On DNF approximators for monotone Boolean functions

Eric Blais*
MIT
eblais@csail.mit.edu

Johan Håstad†
KTH Royal Institute of Technology
johanh@kth.se

Rocco A. Servedio‡
Columbia University
rocco@cs.columbia.edu

Li-Yang Tan§
Columbia University
liyang@cs.columbia.edu

November 27, 2013

Abstract

We study the complexity of approximating monotone Boolean functions with disjunctive normal form (DNF) formulas, exploring two main directions. In the first direction we construct DNF approximators for arbitrary monotone functions achieving one-sided error: we show that every monotone f can be ε-approximated by a DNF g of size $2^{n-\Omega_\varepsilon(\sqrt{n})}$ satisfying $g(x) \leq f(x)$ for all $x \in \{0,1\}^n$. This represents the first non-trivial universal upper bound even for DNF approximators incurring two-sided error.

In the second direction we study the power of negations in DNF approximators for monotone functions. We exhibit monotone functions for which non-monotone DNFs perform better than monotone ones, giving separations with respect to both DNF size and width. Our results, when taken together with a classical theorem of Quine [Qui54], highlight an interesting contrast between approximation and exact computation in the DNF complexity of monotone functions, and they add to a line of work on the surprising role of negations in monotone complexity [Raz85, Oko82, AG87].

*Supported by a Simons Postdoctoral Fellowship.

†Supported by ERC Advanced Investigator Grant 226203. Work done while visiting the Simons Institute.

‡Supported by NSF grants CCF-1115703 and CCF-1319788.

§Supported by NSF grants CCF-1115703 and CCF-1319788. Part of this research was completed while visiting KTH Royal Institute of Technology, supported by ERC Advanced Investigator Grant 226203.

1 Introduction

Monotone Boolean functions constitute a rich and complex class of functions, and their structural and combinatorial properties have been intensively studied for decades; see e.g. the monograph [Kor03] for an in-depth survey. In complexity theory monotone functions play an especially important role in circuit complexity, where Razborov's celebrated result [Raz85] has led to a significant body of work centered around monotone functions and the circuits that compute them [AB87, AG87, KW88, RW90, KRW91, GS95, RR97, GH98, RM99, Pot10, CP12, FPRC13].

In this paper we study the circuit complexity of *approximating* monotone functions, focusing on DNF formulas, one of the simplest and most basic types of circuits. We say that a DNF ε-approximates a function $f : \{0,1\}^n \to \{0,1\}$ if the function g computed by the DNF satisfies $f(x) = g(x)$ on at least a $1-\varepsilon$ fraction of inputs x in $\{0,1\}^n$. Recent works [OW07, BT13] have highlighted interesting qualitative and quantitative differences in the landscape of DNF complexity when the formula is only required to approximate f rather than compute it exactly, and while the DNF complexity of exact computation is fairly well-understood, these papers have also pointed to significant gaps in our understanding of seemingly basic questions regarding the DNF complexity of approximate computation.

We continue this study and explore two main directions. In the first direction we seek a non-trivial upper bound on the DNF complexity of approximating an arbitrary monotone function to high accuracy, in the spirit of the positive results of [BT13]. In the second direction, in the spirit of Razborov's theorem [Raz85] we seek a separation between the relative powers of monotone and non-monotone DNF that approximate monotone functions. As we describe below, our results further illustrate how different DNF complexity can be in the settings of exact versus approximate computation.

Universal bounds on approximability Recent work of [BT13] established the first non-trivial universal upper bound on the DNF complexity of approximating an arbitrary Boolean function, achieving logarithmic savings over the worst-case cost of $\Omega(2^n)$ necessary for exact computation:

Theorem 1 of [BT13]. *Every Boolean function can be ε-approximated by a DNF of size $O_\varepsilon(2^n/\log n)$.*

We begin with the simple observation that this result does not say anything meaningful about the approximation of monotone functions. Since the minterms of a monotone function form a Sperner family, Sperner's classical theorem readily translates into an upper bound on the DNF complexity of *exactly* computing monotone functions that is polynomially stronger:

Fact 1.1. *Every monotone function can be computed exactly by a DNF of size $\binom{n}{\lceil n/2 \rceil} = \Theta(2^n/\sqrt{n})$.*

This bound is exactly tight by considering the n-variable majority function, and in fact an elementary combinatorial argument establishes that a $1-o_n(1)$ fraction of monotone functions do actually require DNFs of size $\Omega(2^n/\sqrt{n})$ to compute. Fact 1.1, taken together with the result of [BT13], raises a basic qualitative question: are there monotone functions that require DNFs of size $\Omega(2^n/\sqrt{n})$ to approximate, or can every monotone function be approximated by a DNF of size $o(2^n/\sqrt{n})$? Despite the vast literature on monotone functions and Sperner families, this question does not appear to have been explicitly studied before. We answer this question in the first half of the paper, constructing DNF approximators for arbitrary monotone functions that achieve exponential savings over the size necessary for exact computation. Our DNF approximators only make one-sided error, and our construction is based on a new structural decomposition of monotone functions.

Power of negations in approximating monotone functions In the second half of the paper we turn our attention to the role of *negations* in the DNF complexity of approximating monotone functions. Recall that a circuit is said to be monotone if it does not contain any NOT gates, and non-monotone otherwise. While every monotone function can be computed by a monotone circuit, there is a body of results showing the remarkable fact that for various circuit classes, the optimal circuit computing a monotone function must be non-monotone. The most prominent example is perhaps Razborov's celebrated lower bound:

Razborov's Theorem [Raz85]. *There is a polynomial-time computable monotone function that requires monotone circuits of quasi-polynomial size.*

This separation of monotone NP from monotone P/poly was subsequently improved from quasi-polynomial to exponential by E. Tardos [Tar88]. An analogue of Razborov's result in the setting of bounded-depth circuits was established by Okol'nishnikova, Ajtai, and Gurevich:

Okol'nishnikova–Ajtai–Gurevich Theorem [Oko82, AG87]. *There is a monotone function in* AC^0 *that is not in monotone* AC^0.

For the class of DNFs, however, it is well-known (and straightforward to verify) that the analogue these separations does not hold:

Quine's Theorem [Qui54]. *The optimal DNF, with respect to both size and width, computing a monotone function is monotone as well.*

In the second half of this paper we investigate the question: does Quine's theorem hold for *approximation* by DNFs? In other words, is the optimal DNF approximator for a monotone function monotone as well, or do negations buy us power in the setting of approximation? We show that the answer is the latter, giving separations with respect to both DNF size and width. Our results, taken in contrast with Quine's theorem, highlight an interesting qualitative difference between the DNF complexity of exact and approximate computation. More broadly, we believe that the role of negations in the circuit complexity of approximating monotone functions is a topic of intrinsic interest, and we view our separations as the first steps in its systematic study.

1.1 Our results

Universal bounds on approximability Our first result is the construction of DNF approximators for arbitrary monotone Boolean functions that achieve one-sided error:

Theorem 1. *Every monotone function* f *can be* ε*-approximated by a monotone function* g *of DNF size* $2^{n-\Omega_\varepsilon(\sqrt{n})}$, *satisfying* $g(x) \le f(x)$ *for all* $x \in \{0,1\}^n$.

Prior to our work the only known universal upper bound, even for approximators incurring two-sided error, was the trivial one of $\binom{n}{\lceil n/2 \rceil} = \Theta(2^n/\sqrt{n})$, the size sufficient for exact computation. A standard information-theoretic argument (see [BT13] for proof) shows that any ε-approximator for a random Boolean function has DNF size $\Omega_\varepsilon(2^n/n)$; Theorem 1 therefore shows that the structure of monotonicity can be leveraged to obtain DNF approximators with complexity exponentially smaller than that required for almost all other functions. Our construction relies on a new structural fact about monotone functions which we believe may be of independent interest:

Lemma 1.2. *Let* f *be a monotone function and* $\varepsilon > 0$. *There is a function* $g = g_1 \vee \cdots \vee g_t$ *that* ε*-approximates* f, *where* $t = O_\varepsilon(1)$ *and each* g_i *is a monotone DNF with terms of width exactly* k_i *and size at least* $(\varepsilon/2)\binom{n}{k_i}$. *Furthermore,* $g(x) \le f(x)$ *for all* $x \in \{0,1\}^n$.

Since $g(x) \leq f(x)$ for all $x \in \{0,1\}^n$, we say that g is a lower ε-approximator for f. We prove Lemma 1.2 in Section 2, and with this structural fact in hand, the task of constructing lower approximators for an arbitrary monotone function reduces to that of constructing lower approximators for the g_i's. Since g comprises only a constant number of these g_i's, taking a naive union bound incurs no more than a constant factor in terms of error and DNF size of the overall approximator. Our lower approximators for the g_i's, presented in Section 3, are obtained via a randomized algorithm that constructs an approximating DNF. We complement our positive result with a lower bound showing that Theorem 1 is essentially optimal:

Theorem 2. *Let g be a $\frac{1}{10}$-approximator for the majority function MAJ_n satisfying $g(x) \leq \mathsf{MAJ}_n(x)$ for all $x \in \{0,1\}^n$. Then g has DNF size $2^{n-O(\sqrt{n}\log n)}$.*

Power of negations in approximating monotone functions The proof of Quine's classical theorem mentioned in the introduction is simple: given a DNF g that computes a monotone function f, if g contains a term T with a negated variable $\bar{x}_i$, it is easy to check that g still computes the same monotone function f if $\bar{x}_i$ is removed from T. Therefore, by removing all occurrences of negated variables in g, we obtain a monotone DNF h computing the same function f, where the size and width of h are at most those of g.

It is natural to suspect that the same would be true for DNF approximators, that the optimal DNF approximator for a monotone function is always monotone as well; indeed, we note that the universal DNF approximators we construct in Theorem 1 are in fact monotone. To be precise, we consider the following question:

Question 1. *Let f be a monotone function that is ε-approximated by a DNF g of size s (resp. width w). Can f be ε-approximated by a* monotone *DNF h of size s (resp. width w)?*

The simple proof of Quine's theorem does not extend to answer this question in the affirmative. In fact, for all three natural ways of "locally monotonizing" the DNF approximator g — removing $\bar{x}_i$ in T (as is done in the proof of Quine's theorem); replacing $\bar{x}_i$ with x_i in T; and removing T from f entirely — it is possible to construct examples showing that these operations increase the distance of g from f (i.e. worsens the quality of approximation).

In the second half of the paper we resolve Question 1 by showing, perhaps somewhat surprisingly, that the answer is "No" for both complexity measures of DNF size and DNF width. In Section 4 we prove the following two theorems:

Theorem 3 (Separation for DNF size)**.** *For all sufficiently large n, there exists an n-variable monotone function f and a value $\varepsilon = \varepsilon(n) > 0$ such that f can be ε-approximated by a DNF of size $O(n)$, but any monotone function that ε-approximates f has DNF size $\Omega(n^2)$.*

Theorem 4 (Separation for DNF width)**.** *For all sufficiently large n, and for all $k = o(n)$, there exists an n-variable monotone function f and a value $\varepsilon = \varepsilon(n) > 0$ such that f can be ε-approximated by a DNF of width $k + \log k$, but any monotone function that ε-approximates f has DNF width at least $2k - 1 - o_n(1)$.*

We view these separations as the first steps in quantifying just how powerful negations can be in the approximation of monotone functions, a question that does not appear to have been explicitly studied before (despite a significant body of results on the power of negations in the computation of monotone functions, as discussed above). We conclude the paper by listing a few interesting questions for future work in this direction.

1.2 Previous work

The explicit study of the DNF complexity of approximating Boolean functions was initiated by O'Donnell and Wimmer [OW07]. They showed that DNF size $2^{O_\varepsilon(\sqrt{n})}$ is both necessary and sufficient for ε-approximating the n-variable majority function, and constructed an explicit n-variable monotone function for which any 0.01-approximating DNF must have size $2^{\Omega(n/\log n)}$. As mentioned above, Blais and Tan [BT13] gave universal upper bounds on DNF size for approximating arbitrary Boolean functions, but [BT13] does not consider monotone functions.

We also note that the earlier work of Bshouty and Tamon [BT96], which established Fourier concentration bounds for monotone Boolean functions, implies that every n-variable monotone function is ε-close to a depth-2 circuit of size $2^{O(\sqrt{n}\log(n)/\varepsilon)}$ in which the bottom-level gates are parity gates and the top gate is a threshold gate (with unbounded weights). Recall that while threshold-of-parity circuits can simulate DNF formulas with only a polynomial size increase [Jac97, KP97], the converse is not true (indeed, even a single parity gate requires exponential DNF size). Thus the [BT96] results do not imply the existence of nontrivial DNF approximators for monotone functions.

1.3 Preliminaries

Throughout this paper all probabilities and expectations are with respect to the uniform distribution unless otherwise stated; we will use boldface (*e.g.* $\boldsymbol{x}$ and $\mathbf{X}$) to denote random variables. For strings $x, y \in \{0,1\}^n$ we write $\|x\|$ to denote the Hamming weight $\#\{i \in [n] \colon x_i = 1\}$ of x, and $x \preceq y$ if $x_i \le y_i$ for all $i \in [n]$, and $x \prec y$ if $x \preceq y$ and $x \neq y$. For $0 \le k \le n$, we write $\mathrm{Vol}(n,k) := \sum_{i=0}^{k} \binom{n}{i}$ to denote the volume of the n-dimensional Hamming ball of radius k.

A monotone Boolean function $f : \{0,1\}^n \to \{0,1\}$ is one that satisfies $f(x) \le f(y)$ whenever $x \preceq y$. A DNF formula is the logical OR of logical ANDs, where we refer to each AND as a *term*. The *size* of a DNF is the number of terms it contains, and the *width* of a DNF is the maximum width of any term. For a term T, we write $|T|$ to denote the width of T, the number of literals occurring in it. For any $x \in \{0,1\}^n$, we write T_x to denote the monotone conjunction that accepts all $y \in \{0,1\}^n$ such that $y \succeq x$. That is, $T_x(y) = 1$ iff $y_i = 1$ for all $i \in [n]$ such that $x_i = 1$. We say that x *defines a minterm in a monotone function* f if T_x is a minterm in the canonical DNF computing f, and we write $\mathrm{minterm}(x, f)$ to denote the indicator for this event.

Definition 5 (canonical DNF)**.** Let f be a monotone Boolean function. The *canonical DNF* for f is the unique monotone DNF whose terms correspond precisely to the minterms of f.

Definition 6 (ε-approximator)**.** Let $f, g : \{0,1\}^n \to \{0,1\}$ be Boolean functions and $\varepsilon \in [0,1]$. We say that g is an ε-approximator for f, or that f and g are ε-close, if $\mathbf{Pr}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})] \le \varepsilon$. We say that g is a lower approximator for f if $g(x) \le f(x)$ for all $x \in \{0,1\}^n$, and an upper approximator for f if $f(x) \le g(x)$ for all $x \in \{0,1\}^n$.

Definition 7 (density)**.** Let $f : \{0,1\}^n \to \{0,1\}$ and $k \in \{0,1,\ldots,n\}$. The density of f at level k, denoted $\mu_k(f)$, is defined to be

$$\mu_k(f) = \mathop{\mathbf{Pr}}_{\|x\|=k}[f(\boldsymbol{x}) = 1] = \#\{x \in \{0,1\}^n \colon \|x\| = k \text{ and } f(x) = 1\} \cdot \binom{n}{k}^{-1}.$$

Fact 1.3. *Let f be a monotone function. Then $\mu_k(f) \ge \mu_{k-1}(f)$ for all $k \in [n]$.*

We recall two basic facts from probability theory.

Fact 1.4 (Chernoff bound)**.** *Let* $\mathbf{X} \sim \mathrm{Binomial}(n, 1/2)$*. Then for any* $0 \le t \le \sqrt{n}$*, we have* $\mathbf{Pr}\left[\mathbf{X} \ge \frac{n}{2} + t\frac{\sqrt{n}}{2}\right] \le e^{-t^2/2}$ *and* $\mathbf{Pr}\left[\mathbf{X} \le \frac{n}{2} - t\frac{\sqrt{n}}{2}\right] \le e^{-t^2/2}$.

Fact 1.5 (anti-concentration of the Binomial)**.** *For every* $\varepsilon \ge 1/\sqrt{n}$ *and interval* $I \subseteq [0, n]$ *of width at most* $\varepsilon\sqrt{n}$*, we have* $\mathbf{Pr}_{\boldsymbol{x}\in\{0,1\}^n}[\|\boldsymbol{x}\| \in I] \le 2\varepsilon$.

2 A Regularity Lemma for Monotone DNFs

We begin with a new structural fact about monotone functions, which states that every monotone DNF f is lower approximated by the disjunction g of a constant number of monotone DNFs that are "dense" and "regular." Here a "regular" DNF is one in which all terms have the same width k, and a "dense" regular DNF is one that contains a constant fraction of the $\binom{n}{k}$ many possible terms of width k. This structural decomposition is useful as it reduces the task of (lower) approximating an arbitrary monotone DNF f to that of (lower) approximating a dense regular one. Sine g is the disjunction of only a constant number of dense regular DNFs, taking a naive union bound incurs only a constant factor in terms of error and DNF size of the overall approximator.

Definition 8 (regular and dense DNFs)**.** Let $k \in [n]$. We say that a monotone DNF f is k-regular if all its terms have width exactly k, and regular if it is k-regular for some k. Additionally, we say that f is (ε, k)-regular if it is a k-regular DNF with at least $\varepsilon\binom{n}{k}$ many terms.

Our structural result says that every monotone function is lower ε-approximated by the disjunction of $O_\varepsilon(1)$ many $(\varepsilon/2, k_i)$-regular DNFs, where each $k_i = (n/2) \pm O(\sqrt{n})$. More precisely:

Lemma 1.2. *For any* $\varepsilon > 0$*,every monotone function* f *is* ε*-close to the disjunction* g *of monotone DNFs,* $g(x) = g_1(x) \vee \cdots \vee g_t(x)$*, where*

1. $t \le 2/\varepsilon$*,*
2. *each* g_i *is* k_i*-regular for some* $k_i \in \left[(n/2) - \sqrt{n\ln(4/\varepsilon)/2}, (n/2) + \sqrt{n\ln(4/\varepsilon)/2}\right]$*,*
3. *the DNF size of* g_i *is at least* $(\varepsilon/2)\binom{n}{k_i}$ *(i.e.,* $\mu_{k_i}(g_i) \ge \varepsilon/2$*).*
4. $g(x) \le f(x)$ *for all* $x \in \{0,1\}^n$*.*

Proof. Fix $\ell := \sqrt{n\ln(4/\varepsilon)/2}$. For each $k = \{0, 1, \ldots, n\}$ we define

$$f_k(x) := \bigvee\{T_x \colon \|x\| = k \text{ and } \mathrm{minterm}(x, f) = \mathsf{true}\},$$

where we recall that T_x is the monotone term that accepts all y such that $y \succeq x$. By the Chernoff bound $\mathbf{Pr}_{\boldsymbol{x}\in\{0,1\}^n}[||\|\boldsymbol{x}\| - n/2| \ge \ell] \le \varepsilon/2$, and so f is $(\varepsilon/2)$-close to

$$f^*(x) = f_{(n/2)-\ell}(x) \vee \cdots \vee f_{(n/2)+\ell}(x).$$

Furthermore, $f^*(x)$ is an lower approximator for f. By the triangle inequality, it suffices to prove that f^* is $(\varepsilon/2)$-close to g satisfying the four claims in the lemma statement.

Consider the algorithm described in Figure 1, and let g be the resulting function when the algorithm terminates. First, since the algorithm only sets $f^*(x) = 0$ for inputs x that define a minterm in f^*, we regularize(f^*):

1. for $k = (n/2) - \ell, \ldots, (n/2) + \ell$:
2. if $\mathbf{Pr}_{\|\boldsymbol{x}\|=k}[\mathrm{minterm}(\boldsymbol{x}, f^*)] < \varepsilon/2$
3. set $f^*(x) = 0$ for all x s.t. $\|x\| = k$ and $\mathrm{minterm}(x, f^*) =$ true.

Figure 1: The regularize algorithm

have that g is a monotone lower approximator for f^*. Second, since the algorithm corrupts less than an $(\varepsilon/2)$-fraction of any layer, g is $(\varepsilon/2)$-close to f^*.

We will argue that g is the disjunction of regular monotone DNFs satisfying the first three claims in the lemma. The algorithm ensures that for every $k \in [(n/2) - \ell, (n/2) + \ell]$, the fraction of inputs at layer k that define a minterm in g is either 0 at least $\varepsilon/2$ — if this fraction is in the range $[0, \varepsilon/2)$, the predicate in Line 2 of the algorithm is satisfied and the fraction is set to 0 by Line 3. Furthermore, since all the minterms of f^* have weight in the range $[(n/2) - \ell, (n/2) + \ell]$, the same is true for g and so

$$\mathop{\mathbf{Pr}}_{\|\boldsymbol{x}\|=k}[\mathrm{minterm}(\boldsymbol{x}, g)] = \begin{cases} 0 & \text{if } k \notin [(n/2) - \ell, (n/2) + \ell] \\ 0 \text{ or } \geq \varepsilon/2 & \text{if } k \in [(n/2) - \ell, (n/2) + \ell]. \end{cases}$$

Each layer $k_i \in [(n/2) - \ell, (n/2) + \ell]$ such that this probability is at least $\varepsilon/2$ naturally defines a k_i-regular monotone DNF g_i satisfying the second and third claims of the lemma: g_i is simply the DNF

$$g_i(x) := \bigvee \{T_x \colon \|x\| = k_i \text{ and } \mathrm{minterm}(x, g)\}.$$

Therefore $g(x) = g_1(x) \vee \cdots \vee g_t(x)$ where each g_i is k_i-regular and $\mu_{k_i}(g_i) \geq \varepsilon/2$, and so it remains to justify the first claim of the lemma, that $t \leq 2/\varepsilon$. We assume without loss of generality that $k_1 < k_2 < \ldots < k_t$, and claim that

$$\mu_{k_i}(g_1 \vee \cdots \vee g_i) \geq i \cdot \frac{\varepsilon}{2} \quad \text{for all } i \in [t]. \tag{1}$$

Note that this implies the first claim of the lemma since $\mu_{k_t}(g_1 \vee \ldots \vee g_t) \leq 1$ holds trivially, and so $t \leq 2/\varepsilon$. We prove (1) by induction on i, noting that the base case holds since $\mu_{k_i}(g_i) \geq \varepsilon/2$ for all i by construction, and in particular when $i = 1$. Suppose $\mu_{k_i}(g_1 \vee \ldots \vee g_i) \geq i \cdot \varepsilon/2$ for some $i < t$. By Fact 1.3, we have

$$\mu_{k_{i+1}}(g_1 \vee \cdots \vee g_i) \geq \mu_{k_i}(g_1 \vee \cdots \vee g_i) \geq i \cdot \frac{\varepsilon}{2}.$$

Since the terms of g_{i+1} are the width-(k_{i+1}) minterms of g, the sets

$$\begin{aligned} A &= \{x \in \{0,1\}^n \colon \|x\| = k_{i+1} \text{ and } g_1(x) \vee \cdots \vee g_i(x) = 1\} \\ B &= \{x \in \{0,1\}^n \colon \|x\| = k_{i+1} \text{ and } g_{i+1}(x) = 1\} \\ &\equiv \{x \in \{0,1\}^n \colon \|x\| = k_{i+1} \text{ and } \mathrm{minterm}(x, g)\} \end{aligned}$$

are disjoint, and so

$$\mu_{k_{i+1}}(g_1 \vee \cdots \vee g_{i+1}) = \mu_{k_{i+1}}(g_1 \vee \ldots \vee g_i) + \mu_{k_{i+1}}(g_{i+1}) = \left(i \cdot \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2} = (i+1) \cdot \frac{\varepsilon}{2}.$$

This completes the proof. □

3 Lower Approximators for Regular DNFs

With Lemma 1.2 in hand it suffices to construct lower approximators for regular DNFs:

Proposition 3.1. *Let f be a regular monotone function. For every $\varepsilon > 0$ there exists a monotone DNF g of size $2^{n-\Omega(\varepsilon\sqrt{n}-\log(n))}$ that is a lower ε-approximator for f.*

Proof of Theorem 1 assuming Proposition 3.1. By Lemma 1.2 every monotone f has an upper $(\varepsilon/2)$-approximator $g(x) = g_1(x) \vee \cdots \vee g_t(x)$ where $t \leq 4/\varepsilon$ and each $g_i(x)$ is a regular monotone function. Next, by Proposition 3.1 each regular $g_i(x)$ has a lower $(\varepsilon/2t)$-approximator $h_i(x)$ of size $2^{n-\Omega((\varepsilon\sqrt{n}/t)-\log(n))}$. Finally, by the union bound and the triangle inequality, we conclude that $h(x) = h_1(x) \vee \cdots \vee h_t(x)$ is a lower ε-approximator for f of size at most $t \cdot 2^{n-\Omega((\varepsilon\sqrt{n}/t)-\log(n))} = 2^{n-\Omega_\varepsilon(\sqrt{n})}$. $\square$

Proof of Proposition 3.1. We may assume that $\varepsilon \geq (C\log n)/\sqrt{n}$ (for some constant $C > 0$ which we will specify below), since otherwise the claimed bound on monotone DNF size is trivial. Let f be a k-regular monotone function for some $k \in [n]$. The minterms of our monotone approximator g will be conjunctions of the form T_y where $y \in f^{-1}(1)$, which guarantees that g will be a lower approximator for f. Furthermore, since $\mathbf{Pr}_{\boldsymbol{x}\in\{0,1\}^n}\left[\|\boldsymbol{x}\| \geq (n/2) + \sqrt{n\ln(3/\varepsilon)/2}\right] \leq \frac{\varepsilon}{3}$, and $\mathbf{Pr}_{\boldsymbol{x}\in\{0,1\}^n}[\|\boldsymbol{x}\| \in [k, k+\varepsilon\sqrt{n}/6]] \leq \frac{\varepsilon}{3}$, by the Chernoff bound and Fact 1.5 respectively, it suffices to ensure that the monotone DNF g we construct additionally satisfies:

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\in A}[g(\boldsymbol{x}) \neq f(\boldsymbol{x})] \leq \frac{\varepsilon}{3}, \quad A := \left\{x \in \{0,1\}^n \colon \|x\| \in \left[k+\varepsilon\sqrt{n}/6, (n/2) + \sqrt{n\ln(3/\varepsilon)/2}\right]\right\}. \tag{2}$$

Note that if $k + \varepsilon\sqrt{n}/6 > (n/2) + \sqrt{n\ln(3/\varepsilon)/2}$ (*i.e.* the interval in the definition of A is empty) then f is $(2\varepsilon/3)$-close to the constant 0 function and the proposition is trivially true.

For every $\ell \in \{0, 1, \ldots, n-k\}$, we write S_ℓ to denote the 1-inputs of f with Hamming weight exactly $k+\ell$; that is, $S_\ell := \{x \in \{0,1\}^n \colon f(x) = 1 \text{ and } \|x\| = k+\ell\}$. The remainder of this proof will be devoted to showing that for each $\ell \geq \varepsilon\sqrt{n}/6$, there exists a monotone DNF g_ℓ satisfying:

i. The minterms of g_ℓ are of the form T_y for some $y \in S_{\ell/2}$ (and hence $g_\ell \leq f$),

ii. DNF-size$[g_\ell] = O(2^{n-\ell/2}) \leq 2^{n-\Omega(\varepsilon\sqrt{n})}$,

iii. $\mathbf{Pr}_{\boldsymbol{x}\in S_\ell}[g_\ell(\boldsymbol{x}) = 0] \leq \varepsilon/3$.

Indeed, taking g to be the disjunction of all g_ℓ where $k + \ell \in \left[k+\varepsilon\sqrt{n}/3, (n/2) + \sqrt{n\ln(3/\varepsilon)/2}\right]$, we obtain a monotone DNF of size at most $n \cdot 2^{n-\Omega(\varepsilon\sqrt{n})} \leq 2^{n-\Omega(\varepsilon\sqrt{n}-\log(n))}$ satisfying (2), which completes the proof.

Consider a random monotone DNF $\boldsymbol{g}_\ell$ sampled according to the following distribution $\mathcal{D}$: for each $y \in S_{\ell/2}$, independently include T_y as a minterm of $\boldsymbol{g}_\ell$ with probability $p := 2^{-\ell/2}$. By definition, every DNF in the support of this distribution satisfies (i), and so it remains to argue that with positive probability, both (ii) and (iii) are satisfied as well. For (ii), we observe that $\mathbf{E}_{\mathcal{D}}[\text{DNF-size}[\boldsymbol{g}_\ell]] = p \cdot |S_\ell| < p \cdot 2^n = 2^{n-\ell/2}$, and so by Markov's inequality,

$$\mathop{\mathbf{Pr}}_{\mathcal{D}}\left[\text{DNF-size}[\boldsymbol{g}_\ell] \leq 3 \cdot 2^{n-\ell/2}\right] \geq \frac{2}{3}. \tag{3}$$

For (iii), consider any fixed $x \in S_\ell$. Since f is k-regular, there must exist some $z \in S_0$ such that $z \prec x$, and therefore $\binom{\ell}{\ell/2} = \Theta(2^\ell/\sqrt{\ell})$ many $y \in S_{\ell/2}$ such that $z \prec y \prec x$. By the definition of $\mathcal{D}$, for each such y the term T_y is independently included as a minterm of $\boldsymbol{g}_\ell$ with probability $p = 2^{-\ell/2}$, and so

$$\Pr_{\mathcal{D}}[\boldsymbol{g}_\ell(x) = 0] \leq (1-p)^{\Theta(2^\ell/\sqrt{\ell})} = \exp\left(-\Omega(2^{\ell/2}/\sqrt{\ell})\right) < \exp\left(-\Omega(2^{\varepsilon\sqrt{n}/12})/\sqrt{n}\right) < \frac{\varepsilon}{9},$$

where we have used $\varepsilon \geq (C \log n)/\sqrt{n}$ for the final inequality. Therfore

$$\mathop{\mathbf{E}}_{\mathcal{D}}\left[\Pr_{\boldsymbol{x}\in S_\ell}[\boldsymbol{g}_\ell(\boldsymbol{x}) = 0]\right] \leq \frac{\varepsilon}{9}, \quad \text{and} \quad \Pr_{\mathcal{D}}\left[\Pr_{\boldsymbol{x}\in S_\ell}[\boldsymbol{g}_\ell(\boldsymbol{x}) = 0] \leq \frac{\varepsilon}{3}\right] \geq \frac{2}{3}. \tag{4}$$

Applying a union bound to the failure probabilities of (3) and (4), we conclude that there is indeed a positive probability that $\boldsymbol{g}_\ell \sim \mathcal{D}$ satisfies all three properties (i), (ii), and (iii), and this completes the proof. □

3.1 Near-Matching Lower Bound

In this section we show that our upper bound in Theorem 1 is essentially tight.

Theorem 2. *Let $\varepsilon \leq \frac{1}{10}$ and g be an s-term DNF that is a lower ε-approximator for the majority function MAJ_n. Then $s \geq 2^{n-O(\sqrt{n}\log n)}$.*

Proof. First we claim that we may assume without loss of generality that g is an $\lceil n/2 \rceil$-regular monotone function. To see this, fix a DNF representation of g and consider any term

$$T(x) = \left(\bigwedge_{i\in S^+} x_i\right) \wedge \left(\bigwedge_{j\in S^-} \overline{x}_j\right), \quad S^+, S^- \subseteq [n]$$

in the DNF. Note that $|S^+| \geq \lceil n/2 \rceil$, since otherwise $g(y) = 1$ and $\mathsf{MAJ}_n(y) = 0$ on the input y where $y_i = 1$ iff $i \in S^+$ (of Hamming weight $\|y\| = |S^+| < \lceil n/2 \rceil$), contradicting our assumption that $g(x) \leq \mathsf{MAJ}_n(x)$ for all $x \in \{0,1\}^n$. Replacing $T(x)$ in g by $T'(x) = \bigwedge_{i\in S} x_i$, where S is an arbitrary subset of S^+ of cardinality exactly $\lceil n/2 \rceil$, we get a function g^* satisfying $g^{-1}(1) \subseteq (g^*)^{-1}(1) \subseteq \mathsf{MAJ}_n^{-1}(1)$. Performing this replacement for every term in g, we obtain an $\lceil n/2 \rceil$-regular monotone DNF of size at most s that lower ε-approximates MAJ_n.

Next we claim that if g is an $\lceil n/2 \rceil$-regular monotone DNF that ε-approximates MAJ_n, then $\mu_\ell(g) \geq 1 - 4\varepsilon$ for $\ell := \lceil n/2 \rceil + \sqrt{n \ln 2}$. If $\mu_\ell(g) < 1 - 4\varepsilon$, then by Fact 1.3 by have $\mu_k(g) \leq \mu_\ell(g) < 1 - 4\varepsilon$ for all $k \leq \ell$, and since

$$\Pr_{\boldsymbol{x}\in\{0,1\}^n}\left[\|\boldsymbol{x}\| \in \left[\lceil n/2 \rceil, \ell\right]\right] \geq \frac{1}{4},$$

by the Chernoff bound, we get $\mathbf{Pr}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})] > \varepsilon$, a contradiction.

Having established that $\mu_\ell(g) \geq 1 - 4\varepsilon$, we complete the proof with a simple counting argument. For every x of weight $\lceil n/2 \rceil$, we have $|\{y \in \{0,1\}^n \colon \|y\| = \ell \text{ and } y \succ x\}| = \binom{\lceil n/2 \rceil}{\sqrt{n \ln 2}}$. Since $\mu_\ell(g) \geq 1 - 4\varepsilon = \Omega(1)$ and there are $\binom{n}{\ell}$ strings of Hamming weight ℓ, we conclude that the number of terms in the canonical DNF for g is at least

$$\mu_\ell(g)\binom{n}{\ell} \cdot \binom{\lceil n/2 \rceil}{\sqrt{n \ln 2}}^{-1} = \Omega(1) \cdot \binom{n}{\lceil n/2 \rceil + \sqrt{n \ln 2}} \cdot \binom{\lceil n/2 \rceil}{\sqrt{n \ln 2}}^{-1} = 2^{n-O(\sqrt{n}\log n)}$$

as claimed. □

4 Power of Negations in Approximating Monotone Functions

In this section we present our constructions showing that non-monotone DNFs can asympototically outperform monotone ones in the approximation of monotone functions. We present our separation for DNF size in Section 4.1, followed by our separation for DNF width in Section 4.2.

4.1 Separation for DNF size

Theorem 3. *Let $f : \{0,1\}^n \times \{0,1\}^{5n} \to \{0,1\}$ be the monotone function:*

$$f(x,y) = (x_1 \vee \ldots \vee x_n) \wedge (y_1 \vee \ldots \vee y_{5n}) = \bigvee_{\substack{i\in[n]\\ j\in[5n]}} (x_i \wedge y_j),$$

and $\varepsilon = (2^{n-1}-1)\cdot 2^{-6n}$. There exists a DNF of size $6n-1$ that ε-approximates f, but any monotone function that ε-approximates f has DNF size at least n^2.

Proof. Consider the function $g = g(x,y)$ defined as

$$g = (x_1 \wedge (y_1 \vee \ldots \vee y_{5n})) \vee (\overline{x}_1 \wedge (x_2 \vee \ldots \vee x_n)) = \left(\bigvee_{j\in[5n]} (x_1 \wedge y_j)\right) \vee \left(\bigvee_{2\le i\le n} (\overline{x}_1 \wedge x_i)\right). \quad (5)$$

This is a non-monotone DNF with $6n-1$ terms that ε-approximates f, since $g(x,y)$ differs from $f(x,y)$ exactly on the $2^{n-1}-1$ inputs satisfying $x_1 = 0$, $y = \mathbf{0}$, and $x_2 \vee \ldots \vee x_n = 1$.

The rest of the proof will be devoted to showing that any monotone function that ε-approximates f has to have more than n^2 terms, asymptotically as many as the canonical DNF for f which has $5n^2$ terms. We will prove the contrapositive: any monotone DNF h with at most n^2 terms differs from f on strictly more than an ε-fraction of inputs.

We group the terms of h into three types: terms with only x-variables, which we call "pure-x"; terms with only y-variables, which we call "pure-y"; and terms with both x- and y-variables, which we call "mixed". We first observe that we may assume that all mixed terms have width exactly two, comprising one x-variable and one y-variable. Indeed, replacing a mixed term $\left(\bigwedge_{i\in S_1} x_i\right) \wedge \left(\bigwedge_{j\in S_2} y_j\right)$, $S_1 \subseteq [n]$ and $S_2 \subseteq [5n]$, in h with $(x_i \wedge y_j)$ for any $i \in S_1$ and $j \in S_2$ yields a DNF h' such that $h'(x,y) \neq h(x,y)$ only on inputs (x,y) such that $h(x,y) = 0$ and $f(x,y) = 1$).

Furthermore, we claim that we may assume all pure-y terms have width greater than $2n$. Indeed, if h contains a term $T(y) = \bigwedge_{i\in S} y_i$ for some $S \subseteq [5n]$ where $|S| \le 2n$, then $f(x,y) = 0$ and $h(x,y) = 1$ on at least $2^{3n} > \varepsilon \cdot 2^{6n}$ inputs (x,y) satisfying $x = \mathbf{0}$ and $T(y) = 1$.

We proceed by considering two cases, depending on the number of x_i's that occur as a singleton term in h. First suppose at least half of the x_i's occur as a singleton term in h, so there is some $S \subseteq [n]$ where $|S| \ge n/2$ such that if $\mathsf{OR}_S(x) = \bigvee_{i\in S} x_i = 1$ then $h(x,y) = 1$. In this case $f(x,y) = 0$ and $h(x,y) = 1$ on at least $2^n - 2^{n/2} > \varepsilon \cdot 2^{6n}$ inputs satisfying $y = \mathbf{0}$ and $\mathsf{OR}_S(x) = 1$. Finally, suppose less than half of the x_i's occur as singleton terms in h. By our first assumption that all mixed terms have width two (in particular, no mixed term contains more than one x-variable), there must be an x_i that does not occur as a singleton term and participates in at most $2n$ mixed terms (since otherwise h would have more than n^2

terms); without loss of generality suppose x_1 is one such variable. Let $S \subseteq [5n]$ be the set of all $j \in [5n]$ such that $(x_1 \wedge y_j)$ is a mixed term in h, and consider the set of inputs

$$E = \{(x, y)\colon x_1 = 1, x_i = 0 \text{ for all } i \geq 2, \text{ and } y_j = 0 \text{ for all } j \in S, \text{ and } \|y\| = (3n)/2\}.$$

Note that $f(x, y) = 1$ for all $(x, y) \in E$, and we claim that $h(x, y) = 0$ on these inputs. To see this, consider the restriction h^* of h obtained by setting $x_1 \leftarrow 1$, $x_i \leftarrow 0$ for all $i \geq 2$, and $y_j \leftarrow 0$ for all $j \in S$. Since x_1 does not occur as a singleton term in h, this partial assignment does not satisfy any terms and the canonical DNF for h^* comprises only of pure-y terms. Since the pure-y terms of h have width greater than $2n$ (by our second assumption), the same is true for h^* and so h^* cannot be satisfied by any assignment of weight $(3n)/2$; hence $h(x, y) = h^*(y) = 0$ for all $(x, y) \in E$. Lastly, we check that $|E| \geq \binom{3n}{(3n)/2} = \Theta\big(2^{3n}/\sqrt{3n}\big) > \varepsilon \cdot 2^{6n}$ and this completes the proof. □

Remark 9. We note that the non-monotone approximator g in (5) is actually computed by a $O(n)$-size decision tree. Recall that every size-s decision tree is a size-s DNF, but not vice versa: there are polynomial-size DNFs that require exponential-size decision trees. Therefore the proof of Theorem 3 in fact establishes a stronger statement: f is a monotone function that can be ε-approximated by a $O(n)$-size decision tree, and yet any monotone function that ε-approximates f has DNF size $\Omega(n^2)$.

4.2 Separation for DNF width

We will need a simple combinatorial lemma concerning shadows in the hypercube.

Lemma 4.1. *Let $k \in [n]$ and $\delta \in (0, 1)$, and f be a monotone DNF of width δk. Then $\mu_{k-1}(f) \geq (1 - \delta) \cdot \mu_k(f)$.*

Proof. Let $\mathcal{C}$ be the collection of all pairs (y, x) satisfying $\|y\| = k - 1, \|x\| = k$, $y \prec x$, and $T(y) = T(x) = 1$ for some term T in f. We first note that every x such that $f(x) = 1$ and $\|x\| = k$ must satisfy some term T of width at most $\delta k < k$, and hence some term of length at most $k - 1$, so

$$|\{x \in \{0,1\}^n \colon \text{ there exists some } y \in \{0,1\}^n \text{ such that } (y, x) \in \mathcal{C}\}| = \mu_k(f) \cdot \binom{n}{k}.$$

Consider any $x^* \in f^{-1}(1)$ where $\|x^*\| = k$, and let T be a term in f such that $T(x) = 1$. Since $|T| \leq \delta k$, there are at least $(1 - \delta) \cdot k$ many y such that $(y, x^*) \in \mathcal{C}$. On the other hand, for any y^* where $\|y^*\| = k - 1$, there are exactly $n - k + 1$ many x such that $\|x\| = k$ and $x \succ y^*$. By double counting, we conclude that

$$\begin{aligned}\mu_{k-1}(f) \cdot \binom{n}{k-1} &= |\{y \in \{0,1\}^n \colon \text{there exists some } x \in \{0,1\}^n \text{ such that } (y, x) \in \mathcal{C}\}| \\ &\geq \frac{\mu_k(f) \cdot \binom{n}{k} \cdot (1 - \delta) \cdot k}{n - k + 1} = \mu_k(f) \cdot (1 - \delta) \cdot \binom{n}{k-1}.\end{aligned}$$

Equivalently, $\mu_{k-1}(f) \geq (1 - \delta) \cdot \mu_k(f)$. □

Theorem 4. *Let $f : \{0,1\}^n \times \{0,1\}^k \times \{0,1\}^\ell \to \{0,1\}$ be the monotone function:*

$$f(x, y, z) = \begin{cases} \|x\| \geq k \text{ and } y = 1^k & \text{if } \|z\| = 0 \\ \|x\| \geq k & \text{otherwise,} \end{cases}$$

and $\varepsilon = \mathrm{Vol}(n, k - 1) \cdot 2^{-(n+k+\ell)}$. There exists a DNF of width $k + \ell$ that ε-approximates f, but for all $k = o(n)$ any monotone function that ε-approximates f has width at least $(2 - 2^{-\ell}(1 + o_n(1))) \cdot k$. In particular, taking $\ell = \log k$ yields a gap of $k + \log k$ versus $2k - 1 - o_n(1)$.

Proof. Consider the function

$$g(x,y,z) = \begin{cases} y = 1^k & \text{if } \|z\| = 0 \\ \|x\| \geq k & \text{otherwise.} \end{cases}$$

This is a non-monotone function that is computed by a DNF of width $k + \ell$:

$$g(x,y,z) = \left(\bigwedge_{i\in[\ell]} \bar{z}_i \wedge \bigwedge_{j\in[k]} y_j \right) \vee \bigvee_{\substack{i\in[\ell] \\ S\subseteq[n]\colon |S|=k}} \left(z_i \wedge \bigwedge_{j\in S} x_j \right),$$

and we observe that g is indeed an ε-approximator for f since f and g differ on the $\mathrm{Vol}(n, k-1)$ inputs (x,y,z) where $z = \mathbf{0}$, $y = \mathbf{1}$, and $\|x\| \leq k-1$.

The rest of this proof will be devoted to showing that for all $k = o(n)$, any monotone DNF h that ε-approximates f has width at least $(2 - 2^{-\ell}(1 + o_n(1))) \cdot k$. Consider the monotone DNF h^* obtained by restricting $z_i \leftarrow 0$ for all $i \in [\ell]$ in h. We claim that *every* term in h^* has to contain *all* of $y_1, \dots, y_k$. Suppose not, and suppose without loss of generality that there exists a term T in h^* that does not contain y_1. If $|T| \geq 2k$ the overall claimed lower bound on $\mathrm{width}(h)$ is true; otherwise h errs on at least

$$2^{n+(k-1)-|T|} \geq 2^{n+(k-1)-2k} = 2^{\Omega(n)} \gg \mathrm{Vol}(n, k-1)$$

many inputs (x,y,z) where $z = \mathbf{0}, y_1 = 0$ and $T(x,y,z) = 1$, since $h(x,y,z) = h^*(x,y,z) = 1$ and $f(x,y,z) = 0$ on these inputs.

Let $h^\dagger$ be h^* with $y_i \leftarrow 1$ for all $i \in [k]$. Since every term in h^* contains all of $y_1, \dots, y_k$, it follows that

$$\mathrm{width}(h) \geq \mathrm{width}(h^*) \geq k + \mathrm{width}(h^\dagger), \tag{6}$$

and so it suffices to prove that $\mathrm{width}(h^\dagger) \geq (1 - 2^{-\ell}(1 + o_n(1))) \cdot k$. First, since $f(x,y,z) = 1$ on the $\binom{n}{k}$ inputs satisfying $z = \mathbf{0}, y = \mathbf{1}$, and $\|x\| = k$, we have that

$$(1 - \mu_k(h^\dagger))\binom{n}{k} \leq \mathrm{Vol}(n, k-1). \tag{7}$$

Next, since $h(x, \mathbf{1}, \mathbf{0}) \leq h(x, \mathbf{1}, z)$ for all $z \in \{0,1\}^\ell$ by monotonicity, every error h incurs on an input $(x, \mathbf{1}, \mathbf{0})$ where $\|x\| = k-1$ implies an error on $(x, \mathbf{1}, z)$ for every $z \in \{0,1\}^\ell$, and so

$$2^\ell \cdot \mu_{k-1}(h^\dagger)\binom{n}{k-1} \leq \mathrm{Vol}(n, k-1). \tag{8}$$

Using our assumption that $k = o(n)$, the bound of (7) implies that $\mu_k(h^\dagger) \geq 1 - o_n(1)$, and (8) that $\mu_{k-1}(h^\dagger) \leq 2^{-\ell}(1 + o_n(1))$. Applying Lemma 4.1 to $h^\dagger$ we conclude that $\mathrm{width}(h^\dagger) \geq (1 - 2^{-\ell}(1 + o_n(1))) \cdot k$, which along with (6) completes the proof. $\square$

4.3 Upper bounds

Given the separations between monotone and non-monotone DNFs established in the previous subsections, it is natural to explore bounds in the other direction which show that the existence of (non-monotone) DNF approximators implies the existence of monotone DNF approximators of related size, width, and accuracy. We first recall a few standard definitions and useful facts from the analysis of Boolean functions:

Definition 10 (influence)**.** The *total influence* of a Boolean function f, denoted $\mathbf{Inf}[f]$, is defined to be

$$\mathbf{Inf}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f] \quad \text{where} \quad \mathbf{Inf}_i[f] = \mathop{\mathbf{Pr}}_{\boldsymbol{x}\in\{0,1\}^n}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{\oplus i})],$$

and $x^{\oplus i}$ denotes x with its i-th coordinate flipped.

Definition 11. A coordinate $i \in [n]$ is *relevant* in a Boolean function f if $\mathbf{Inf}_i[f] > 0$. For $k \in \{0, 1, \ldots, n\}$, we say that f is a *k-junta* if it has at most k relevant coordinates.

Friedgut's Junta Theorem [Fri98]. *For every $\delta > 0$, every Boolean function f is δ-close to a $2^{O(\mathbf{Inf}[f]/\delta)}$-junta.*

Amano's Influence Bound [Ama11]. *Let f be computed by a width-w DNF. Then $\mathbf{Inf}[f] \leq w$.*

Gopalan–Meka–Reingold Junta Bound [GMR13]. *Let f be computed by a width-w DNF. Then f is δ-close to a $(w \log(1/\delta))^{O(w)}$-junta.*

Folklore Junta Bound. *Let f be computed by a size-s DNF. Then f is δ-close to a $(s \log(s/\delta))$-junta.*

Perhaps the most common way to obtain a monotone function from a non-monotone one is via the combinatorial shifting operators introduced by Kleitman:

Definition 12 (combinatorial shifting)**.** For every $i \in [n]$, the i-th shifting operator κ_i acts on Boolean functions as follows:

$$(\kappa_i f)(x) = \begin{cases} f(x) & \text{if } f(x) = f(x^{\oplus i}) \\ x_i & \text{otherwise.} \end{cases}$$

It is straightforward to verify that $\mathrm{shift}(f) := \kappa_1\kappa_2\cdots\kappa_n f$ is a monotone function. We will use additional basic facts concerning the shifting operators. The first is that they can only improve approximation with respect to a monotone function, and the second is that they do not increase the number of relevant coordinates.

Fact 4.2. *Let f be a monotone function and suppose f is ε-close to g. Then for all coordinates $i \in [n]$*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\in\{0,1\}^n}[f(\boldsymbol{x}) \neq \mathrm{shift}(g)(\boldsymbol{x})] \leq \mathop{\mathbf{Pr}}_{\boldsymbol{x}\in\{0,1\}^n}[f(\boldsymbol{x}) \neq (\kappa_i g)(\boldsymbol{x})] \leq \varepsilon.$$

Fact 4.3. *For every Boolean function f and coordinate $i \in [n]$, the number of relevant coordinates in $\kappa_i f$ is at most that in f. Consequently, the number of relevant coordinates in $\mathrm{shift}(f)$ is at most that in f.*

With these facts in hand we are now ready to prove our upper bounds showing that the existence of (non-monotone) DNF approximators implies the existence of monotone DNF approximators of related size, width, and accuracy.

Theorem 13. *Let f be a monotone function and suppose f is ε-approximated by a width-w DNF. For every $\delta > 0$ there is a monotone DNF of width $\min\{2^{O(w/\delta)}, (w\log(1/\delta))^{O(w)}\}$ that $(\varepsilon + \delta)$-approximates f.*

Proof. Let f^* be the width-w DNF that ε-approximates f. Combining Amano's influence bound and Friedgut's junta theorem, we know that f^* is δ-close to a $2^{O(w/\delta)}$-junta g. Next, by Facts 4.2 and 4.3, along with the triangle inequality, we get that $\mathrm{shift}(g)$ is a monotone $2^{O(w/\delta)}$-junta that $(\varepsilon + \delta)$-approximates f. This yields the first bound of $2^{O(w/\delta)}$ since every monotone k-junta is trivially computed by a monotone DNF of width at most k. A similar argument, using the Gopalan–Meka–Reingold junta bound in place of Amano's influence bound and Friedgut's junta theorem, yields the incomparable second bound of $(w\log(1/\delta))^{O(w)}$. □

A similar argument using the folklore junta bound in place of Amano's influence bound and Friedgut's junta theorem establishes an analogous result for DNF size:

Theorem 14. *Let f be a monotone function and suppose f is ε-approximated by a size-s DNF. For every $\delta > 0$ there is a monotone DNF of size $2^{s \log(s/\delta)}$ that $(\varepsilon + \delta)$-approximates f.*

5 Conclusion

Having obtained near-matching upper and lower bounds on the size of universal lower approximators in this paper, the natural next step is to consider *upper* approximators and approximators incurring error on both sides. The task of constructing universal upper approximators appears to be qualitatively different from that of lower approximators, and we are not aware of any construction achieving size better than the trivial one of $O(2^n/\sqrt{n})$ sufficient for exact computation. For approximators incurring two-sided error, our universal lower approximators of size $2^{n-\Omega_\varepsilon(\sqrt{n})}$ represent the current best upper bound. The strongest known lower bound for two-sided approximators is the $2^{\Omega(n/\log n)}$ lower bound of [OW07]; it would be interesting to find out whether this or the current $2^{n-\Omega_\varepsilon(\sqrt{n})}$ upper bound is closer to the truth.

As for the power of negations in the approximation of monotone functions, we believe that our results in Section 4 suggest a number of interesting avenues for further exploration. We suspect that the separations we presented in Sections 4.1 and 4.2 can be improved, perhaps even to super-polynomial for DNF size and super-constant for DNF width, and likewise our upper bounds in Section 4.3. We remark that in addition to the complexity measures of DNF size and width, the quantitative difference between the accuracy of monotone versus general DNFs is also an aspect in which our separations can be strengthened. In other words, we may view our separations as instantiations of the following general template:

> *There exists a monotone function f and a value $\varepsilon = \varepsilon(n) > 0$ such that f can be ε-approximated by a DNF of size s (resp. width w), but any monotone function that $\varphi(\varepsilon)$-approximates f requires DNF size $\Psi(s)$ (resp. width $\Psi(w)$).*

In Theorems 3 and 4, φ is simply the identity function, but one can consider the possibility of stronger statements where $\varphi(\varepsilon) \gg \varepsilon$.

Beyond DNFs, one may ask quantitatively just how powerful negations can be in the approximation of monotone functions for many other classes of circuits. We conclude by restating an open problem, due to Kalai, on the possibility of strengthening the Okol'nishnikova–Ajtai–Gurevich theorem:

Open Problem 1 ([Kal10]). *Is there a monotone function in AC^0 that cannot be approximated by monotone AC^0?*

References

[AB87] N. Alon and R. Boppana. The monotone circuit complexity of Boolean functions. *Combinatorica*, 7:1–22, 1987. 1

[AG87] Miklós Ajtai and Yuri Grevich. Monotone versus positive. *Journal of the ACM*, 34(4):1004–1015, 1987. (document), 1, 1

[Ama11] Kazuyuki Amano. Tight bounds on the average sensitivity of k-CNF. *Theory of Computing*, 7(1):45–48, 2011. 4.3

[BT96] Nader Bshouty and Christino Tamon. On the Fourier spectrum of monotone functions. *Journal of the ACM*, 43(4):747–770, 1996. 1.2

[BT13] Eric Blais and Li-Yang Tan. Approximating Boolean functions with depth-2 circuits. In *Proceedings of the 28th Annual IEEE Conference on Computational Complexity*, pages 74–85, 2013. 1, 1, 1, 1.1, 1.2

[CP12] Siu Man Chan and Aaron Potechin. Tight bounds for monotone switching networks via fourier analysis. In *Symposium on Theory of Computing (STOC)*, pages 495–504, 2012. 1

[FPRC13] Yuval Filmus, Toniann Pitassi, Robert Robere, and Stephen A. Cook. Average case lower bounds for monotone switching networks. In *Symposium on Foundations of Computer Science (FOCS)*, 2013. 1

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–36, 1998. 4.3

[GH98] Mikael Goldmann and Johan Håstad. Monotone circuits for connectivity have depth $(\log n)^{2-o(1)}$. *SIAM J. Comput.*, 27(5):1283–1294, 1998. 1

[GMR13] Parikshit Gopalan, Raghu Meka, and Omer Reingold. Dnf sparsification and a faster deterministic counting algorithm. *Computational Complexity*, 22(2):275–310, 2013. 4.3

[GS95] Michelangelo Grigni and Michael Sipser. Monotone separation of logarithmic space from logarithmic depth. *J. Comput. Syst. Sci.*, 50(3):433–437, 1995. 1

[Jac97] Jeffrey Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *Journal of Computer and System Sciences*, 55(3):414–440, 1997. 1.2

[Kal10] Gil Kalai. Noise stability and threshold circuits. Gil Kalai's *Combinatorics and more* blog, http://gilkalai.wordpress.com/2010/02/10/noise-stability-and-threshold-circuits/, 2010. 1

[Kor03] A. D. Korshunov. Monotone Boolean functions. *Russian Math. Surveys (Uspekhi Mat. Nauk)*, 58(5):929–1001, 2003. 1

[KP97] Matthias Krause and Pavel Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theoretical Computer Science*, 174(1–2):137–156, 1997. 1.2

[KRW91] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via direct sum in communication coplexity. In *Structure in Complexity Theory Conference*, pages 299–304, 1991. 1

[KW88] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. In *STOC '88: Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 539–550, New York, NY, USA, 1988. ACM. 1

[Oko82] E. Okol'nishnikova. On the influence of negations on the complexity of a realization of monotone Boolean functions by formulas of bounded depth (in Russian). *Metody Diskret. Analiz.*, 38:74–80, 1982. (document), 1

[OW07] Ryan O'Donnell and Karl Wimmer. Approximation by DNF: examples and counterexamples. In *Proceedings of the 34th Annual International Colloquium on Automata, Languages and Programming*, pages 195–206, 2007. 1, 1.2, 5

[Pot10] Aaron Potechin. Bounds on monotone switching networks for directed connectivity. In *Symposium on Foundations of Computer Science (FOCS)*, pages 553–562, 2010. 1

[Qui54] Willard Van Orman Quine. Two theorems about truth functions. *Bol. Soc. Math. Mexicana*, 10:64–70, 1954. (document), 1

[Raz85] Aleksandr A. Razborov. Lower bounds for the monotone complexity of some boolean functions. *Soviet Mathematics Doklady*, 31:354–357, 1985. (document), 1, 1

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999. 1

[RR97] A. Razborov and S. Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, 1997. 1

[RW90] R. Raz and A. Wigderson. Monotone circuits for matching require linear depth. In *Proceedings of the 22nd ACM Symposium on Theory of Computing*, pages 287–292, 1990. 1

[Tar88] Éva Tardos. The gap between monotone and non-monotone circuit complexity is exponential. *Combinatorica*, 8(1):141–142, 1988. 1