d(x,y)$, is defined as $d(x,y) \triangleq |\{i : x_i \neq y_i\}|$. The Hamming distance between any two sets $A, B \subseteq \{0,1\}^n$, denoted $d(A,B)$, is the minimum pairwise distance; $d(A,B) = \min_{x \in A, y \in B} d(x,y)$.

Definition 2.10 (Hamming balls)**.** Let $c \in \{0,1\}^n$ and $k \in \{1, \dots, n\}$. A set $A \subseteq \{0,1\}^n$ is called a Hamming ball with centre c and radius k if

$$\{s \in \{0,1\}^n \mid d(s,c) \leq k-1\} \subset A \subseteq \{s \in \{0,1\}^n \mid d(s,c) \leq k\}.$$

A singleton set $A = \{c\}$ is a Hamming ball with centre c and radius 0.

For a set $A \subseteq \{0,1\}^n$, the boundary of A is the set $\{s \in \{0,1\}^n \mid d(s,A) = 1\}$. In [13], Harper proved an isoperimetry result: among all sets of a given cardinality, Hamming balls have the smallest boundary set size. A simplified proof was given by Frankl and Füredi [8], who also stated the theorem in the equivalent form we mention below. (See also the presentation in [1]).

Theorem 2.11 (Harper's Theorem)**.** *Let $A, B \subseteq \{0,1\}^n$ be non-empty sets. Then, there exists a Hamming ball A_0 with centre 0^n and a Hamming ball B_0 with centre 1^n such that $|A_0| = |A|$, $|B_0| = B$, and $d(A_0, B_0) \geq d(A, B)$.*

Definition 2.12 (Binary Entropy)**.** The binary entropy function $\mathbb{H} : [0,1] \to [0,1]$ is defined as follows: $\mathbb{H}(p) = -p \log p - (1-p) \log(1-p)$.

Fact 2.13. $\mathbb{H}(1/4) < 0.82$.

3 Linear decision lists contain large monochromatic rectangles

The argument of Turán and Vatan from [19] implicitly showed that any function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with no large monochromatic squares cannot be computed by small linear decision lists. Their argument was presented specific to the Inner Product function (Theorem 1 in [19]). However, it is not too hard to see that their proof in fact works for any function as long as it has no large monochromatic squares. In this section, we generalize their argument to show that all functions computable by small size linear decision lists must contain, under *any product distribution*, a monochromatic rectangle of large weight with respect to that distribution.

We first establish a technical lemma that can be seen as a generalization of Lemma 2 in [19].

Lemma 3.1. *Let f be an* LTF *over the input variables $x_1, \ldots, x_n, y_1, \ldots, y_n$. Let μ, ν be distributions over $\{0,1\}^n$, and $X, Y \subseteq \{0,1\}^n$. Define $m := \min\{\mu(X), \nu(Y)\}$, and let $t \in (0, m]$. Then, one of the following is true.*

1. *There exists a monochromatic 1-rectangle (X', Y') within $X \times Y$ (i.e., $X' \subseteq X$ and $Y' \subseteq Y$) such that $\mu(X') \geq t$ and $\nu(Y') \geq t$.*
2. *There exists a monochromatic 0-rectangle (X', Y') within $X \times Y$ such that $\mu(X') > m - t$ and $\nu(Y') > m - t$.*

Proof. Let M be the submatrix of M_f restricted to $X \times Y$. Let the LTF f be given by $\mathsf{sign}(a + \langle \alpha \cdot x \rangle + \langle \beta \cdot y \rangle)$. Reorder the rows and columns of M in decreasing order of $a + \langle \alpha \cdot x \rangle$ and $\langle \beta \cdot y \rangle$ to get the matrix $B = R \times C$. Let i denote the least index of a row in B such that $\mu(\{R_1, \ldots R_i\}) \geq t$, and j denote the least index of a column in B such that $\mu(\{C_1, \ldots C_j\}) \geq t$. Note that these indices are well-defined since $t \in (0, m]$. If the $[i, j]$'th entry of B is 1, then the top-left submatrix of B satisfies item (1) in the lemma. If the $[i, j]$'th entry of B is 0, then the bottom-right submatrix of B satisfies item (2) in the lemma. ☐

We now prove the main lemma.

Lemma 3.2. *Let μ, ν be distributions on $\{0,1\}^n$. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be any function with no monochromatic rectangle of weight greater than w under the distribution $\mu \times \nu$. Then, any linear decision list computing f must have size at least $1/\sqrt{w}$.*

Proof. Towards a contradiction, let $(L_1,a_1),(L_2,a_2),\ldots,(L_k,a_k)$ be an LDL of size k computing f, where $k<1/\sqrt{w}$. Pick any $t\in(\sqrt{w},1/k]$. We construct, for each $i\in[k-1]$, a rectangle $S_i=X_i\times Y_i$ which is a 0-rectangle for all L_j with $j\le i$, and furthermore $\mu(X_i),\nu(Y_i)\ge 1-i\cdot t$. We proceed by induction on i.

For the base case $i=1$, let $S_0=(X_0,Y_0)$ be the entire $2^n\times 2^n$ matrix. Suppose S_0 has a rectangle (X',Y') that is a 1-rectangle of L_1 and moreover, $\mu(X')\ge t$, $\nu(Y')\ge t$. Then everywhere in this rectangle, f will be a_1. But f has no monochromatic rectangle of weight as large as $t^2>w$. So S_0 has no rectangle (X',Y') with $\mu(X')\ge t$, $\nu(Y')\ge t$ that is a 1-rectangle of L_1. By Lemma 3.1, S_0 must then contain a 0-rectangle (X_1,Y_1) of L_1 such that both $\mu(X_1)$ and $\nu(Y_1)$ are at least $1-t$. This establishes the base case.

For the inductive step, we have a rectangle $S_{i-1}=(X_{i-1},Y_{i-1})$ which is a 0-rectangle for $L_1,L_2,\ldots,L_{i-1}$ and, moreover, $\min\{\mu(X_{i-1}),\nu(Y_{i-1})\}\ge 1-(i-1)t$. Within this rectangle, suppose L_i has a 1-rectangle (X',Y') such that $\mu(X')\ge t$ and $\nu(Y')\ge t$. Then $f=a_i$ in this rectangle, giving a monochromatic rectangle of f of weight greater than w. But we know that such rectangles do not exist. Since $kt\le 1$ and $i<k$, we have $t\le 1-(i-1)t$ and hence Lemma 3.1 is applicable. Hence we conclude that S_{i-1} must contain a 0-rectangle (X_i,Y_i) of L_i with $\min\{\mu(X_i),\nu(Y_i)\}\ge 1-(i-1)t-t=1-it$. Since this rectangle, say S_i, is contained in S_{i-1}, it is a 0-rectangle for all L_j with $j\le i$.

Thus, we have a rectangle $S_{k-1}=(X_{k-1},Y_{k-1})$ on which $L_1,L_2,\ldots,L_{k-1}$ are 0, and $L_k=1$ because L_k is the constant function 1. Furthermore, $\mu(X_{k-1})$ and $\nu(Y_{k-1})\ge 1-(k-1)t$. Everywhere on this rectangle, f evaluates to a_k. So S_{k-1} is a monochromatic rectangle for f. Hence it cannot have weight more than w. Thus $1-(k-1)t\le\sqrt{w}<t$; that is, $1<kt$, contradicting our choice of t. □

4 MAJ ∘ XOR has no large monochromatic squares

In this section, we show an upper bound and a matching tight lower bound on the size of a largest monochromatic square in the communication matrix of the MAJ ∘ XOR function.

Definition 4.1 (XOR functions)**.** For a function $f:\{0,1\}^n\to\{0,1\}$, let $f\circ\mathsf{XOR}$ denote the function defined by $f\circ\mathsf{XOR}(x_1,\ldots,x_n,y_1,\ldots y_n)=f(x_1\oplus y_1,\ldots,x_n\oplus y_n)$.

Lemma 4.2. *Let* $F:\{0,1\}^n\times\{0,1\}^n\to\{0,1\}$ *be the function* $\mathsf{MAJ}_n\circ\mathsf{XOR}$*. Then, for any* $b\in\{0,1\}$*,* M_F *has a monochromatic b-square of size at least* $\sum_{i=0}^{\lfloor n/4\rfloor}\binom{n}{i}$.

Proof. Define the sets X,Y,Z as follows:

$$\begin{aligned} X=Y &= \{x\in\{0,1\}^n : |x|\le\lfloor n/4\rfloor\}.\\ Z &= \{x\in\{0,1\}^n : |x|\ge n-\lfloor n/4\rfloor\}. \end{aligned}$$

Note that $F(x,y)=0$ for all $x\in X, y\in Y$, and $F(x,z)=1$ for all $x\in X$, $z\in Z$. Thus (X,Y) and (X,Z) are a monochromatic 0-square and 1-square, respectively, each of size $\sum_{i=0}^{\lfloor n/4\rfloor}\binom{n}{i}$. □

Remark 4.3. We remark that when $n\equiv 3 \pmod 4$ the above construction can be improved if we consider monochromatic rectangles. That is, for any $b\in\{0,1\}$, M_F has a monochromatic b-rectangle

(X_1,X_2) such that $|X_1| = \sum_{i=0}^{\lceil n/4\rceil}\binom{n}{i}$ and $|X_2| = \sum_{i=0}^{\lfloor n/4\rfloor}\binom{n}{i}$. Indeed, let $X = \{x \in \{0,1\}^n : |x| \leq \lceil n/4\rceil\}$, $Y = \{x \in \{0,1\}^n : |x| \leq \lfloor n/4\rfloor\}$ and $Z = \{x \in \{0,1\}^n : |x| \geq n - \lfloor n/4\rfloor\}$. Then, it is easily seen that (X,Z) (resp., (X,Y)) is a monochromatic 1-rectangle (resp., 0-rectangle) of the claimed size.

We now show that this bound is tight.

Theorem 4.4. *Let $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be the function $\mathsf{MAJ}_n \circ \mathsf{XOR}$. For any n, M_F has no monochromatic squares of size greater than* $\sum_{i=0}^{\lceil n/4\rceil}\binom{n}{i}$.

Proof. Suppose, to the contrary, that there are sets $A,B \subseteq \{0,1\}^n$ such that $|A| = |B| > \sum_{i=0}^{\lceil n/4\rceil}\binom{n}{i}$ and $A \times B$ is a monochromatic 1-square in M_F. By the definition of F, this implies $d(A,B) > \lfloor n/2\rfloor$. By Theorem 2.11, there exist Hamming balls A_0 around 0^n, and B_0 around 1^n such that $|A_0| = |A|, |B_0| = |B|$ and $d(A_0,B_0) \geq d(A,B)$. The size lower bound enforces that the radius of A_0 and B_0 must be greater than $\lceil n/4\rceil$, and since they are centered on 0^n and 1^n, it follows that $d(A_0,B_0) \leq \lfloor n/2\rfloor$. But then $d(A,B)$ is also at most $\lfloor n/2\rfloor$. Hence, there exist $x \in A, y \in B$ such that $d(x,y) \leq \lfloor n/2\rfloor$, which means $F(x,y) = \mathsf{MAJ}_n \circ \mathsf{XOR}(x,y) = 0$, which contradicts our assumption. Therefore, any monochromatic 1-square in M_F has size at most $\sum_{i=0}^{\lceil n/4\rceil}\binom{n}{i}$.

A similar argument shows the same upper bound on the size of monochromatic 0-squares. □

Now we can put things together to prove our main theorem.

Proof of Theorem 1.1. Let s_n be the minimum size of an LDL computing $\mathsf{MAJ}_n \circ \mathsf{XOR}$. Further let μ and ν be uniform distributions over $\{0,1\}^n$. Then, by Lemma 3.2 and Theorem 4.4, for all n sufficiently large,

$$\begin{aligned} s_n &\geq \frac{2^n}{\sum_{i=0}^{\lceil n/4\rceil}\binom{n}{i}} \\ &\geq \frac{2^n}{2^{n\cdot H(1/4)}} && \text{using Stirling's approximation} \\ &\geq 2^{0.18n}. && \text{using Fact 2.13} \end{aligned}$$

□

5 LDLs and the threshold circuit hierarchy

In this section, we see how the class of functions computable by polynomial sized LDLs fits into the low depth threshold circuit hierarchy. The reader is referred to Razborov's survey [17] for a detailed exposition on the low depth threshold circuits hierarchy.

5.1 Definitions

Definition 5.1 (MAJ)**.** Define MAJ to be the class of all functions computable by polynomial sized MAJ gates. Each input to the MAJ gate may be a constant 0 or 1, or a variable x_i, or its negation $\neg x_i$.

Definition 5.2 (LTF)**.** We denote the class of all linear threshold functions by LTF.

(We note that this is an overload of notation. LTF can now denote a single linear threshold function as well as the class of linear threshold functions. However, we remark that resolving this ambiguity will be clear from the context.)

Definition 5.3 (LDL)**.** Define LDL to be the class of all functions computable by polynomial sized linear decision lists.

Definition 5.4 ($\widehat{\mathsf{LDL}}$)**.** Define $\widehat{\mathsf{LDL}}$ to be the class of all functions computable by polynomial sized linear decision lists where, furthermore, weights of the linear threshold queries are integers with values bounded by a polynomial in the number of variables.

Definition 5.5 (Depth-2 classes)**.** For classes of functions $\mathcal{C}, \mathcal{D}$, define $\mathcal{C} \circ \mathcal{D}$ to be the class of functions computable by polynomial-sized depth-2 circuits, where the top gate computes a function from the class $\mathcal{C}$, and the bottom layer contains gates computing functions in $\mathcal{D}$.

Definition 5.6 ($\widehat{\mathsf{PT}_1}$)**.** The class $\widehat{\mathsf{PT}_1}$ consists of all functions $f : \{0,1\}^n \to \{0,1\}$ which can be represented by polynomial sized $\mathsf{MAJ} \circ \mathsf{PARITY}$ circuits.

Definition 5.7 (PT_1)**.** The class PT_1 consists of all functions $f : \{0,1\}^n \to \{0,1\}$ which can be represented by polynomial sized $\mathsf{LTF} \circ \mathsf{PARITY}$ circuits.

(These are precisely the classes of polynomial threshold functions [2]; it is more convenient for us here to use the equivalent formulation as depth-2 circuits.)

In order to define classes given by the spectral representation of functions, we first recall a few preliminaries from Boolean function analysis.

Consider the real vector space of functions from $\{0,1\}^n \to \mathbb{R}$, equipped with the following inner product.

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x) = \mathbb{E}_{x \in \{0,1\}^n}[f(x)g(x)].$$

For each $S \subseteq [n]$, define $\chi_S : \{0,1\}^n \to \{-1,1\}$ by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. It is not hard to verify that $\{\chi_S : S \subseteq [n]\}$ forms an orthonormal basis for this vector space. Thus, every $f : \{0,1\}^n \to \mathbb{R}$ has a unique representation as $f = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S$, where

$$\widehat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}_{x \in \{0,1\}^n}[f(x)\chi_S(x)].$$

Definition 5.8 (PL_1)**.** The class PL_1 consists of all functions $f : \{0,1\}^n \to \{0,1\}$ for which $\sum_{S \subseteq [n]} |\widehat{f}(S)| \leq \text{poly}(n)$.

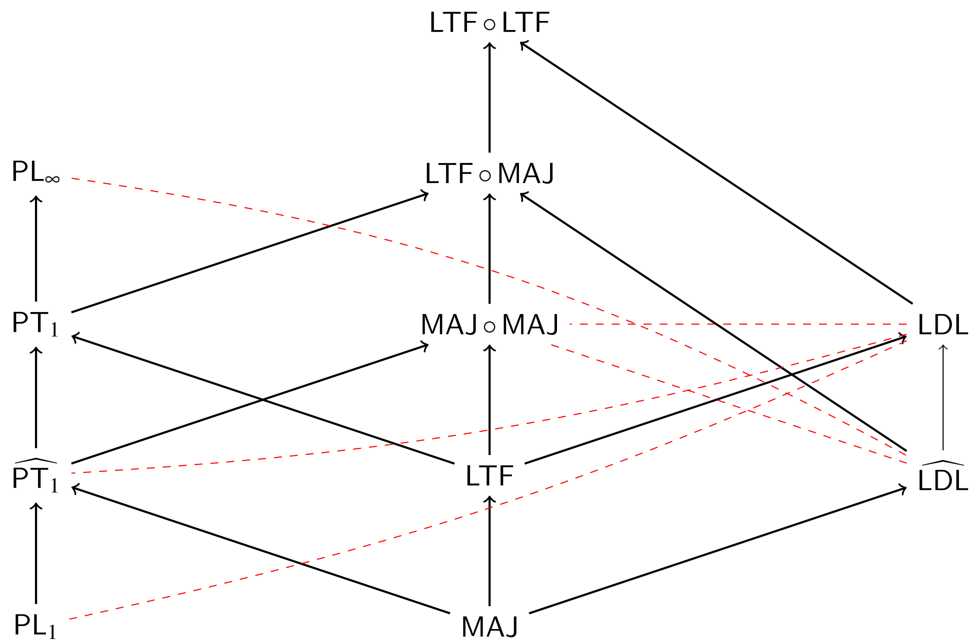

Figure 1: Low depth threshold circuit hierarchy

Definition 5.9 (PL_∞)**.** The class PL_∞ consists of all functions $f : \{0,1\}^n \to \{0,1\}$ for which $\max_{S \subseteq [n]} |\widehat{f}(S)| \geq \frac{1}{\mathrm{poly}(n)}$.

Figure 1 depicts the currently known status of low depth circuit class containments, and shows where linear decision lists fit in this hierarchy.

A thick solid arrow from $\mathcal{C}_1$ to $\mathcal{C}_2$ denotes $\mathcal{C}_1 \subsetneq \mathcal{C}_2$, a thin solid arrow from $\mathcal{C}_1$ to $\mathcal{C}_2$ denotes $\mathcal{C}_1 \subseteq \mathcal{C}_2$[1], and a dashed line between $\mathcal{C}_1$ and $\mathcal{C}_2$ denotes incomparability. In the figure, we only show the newly established incomparabilities.

The leftmost column has the classes defined based on spectral representation, and the middle column has the classes based on depth-2 circuits. Concerning these classes, the picture was already completely clear: All containments shown among classes in these columns are known to be strict, and wherever no arrow connects two classes, they are known to be incomparable. Essentially this part of the figure appears in [9]; a subsequent refinement is the insertion of the class $\mathsf{LTF} \circ \mathsf{MAJ}$, separated from $\mathsf{MAJ} \circ \mathsf{MAJ}$ in [9], from PT_1 in [2] and most recently from $\mathsf{LTF} \circ \mathsf{LTF}$ in [4].

The two classes $\widehat{\mathsf{LDL}}$ and LDL form the new column on the right. In the following subsection we explain their position with respect to the other two columns. However here the picture is not yet completely clear, and there are still several open questions.

[1]There is only one thin arrow, between $\widehat{\mathsf{LDL}}$ and LDL.

5.2 New results

By definition, $\mathsf{MAJ} \subseteq \widehat{\mathsf{LDL}}$ and $\mathsf{LTF} \subseteq \mathsf{LDL}$ via lists of size 2. The parity function is known to not be in LTF, and it has a simple LDL with 0-1 weights in the query functions. Thus both these containments are proper, and $\widehat{\mathsf{LDL}}$ is not contained in LTF. We now observe that, implicit from prior work, $\widehat{\mathsf{LDL}}$ is not even contained in $\mathsf{MAJ} \circ \mathsf{MAJ}$.

Theorem 5.10.

$$\widehat{\mathsf{LDL}} \not\subseteq \mathsf{MAJ} \circ \mathsf{MAJ}.$$

Proof. Define the ODD-MAX-BIT function by $\mathsf{OMB}(x) = 1$ iff the largest index i where $x_i = 1$ is odd ($\mathsf{OMB}(0^n) = 0$). Buhrman, Vereshchagin and de Wolf [3] showed that $\mathsf{OMB} \circ \mathsf{AND}$ is hard, in the sense that it has exponentially small *discrepancy*. By a result of Hajnal, Maass, Pudlák, Szegedy and Turán [11], this implies that $\mathsf{OMB} \circ \mathsf{AND}$ cannot be computed by polynomial sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits.

Note that OMB can be computed by a linear sized decision list by querying the variables in decreasing order of their indices. Thus $\mathsf{OMB} \circ \mathsf{AND}$ can be computed by a linear sized decision list of AND's, and hence by a linear decision list with 0-1 weights. □

On the other hand, it is easily seen that $\mathsf{MAJ}_n \circ \mathsf{XOR}$ is in $\mathsf{MAJ} \circ \mathsf{MAJ}$, and even in $\widehat{\mathsf{PT}}_1$ (see for instance [2]). Combining this with Theorem 1.1, we obtain:

Theorem 5.11.

$$\widehat{\mathsf{PT}}_1 \not\subseteq \mathsf{LDL}.$$

The following strengthening of Theorem 5.11 is implicit from a recent result of Chattopadhyay, Mande and Sherif [5].

Theorem 5.12.

$$\mathsf{PL}_1 \not\subseteq \mathsf{LDL}.$$

(We defer a discussion of why Theorem 5.12 holds to Section 5.3.) Putting together these separations with the known containments $\mathsf{PL}_1 \subseteq \widehat{\mathsf{PT}}_1 \subseteq \mathsf{MAJ} \circ \mathsf{MAJ}$, we obtain a slew of incomparability results.

Corollary 5.13. *For any class $A \in \{\widehat{\mathsf{LDL}}, \mathsf{LDL}\}$ and $B \in \{\mathsf{PL}_1, \mathsf{MAJ} \circ \mathsf{MAJ}\}$, the classes A and B are incomparable.*

In particular, the classes LDL and $\mathsf{MAJ} \circ \mathsf{MAJ}$ are incomparable. This completely answers the open question posed by Turán and Vatan [19].

Impagliazzo and Williams [14, Theorem 4.8] showed that the function $\mathsf{OR}_n \circ \mathsf{EQ}_n$ (also called Block-Equality) does not contain large monochromatic rectangles (in fact they showed that it does not contain large monochromatic rectangles under any product distribution). Thus, by Lemma 3.2, any linear decision list computing $\mathsf{OR}_n \circ \mathsf{EQ}_n$ must be of size at least $2^{\Omega(n)}$. We now observe that $\mathsf{OR} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{MAJ}$. Consequently, $\mathsf{OR} \circ \mathsf{EQ}$ also witnesses $\mathsf{MAJ} \circ \mathsf{MAJ} \not\subseteq \mathsf{LDL}$. However, in contrast to Theorem 1.1, note that the lower bound is subexponential since $\mathsf{OR} \circ \mathsf{EQ}$ is defined on $2n^2$ variables. Moreover, $\mathsf{OR} \circ \mathsf{EQ}$ seems to incur a significant polynomial blow up in size when simulated by $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits, whereas $\mathsf{MAJ}_n \circ \mathsf{XOR}$ has linear sized $\mathsf{MAJ} \circ \mathsf{MAJ}$ circuits.

Theorem 5.14.

$$\mathsf{OR} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{MAJ}.$$

Proof. First observe that $\mathsf{OR} \circ \mathsf{EQ}$ can be computed by a $\mathsf{MAJ} \circ \mathsf{EQ}$ circuit by suitably padding constants to the input. Next, note that EQ is an *exact threshold function*, that is there exist reals $a_1, \ldots, a_n, b_1, \ldots, b_n, c$ such that $\mathsf{EQ}(x,y) = 1$ iff $\sum_{i=1}^{n} a_i x_i + b_i y_i = c$. Hansen and Podolskii [12] showed that such functions can be efficiently simulated by $\mathsf{MAJ} \circ \mathsf{LTF}$ circuits. However, we do not need the full strength of their result, so we give a direct construction below.

For an equality on $2n$ bits, say $x_1, \ldots, x_n, y_1, \ldots, y_n$, note that

$$\mathsf{EQ}_n(x_1, \ldots, x_n, y_1, \ldots, y_n) = 1 \iff \sum_{i=1}^{n} 2^i (x_i - y_i) = 0.$$

Consider the following linear threshold functions.

$$g_1(x,y) = \mathsf{sign}\left(\sum_{i=1}^{n} 2^i (x_i - y_i) + 1/2 \right) \text{ and}$$
$$g_2(x,y) = \mathsf{sign}\left(\sum_{i=1}^{n} 2^i (x_i - y_i) - 1/2 \right).$$

Observe that $g_1(x,y) - g_2(x,y) = \mathsf{EQ}_n(x,y)$.

Let $g_1^{(i)}$ and $g_2^{(i)}$ denote these LTFs for the ith block on which we test equality. The function $\mathsf{OR}_n \circ \mathsf{EQ}_n$ is just

$$\mathsf{OR}_n \circ \mathsf{EQ}_n = \mathsf{sign}\left((g_1^{(1)} - g_2^{(1)}) + (g_1^{(2)} - g_2^{(2)}) + \ldots + (g_1^{(n)} - g_2^{(n)}) \right);$$

this formulation puts it in $\mathsf{MAJ} \circ \mathsf{LTF}$.

Finally, Goldmann, Håstad and Razborov [9] showed that $\mathsf{MAJ} \circ \mathsf{LTF} = \mathsf{MAJ} \circ \mathsf{MAJ}$. Thus, $\mathsf{OR} \circ \mathsf{EQ} \in \mathsf{MAJ} \circ \mathsf{MAJ}$. □

Theorem 5.15.

$$\widehat{\mathsf{LDL}} \not\subseteq \mathsf{PL}_\infty.$$

Proof. It is easy to see that any symmetric function (a function that only depends on the Hamming weight of the input) can be computed by linear sized linear decision lists where query functions are majority: the linear threshold queries can be used to determine the Hamming weight of the input, and the decision list outputs the appropriate answer at each decision.

Bruck [2] showed that the *Complete Quadratic* function, which is a symmetric function, is not in PL_∞. This function yields the required separation. □

Combining Corollary 5.13 and Theorem 5.15 yields more incomparability results.

Corollary 5.16. *For any class $A \in \{\widehat{\mathsf{LDL}}, \mathsf{LDL}\}$ and $B \in \{\mathsf{PL}_1, \mathsf{PL}_\infty\}$, the classes A and B are incomparable. In other words, all spectral classes in the first column (see Figure 1) are incomparable to all classes in the third column.*

Finally, as noted in [19], LDL is contained in $\mathsf{LTF} \circ \mathsf{LTF}$. The same argument shows that $\widehat{\mathsf{LDL}}$ is contained in $\mathsf{LTF} \circ \mathsf{MAJ}$. Corollary 5.13 implies that these containments are strict.

5.3 Proving Theorem 5.12

As mentioned earlier, it is implicit from a recent result of Chattopadhyay et al. [5] that $\mathsf{PL}_1 \not\subseteq \mathsf{LDL}$. We first define the function used to achieve the separation and introduce some background required.

Definition 5.17 (SINK, [5])**.** Consider a complete undirected graph on n vertices with variables $x_{i,j}$ for $i < j \in [n]$. The variable $x_{i,j}$ assigns a direction to the edge between v_i and v_j in the following way: $x_{i,j} = 0$ implies the edge points towards v_i, and $x_{i,j} = 1$ implies the edge points towards v_j. The function SINK computes whether or not there is a sink in the graph. In other words,

$$\mathsf{SINK}(x) = 1 \iff \exists i \in [n] \text{ such that all edges adjacent to } i \text{ are incoming.}$$

We now define the notion of *projections* of strings to certain subsets of coordinates. Let $X \in \{0,1\}^{\binom{n}{2}}$. For any vertex v_i, let E_{v_i} be the set of $n-1$ coordinates corresponding to the $n-1$ edges adjacent to v_i. Let X_{v_i} denote the $(n-1)$-bit string obtained by projecting X to the coordinates in E_{v_i}.

Definition 5.18 (Entropy)**.** Let X be a discrete random variable. The entropy $H(X)$ is defined as

$$H(X) = \sum_{s \in \mathrm{supp(X)}} \Pr[X = s] \log \frac{1}{\Pr[X = s]}.$$

Fact 5.19. *[7, Theorem 2.6.4]* $\mathrm{supp}(X) = k \implies H(X) \le \log k$*, with equality if and only if X is uniform.*

Lemma 5.20 (Shearer's Lemma [6] (see also [16]))**.** *Let $X = (X_1, \ldots, X_t)$ be a random variable. If S is a set of projections such that for each $i \in [t]$, i appears in at least k projections, then $\sum_{P \in S}[H_{X_P}] \ge kH(X)$.*

Chattopadhyay et al. [5] introduced and used the function $\mathsf{SINK} \circ \mathsf{XOR}$ to refute the long-standing Log-Approximate-Rank Conjecture, along with several other conjectures. They observe that $\mathsf{SINK} \circ \mathsf{XOR} \in \mathsf{PL}_1$ [5, Theorem 1.9].

Lemma 5.21 (Part 1 of Theorem 1.9 in [5])**.**

$$\mathsf{SINK} \circ \mathsf{XOR} \in \mathsf{PL}_1.$$

It is also implicit from their work that $\mathsf{SINK} \circ \mathsf{XOR}$ does not contain large monochromatic rectangles under the uniform distribution. More precisely, plugging the value $\varepsilon = 0$ in [5, Claim 6.4] implies that any monochromatic rectangle in the communication matrix of $\mathsf{SINK} \circ \mathsf{XOR}$ on $2\binom{n}{2}$ variables must have weight at most $2^{2\binom{n}{2} - \Omega(n)}$. However, we do not require the full power of their proof for our purpose, and therefore produce a self-contained proof below.

Theorem 5.22. *Any monochromatic rectangle $R = A \times B$ in the communication matrix of* $\mathsf{SINK} \circ \mathsf{XOR}$ *must satisfy $|R| \le 2^{2\binom{n}{2} - n + \log n + 1}$.*

Proof. It is easy to verify that the probability of a 1-input under the uniform distribution equals $n/2^{n-1}$. Hence if R is a 1-monochromatic rectangle, then $|R| \le 2^{2\binom{n}{2}} \times n/2^{n-1}$, as claimed in the theorem.

Let $R = A \times B$ be a 0-monochromatic rectangle. Consider the random variable XY (X concatenated with Y) over $2\binom{n}{2}$ coordinates, when X and Y are sampled uniformly from A and B, respectively. From Fact 5.19 we have $H(XY) = \log |R|$.

Let S be the set of projections $S := \{E_{v_i} \mid 1 \le i \le n\}$. Then each coordinate appears in exactly two projections. Hence by Lemma 5.20,

$$2H(XY) \le \sum_{P \in S} H((XY)_P) = \sum_{i \in [n]} H((XY)_{v_i}).$$

We now bound the entropy in XY restricted to each of the projections. Let A_{v_i} and B_{v_i} be the projections of A and B on E_{v_i}, respectively. Since there is no input in R which is a sink, we have $|\mathrm{supp}(A_{v_i})| + |\mathrm{supp}(B_{v_i})| \le 2^{n-1}$. (Each string in A_{v_i} rules out one string from B_{v_i} and vice versa.) By the AM-GM inequality, $|\mathrm{supp}(A_{v_i})| \cdot |\mathrm{supp}(B_{v_i})| \le 2^{2n-4}$. Hence Fact 5.19 implies that $H((XY)_{v_i}) \le 2n-4$.

Returning to our use of Lemma 5.20, we obtain

$$\begin{aligned} 2H(XY) &\le \sum_{P \in S} H((XY)_P) \le n(2n-4) \\ \implies H(XY) &\le 2\binom{n}{2} - n \\ \implies |R| &\le 2^{2\binom{n}{2} - n}. \end{aligned}$$

□

Along with Lemma 3.2, Theorem 5.22 shows that any linear decision list computing the function $\mathsf{SINK} \circ \mathsf{XOR}$ on $2\binom{n}{2}$ variables (which is in PL_1) must have size at least $2^{n/2}$. This completes the proof of Theorem 5.12.

Clearly, $\mathsf{SINK} \circ \mathsf{XOR}$ also witnesses $\mathsf{MAJ} \circ \mathsf{MAJ} \not\subseteq \mathsf{LDL}$. However, the lower bound against LDL is again only subexponential.

6 Conclusions

We show that $\mathsf{MAJ} \circ \mathsf{XOR}$ cannot be computed by polynomial sized linear decision lists, resolving an open question of Turán and Vatan [19]. We also show that several spectral classes and polynomial threshold function classes are incomparable to linear decision lists. Figure 1 depicts where the class LDL, and its small-weight version $\widehat{\mathsf{LDL}}$, fit in the low depth threshold circuit hierarchy.

A subset of the authors [4] showed that a decision list of *exact threshold functions* cannot be computed by $\mathsf{LTF} \circ \mathsf{MAJ}$. A natural question that arises is whether LDL is incomparable with $\mathsf{LTF} \circ \mathsf{MAJ}$. (Note that the function from [4] separating $\mathsf{LTF} \circ \mathsf{LTF}$ from $\mathsf{LTF} \circ \mathsf{MAJ}$ does not settle this question as it is also not in LDL – it contains the function $\mathsf{OR} \circ \mathsf{EQ}$ as a subfunction.)

Another natural question is whether $\widehat{\mathsf{LDL}}$ is strictly contained in LDL; that is, whether weights matter in linear decision lists.

7 Acknowledgments

We thank Rahul Santhanam for discussions concerning decision lists. We thank Jaikumar Radhakrishnan for referring us to Harper's theorem.

References

[1] BÉLA BOLLOBÁS: *Combinatorics: set systems, hypergraphs, families of vectors, and combinatorial probability*. Cambridge University Press, 1986. 4

[2] JEHOSHUA BRUCK: Harmonic analysis of polynomial threshold functions. *SIAM J. Discrete Math.*, 3(2):168–177, 1990. 7, 8, 9, 10

[3] 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, pp. 24–32. IEEE Computer Society, 2007. 9

[4] ARKADEV CHATTOPADHYAY AND NIKHIL S. MANDE: A short list of equalities induces large sign rank. In *Proc. 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2018. Preliminary version in ECCC TR 2017-083. 8, 12

[5] ARKADEV CHATTOPADHYAY, NIKHIL S. MANDE, AND SUHAIL SHERIF: The log-approximate-rank conjecture is false. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pp. 42–53. ACM, 2019. Preliminary version in ECCC TR 2018-176. 2, 9, 11

[6] FAN R. K. CHUNG, RONALD L. GRAHAM, PETER FRANKL, AND JAMES B. SHEARER: Some intersection theorems for ordered sets and graphs. *Journal of Combinatorial Theory, Ser. A*, 43(1):23–37, 1986. 11

[7] THOMAS M. COVER AND JOY A. THOMAS: *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006. 11

[8] PETER FRANKL AND ZOLTÁN FÜREDI: A short proof for a theorem of Harper about Hamming-spheres. *Discrete Mathematics*, 34(3):311–313, 1981. 4

[9] MIKAEL GOLDMANN, JOHAN HÅSTAD, AND ALEXANDER A. RAZBOROV: Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992. 8, 10

[10] HANS DIETMAR GRÖGER AND GYÖRGY TURÁN: On linear decision trees computing boolean functions. In *Automata, Languages and Programming, 18th International Colloquium, ICALP91, Madrid, Spain, July 8-12, 1991, Proceedings*, pp. 707–718, 1991. 2

[11] ANDRÁS HAJNAL, WOLFGANG MAASS, PAVEL PUDLÁK, MARIO SZEGEDY, AND GYÖRGY TURÁN: Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993. 9

[12] 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*, pp. 270–279, 2010. 10

[13] L. H. HARPER: Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1:385–393, 1966. 4

[14] 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*, pp. 259–269, 2010. 1, 2, 9

[15] MATTHIAS KRAUSE: On the computational power of boolean decision lists. *Computational Complexity*, 14(4):362–375, 2006. 1

[16] JAIKUMAR RADHAKRISHNAN: Entropy and counting. In J.C. MISRA, editor, *Computational Mathematics, Modelling and Algorithms*. Narosa Publishers, New Delhi, 2001. IIT Kharagpur Golden Jubilee Volume. 11

[17] ALEXANDER A. RAZBOROV: On small depth threshold circuits. In *Third Scandinavian Workshop on Algorithm Theory (SWAT)*, pp. 42–52, 1992. 6

[18] RONALD L. RIVEST: Learning decision lists. *Machine Learning*, 2(3):229–246, 1987. 1

[19] GYÖRGY TURÁN AND FARROKH VATAN: Linear decision lists and partitioning algorithms for the construction of neural networks. In *Foundations of Computational Mathematics*, pp. 414–423, Berlin, Heidelberg, 1997. Springer Berlin Heidelberg. 1, 2, 4, 9, 10, 12

[20] KEI UCHIZAWA AND EIJI TAKIMOTO: Lower bounds for linear decision trees via an energy complexity argument. In *Mathematical Foundations of Computer Science 2011 - 36th International Symposium, MFCS 2011, Warsaw, Poland, August 22-26, 2011. Proceedings*, pp. 568–579, 2011. 2

[21] KEI UCHIZAWA AND EIJI TAKIMOTO: Lower bounds for linear decision trees with bounded weights. In *SOFSEM 2015: Theory and Practice of Computer Science*, pp. 412–422, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg. 2

AUTHORS

Arkadev Chattopadhyay
Professor
Tata Institute of Fundamental Research, Mumbai, India
arkadev.c@tifr.res.in
https://www.tcs.tifr.res.in/~arkadev/

Meena Mahajan
Professor
The Institute of Mathematical Sciences, HBNI, Chennai, India
meena@imsc.res.in
https://www.imsc.res.in/~meena/

Nikhil Mande
Postdoctoral Fellow
Georgetown University, Washington, D.C. USA
nikhil.mande@georgetown.edu
http://nikhil.georgetown.domains/

Nitin Saurabh
Postdoctoral Fellow
Technion - I.I.T, Haifa, Israel
nitinsau@cs.technion.ac.il
https://people.mpi-inf.mpg.de/~nsaurabh/