Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 85 (2019)

Approximate Degree-Weight and Indistinguishability

Xuangui Huang
Northeastern University
stslxg@ccs.neu.edu

Emanuele Viola
Northeastern University
viola@ccs.neu.edu

June 19, 2019

Abstract

We prove that the OR function on n bits can be point-wise approximated with error ε by a polynomial of degree $O(k)$ and weight $2^{O(n\log(1/\varepsilon)/k)}$, for any $k \geq \sqrt{n\log(1/\varepsilon)}$. This result is tight for $k=(1-\Omega(1))n$. Previous results were either not tight or had $\varepsilon=\Omega(1)$. In general we obtain a tight approximation result for any symmetric function. Building on this we also obtain an approximation result for bounded-width CNF. For these two classes no such result was known.

One motivation for such results comes from the study of indistinguishability. Two distributions P, Q over n-bit strings are (k,δ)-indistinguishable if their projections on any k bits have statistical distance at most δ. The above approximations give values of (k,δ) that suffice to fool OR, symmetric functions and bounded-width CNF, and the first result is tight for all k while the second result is tight for $k=(1-\Omega(1))n$. Finally, we show that any two (k,δ)-indistinguishable distributions are $O(n)^{k/2}\delta$-close to two distributions that are $(k,0)$-indistinguishable, improving the previous bound of $O(n)^k\delta$.

1 Introduction

The idea of approximating boolean functions point-wise using real-valued polynomials of "low complexity" has been a powerful tool in theoretical computer science. A natural notion of complexity of the polynomial is its degree, extensively studied since the seminal work by Nisan and Szegedy [NS94]. We can also consider the *weight* of the approximating polynomial, that is the sum of absolute value of its coefficients. This is a more interesting quantity when studied in conjunction with degree, because for example the parity function has weight 1 when we allow the degree to be n, but it requires exponential weight for degree $0.001n$ as our results will show. The weight also depends on the basis we use: the parity function and the OR function have constant weights over $\{-1,1\}^n$ for degree n but they require exponential weights over $\{0,1\}^n$ (see Section 7). But interestingly we have low-degree low-weight approximations of AND over $\{0,1\}^n$, as discussed later.

Bounds on the weight of approximations have several applications, ranging from differential privacy [CTUW14], to attribute efficient learning [KS06, STT12], and to indistinguishability [BW17].

Bogdanov and Williamson [BW17] showed tight tradeoffs for approximating OR over $\{-1,1\}^n$ within constant error. To set context, recall that the approximate degree of OR is $\Theta(\sqrt{n})$ [NS94]. They showed that OR can be approximated in degree $O(k)$ and weight $2^{O(n/k)}$ for $k \geq \sqrt{n}$, and this is tight. Prior to their results, Servedio et al. [STT12] showed

ISSN 1433-8092

$w = 2^{\widetilde{O}(n/k)}$; Chandrasekaran et al. [CTUW14] showed $w = 2^{\widetilde{O}\left(n \log^2(1/\varepsilon)/k\right)}$ for error ε when $k \geq \sqrt{n}\log(1/\varepsilon)$; Bun and Thaler [BT15] showed $w = 2^{\Omega(n/k)}$.

Apparently little was known for arbitrary symmetric functions and non-symmetric functions.

1.1 Our results: approximate degree-weight

We prove a tight result for approximating the OR function. This refines the result in [BW17] by including the dependency on the error ε, and the upper bound improves on the result in [CTUW14] by getting a better dependence on $\log(1/\varepsilon)$ and removing the other log terms. Jumping ahead, this improvement is critical to obtain our result for t-CNF, because we will need to set ε exponentially small.

We state the upper and lower bounds as separate theorems. Note that in the lower bound for $k \leq \frac{c}{2}n$ we have $2^{(cn/k-1)\log(1/\varepsilon)} \geq 2^{cn\log(1/\varepsilon)/2k}$, so the upper bound is tight for $k = (1-\Omega(1))n$.

Theorem 1.1. *For every ε, n, k satisfying $\sqrt{n\log(1/\varepsilon)} \leq k \leq n$, OR_n can be ε-approximated over $\{-1,1\}^n$ in degree $O(k)$ and weight $2^{O(n\log(1/\varepsilon)/k)}$.*

Theorem 1.2. *There exists a constant $c < 1$ such that for every ε, n, $k \leq n$, if a polynomial p ε-approximates OR_n over $\{-1,1\}^n$ in degree k, then its weight is at least $2^{(cn/k-1)\log(1/\varepsilon)}$.*

These theorems are special cases of the following results for *symmetric functions* (a function is symmetric if its value only depends on the Hamming weight of the input). For $0 \leq t \leq \frac{n}{2}$, let $\mathsf{SYM}_{n,t}$ denote the class of symmetric functions such that t is the smallest number to make these functions constant on inputs of Hamming weight in $(t, n-t)$. To set context, the ε-approximate degree of any $f \in \mathsf{SYM}_{n,t}$ is $\Theta\left(\sqrt{n(\log(1/\varepsilon)+t)}\right)$ [Pat92, dW08].

Again we state the upper and lower boundes separately.

Theorem 1.3. *For every ε, n, t, k satisfying $\sqrt{n(\log(1/\varepsilon)+t)} \leq k \leq n$, every function $f \in \mathsf{SYM}_{n,t}$ can be ε-approximated over $\{-1,1\}^n$ in degree $O(k)$ and weight $2^{O(n(\log(1/\varepsilon)+t)/k)}$.*

Theorem 1.4. *There exists a constant $c < 1$ such that for every ε, n, t, and k with $k \leq \frac{c}{2}n$ and $0 \leq t \leq \frac{n}{2}$, if a polynomial p ε-approximates $f \in \mathsf{SYM}_{n,t}$ over $\{-1,1\}^n$ in degree k, then its weight is at least $2^{\Omega(n(\log(1/\varepsilon)+t)/k)}$.*

Independently and concurrently, Bogdanov et al. [BMTW19] obtained a similar result with a slightly weaker upper bound.

We then move to non-symmetric functions. A t-CNF is a CNF with clauses of size t. Sherstov [She18] proved that the ε-approximate degree of t-CNF is $O_t\left(n^{\frac{t}{t+1}}(\log(1/\varepsilon))^{\frac{1}{t+1}}\right)$. For $t = 2$ and constant ε this is $O(n^{2/3})$. We prove the following degree-weight approximation for t-CNF, which recovers [She18] and shows that the larger the degree k the smaller the weight w we can have, up to about $w = 2^{O(n^{1-1/t})}$. For $t = 2$ the latter is $2^{O(\sqrt{n})}$.

Theorem 1.5. *For every ε, n, t, k satisfying $n^{\frac{t}{t+1}}(\log(1/\varepsilon))^{\frac{1}{t+1}} \leq k \leq n$, there exists constant c_t depending on t such that any t-CNF can be ε-approximated over $\{-1,1\}^n$ in degree $c_t \cdot k$ and weight $2^{c_t \cdot n(\log(1/\varepsilon))^{1/t}/k^{1/t}}$.*

1.2 Indistinguishability

One of our motivations for these approximation results comes from our interest in indistinguishability. Two distributions on n bits are called *k-wise indistinguishable* if the marginals on any k bits are identical. It seems natural to ask which functions are *fooled* by k-wise indistinguishability, or in other words cannot distinguish any two k-wise indistinguishable distributions. Linear programming duality shows that k-wise indistinguishability fools a function f if and only if the approximate degree of f is at most k [BIVW16]. It could be a convenient framework to prove approximate degree results, see for example the proof in [Vio17, Lecture 6, 7] of the approximate degree of the so-called $\mathsf{AND\text{-}OR}$ tree [BT13, She13].

We study a natural relaxation of indistinguishability [BIVW16], defined next.

Definition 1.6. *Two distributions on n bits are (k,δ)-indistinguishable if the marginals on any k bits are δ-close in statistical distance.*

*A function $f\colon \{0,1\}^n \to \mathbb{R}$ is ε-*fooled *by (k,δ)-indistinguishability if for any two (k,δ)-indistinguishable distributions P and Q we have $|\mathbb{E}[f(P)] - \mathbb{E}[f(Q)]| \le \varepsilon$.*

Actually in the aforementioned paper, Bogdanov and Williamson [BW17] proved tradeoff results in terms of (k,δ)-indistinguishability. They showed that if f can be ε-approximated over $\{-1,1\}^n$ in degree k and weight w, then f is ε-fooled by (k,δ)-indistinguishability for $\delta = \varepsilon/w$. Using this they showed that $k \ge \sqrt{n}$, (k,δ)-indistinguishable fools OR for any $\delta = 2^{-O(n/k)}$. They also show that this is tight.

1.3 Our results: (k,δ)-indistinguishability

Using the above-mentioned connection in [BW17] (see also Theorem 2.5), tradeoffs between degree and weight imply tradeoffs between k and δ. Therefore the following "fools" theorems for OR, symmetric functions, and t-$\mathsf{CNF}/\mathsf{DNF}$ follow from our degree-weight tradeoffs upper bounds for approximating these functions.

Theorem 1.7. *For every ε, n, and k satisfying $\Omega\Big(\sqrt{n\log(1/\varepsilon)}\Big) \le k \le n$, $\big(k, 2^{-O(n\log(1/\varepsilon)/k)}\big)$-indistinguishability ε-fools OR_n.*

Theorem 1.8. *For every ε, n, t, k satisfying $\Omega\Big(\sqrt{n(\log(1/\varepsilon)+t)}\Big) \le k \le n$, $\big(k, 2^{-O(n(\log(1/\varepsilon)+t)/k)}\big)$-indistinguishability ε-fools any function $f \in \mathsf{SYM}_{n,t}$.*

Theorem 1.9. *For every ε, n, t, k satisfying $n^{\frac{t}{t+1}}(\log(1/\varepsilon))^{\frac{1}{t+1}} \le k \le n$, there exists constant c_t depending on t such that $\Big(c_t \cdot k, 2^{-c_t \cdot n(\log(1/\varepsilon))^{1/t}/k^{1/t}}\Big)$-indistinguishability ε-fools t-$\mathsf{CNF}/\mathsf{DNF}$ on n variables.*

We also prove the following "does not fool" theorems, matching the first two "fools" results. Theorem 1.10 shows that Theorem 1.7 is tight for all $k \le n$, while Theorem 1.11 shows that Theorem 1.8 is tight for $k = (1-\Omega(1))n$. Using Theorem 2.5, they imply the degree-weight tradeoff lower bounds in Theorem 1.2 and 1.4. This is how the latter are proved in this paper.

Theorem 1.10. *For every ε, n, and k, $\big(k, 2^{-\Omega(n\log(1/\varepsilon)/k)}\big)$-indistinguishability doesn't ε-fool OR_n.*

Theorem 1.11. *There exists a constant $c' < 1$ such that for every ε, n, t, and k with $k \le c'n$ and $0 \le t \le \frac{n}{2}$, there exists function $f \in \mathsf{SYM}_{n,t}$ such that $\big(k, 2^{-\Omega(n(\log(1/\varepsilon)+t)/k)}\big)$-indistinguishability doesn't ε-fool f.*

The independent work by Bogdanov et al. [BMTW19], mentioned earlier, also obtained similar "does not fool" result for symmetric functions with constant ε (moreover, unlike ours, their distributions do not depend on k).

Finally, we improve the result in [BIVW16] about k-wise indistinguishability vs. (k,δ)-indistinguishability, analogous to the k-wise independence vs. almost k-wise independence results in [AGM03, Theorem 2.1], [OZ18]. This result is tight because of the distributions given in [O'D14, Theorem 1.2] when k is constant.

Theorem 1.12. *If P and Q are (k,δ)-indistinguishable, then they are $O(e^k n^{k/2}\delta)$-close to P' and Q' that are k-wise indistinguishable.*

1.4 Techniques

The proof of Theorem 1.12 follows the proof in [OZ18], which applies to independence rather than indistinguishability.

Existence of low degree-weight polynomials. As observed in [STT12, BW17], the Chebyshev polynomials T_d has degree d and weight $2^{O(d)}$, and by composing it with the monomial $x^{k/d}$, which has *high-degree but weight just one*, we can get a polynomial $T_d(x^{k/d})$ with larger degree $O(k)$, whose weight is still $2^{O(d)}$, and maintains some of the properties of Chebyshev polynomials. For example it is bounded on $[0,1]$ and has derivative $\geq d^2 \cdot \frac{k}{d} \geq kd$ for $x \geq 1$. At a high level, Theorem 1.1, 1.3, and 1.5 follow by applying such an idea to the constructions by Sherstov [She18].

"Does not fool" theorems. The notion of fooling by (k,δ)-indistinguishability does not seem to have a dual characterization, because there does not seem to be a way to express statistical tests in the dual LP. Indeed in Theorem 2.5 while we are considering the weight of the approximations in the dual, we are essentially restricting the statistical tests to the parity tests in the primal, which are not equivalent and can be separated easily for small-bias distributions [NN93]. Therefore degree-weight tradeoff lower bounds do not imply "does not fool" results. Instead we use a different method.

For Theorem 1.10 and 1.11 we reduce it to the case of k-wise indistinguishability (that is $\delta = 0$) by Lemma 5.1, generalizing the proof in [BW17]. By inserting 0's into some random indices, we generate (k,δ)-indistinguishable distributions from k'-wise indistinguishable distributions while keeping their Hamming weight, for suitable settings of k, δ, and k'. Then the result follows from approximate degree lower bound of symmetric functions.

1.5 Organization

In Section 2 we provide useful facts about Chebyshev polynomials, weights, and its connection to (k,δ)-indistinguishability. In Section 3 we prove Theorem 1.12. In Section 4 we prove Theorem 1.3 for symmetric functions and in particular Theorem 1.1 for OR, thus also proving Theorem 1.8 and 1.7. We prove matching lower bounds (Theorem 1.2, 1.4, 1.10, 1.11) in Section 5. In Section 6 we prove Theorem 1.5 and 1.9 for t-CNF. We show that PARITY and OR require exponential weight over $\{-1,1\}^n$ in Section 7. Finally we list some open problems in Section 8.

2 Preliminaries

Let $[n]$ denote the set $\{1,2,\ldots,n\}$.

Weight of polynomials. We denote the *weight* of polynomial p by $\|p\|$. Over $\{-1,1\}^n$, $\|p\|$ is the ℓ_1 Fourier weight of p [O'D14]. It has the following properties.

Claim 2.1 ([She18, Fact 2.7])**.** *For any polynomials p and q,*

- $\|ap\| = |a| \cdot \|p\|$ *for any* $a \in \mathbb{R}$*;*

- $\|p+q\| \le \|p\| + \|q\|$;
- $\|p \cdot q\| \le \|p\| \cdot \|q\|$.

Claim 2.2 ([BW17, Fact 8, 9])**.** *For any univariate polynomial p of degree k,*

(i) if $q(x) = p(ax^t + b)$ where $|a| + |b| \ge 1$ and $t \ge 1$, then $\|q\| \le (|a| + |b|)^k \|p\|$;

(ii) if $q(x_1, \ldots, x_n) = p(\sum_{i=1}^n x_i/n)$ then $\|q\| \le \|p\|$.

Boolean basis and Fourier basis. For Boolean functions, traditionally we use 0 for False, and 1 for True, which is called *the Boolean basis*. Alternatively we can also use *the Fourier basis*, where 1 is used for False and -1 for True. In some cases the Fourier basis is more convenient as negation of variables becomes negation of values; while in some other case the Boolean basis is more convenient as multiplication is equivalent to AND. The *Hamming weight* of a string is the number of bits representing true.

The degree of a polynomial is basis invariant, but the weight is not. The following lemma shows one direction.

Lemma 2.3. *For any polynomial $f\colon \{0,1\}^n \to \mathbb{R}$ on the Boolean input basis, we have a polynomial representing the same function on the Fourier input basis with the same weight.*

Proof. Define $g\colon \{-1,1\}^n \to \mathbb{R}$ by $g(x) = f\left(\frac{1-x_1}{2}, \frac{1-x_2}{2}, \ldots, \frac{1-x_n}{2}\right)$, and the result follows from a multivariate version of Claim 2.2 (i) with $|a| + |b| = 1$. $\square$

Chebyshev polynomials. Chebyshev polynomials (c.f. [Che98]), denoted as T_d for each degree d, is a sequence of orthogonal univariate polynomials that can be uniquely defined by $T_d(\cos x) = \cos dx$ for each d. Its value is given by $T_d(x) = \frac{1}{2}\left(\left(x + \sqrt{x^2-1}\right)^d + \left(x - \sqrt{x^2-1}\right)^d\right)$.

Claim 2.4 (c.f. [She18, BW17])**.** *For degree-d Chebyshev polynomial T_d, we have the following properties:*

(i) $T_d(1) = 1$;

(ii) $T_d\left(\cos\left(\frac{2i-1}{2d}\pi\right)\right) = 0$, for $i \in [d]$;

(iii) $|T_d(z)| \le 1$ for $z \in [-1,1]$;

(iv) $T_d'(t) \ge d^2$ for $t \in [1,\infty)$, so T_d is monotonically increasing on $[1,\infty)$;

(v) $T_d(1+\delta) \ge 2^{d\sqrt{\delta}-1}$ for $\delta \in [0,1]$;

(vi) $\|T_d\| \le 2^{2d}$.

Fooling by (k,δ)-Indistinguishability. The following theorem shows that low-degree low-weight approximation implies fooling by (k,δ)-indistinguishability.

Theorem 2.5 ([BW17])**.** *Given any function $f\colon \{-1,1\}^n \to \mathbb{R}$, for any k and δ we have*

$$\max_{P,Q\colon (k,\delta)\text{-indist.}} |\mathbb{E}[f(P)] - \mathbb{E}[f(Q)]| \le \min_{\substack{g\colon \varepsilon\text{-approx. } f \\ \deg(g) \le k}} 2\varepsilon + 2\delta\|g\|.$$

3 Proof of Theorem 1.12

More generally we are going to prove that for any $k \le n$, any two distributions P and Q over $\{-1,1\}^n$ are w-close to some k-wise indistinguishable distributions P' and Q', where $w = e^k \sqrt{\sum_{|S| \le k} (\mathbb{E}[\chi_S(P)] - \mathbb{E}[\chi_S(Q)])^2}$.

To prove this theorem, we need the following lemmas.

Lemma 3.1. *Let* $\phi\colon \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ *be a polynomial of degree at most* k*, then*

$$\|\widehat{\phi}\|_2 = \sqrt{\sum_{|S|\leq k} \widehat{\phi}(S)^2} \leq e^k \widehat{\phi}(\varnothing).$$

Proof. This lemma is essentially the same as Lemma 2.1 in [OZ18]. We have

$$\|\widehat{\phi}\|_2 = \mathbb{E}[\phi^2(x)] \leq e^k \mathbb{E}[|\phi(x)|] = e^k \mathbb{E}[\phi(x)] = e^k \widehat{\phi}(\varnothing),$$

where the first step comes from Parseval's Theorem ([O'D14, Section 1.4]), the second step holds by hypercontractivity ([O'D14, Theorem 9.22]), the third step follows from the fact that $\phi(x) \geq 0$ for all x, and the last step follows from standard Fourier analysis ([O'D14, Proposition 1.8]). $\square$

Lemma 3.2 (Farkas' Lemma)**.** *Let* $A \in \mathbb{R}^{m\times n}$ *and* $b \in \mathbb{R}^m$*. Then* exactly one *of the following two statements is true:*

1. *There exists an* $x \in \mathbb{R}^n$ *such that* $Ax = b$ *and* $x \geq 0$*.*
2. *There exists a* $y \in \mathbb{R}^m$ *such that* $A^\top y \geq 0$ *and* $b^\top y < 0$*.*

Proof of Theorem 1.12. Given distributions P and Q, let $b_S = \mathbb{E}[\chi_S(P)] - \mathbb{E}[\chi_S(Q)]$ for each $|S| \leq k$. We are going to find distributions P'' and Q'' such that for all $1 \leq |S| \leq k$,

$$\mathbb{E}[\chi_S(P'')] - \mathbb{E}[\chi_S(Q'')] = -\frac{b_S}{w}.$$

Suppose such P'' and Q'' exist, then we set $P' = \frac{P+wP''}{1+w}$ and $Q' = \frac{Q+wQ''}{1+w}$. Then the statistical distance $\Delta(P,P') \leq \frac{w}{1+w} \leq w$ and similarly for Q and Q'. We also have

$$\mathbb{E}[\chi_S(P')] - \mathbb{E}[\chi_S(Q')] = \frac{1}{1+w}\big(\mathbb{E}[\chi_S(P)] - \mathbb{E}[\chi_S(Q)] + w(\mathbb{E}[\chi_S(P'')] - \mathbb{E}[\chi_S(Q'')])\big) = 0,$$

for all $1 \leq |S| \leq k$, therefore P' and Q' are k-wise indistinguishable.

To prove the existence of P'' and Q'', we write it as an LP feasibility problem with variables $p(x)$ for $x \in \{-1,1\}^n$ corresponding to $P''(x)$ and $q(x)$ for $Q''(x)$, and the following constraints.

$\sum_{x\in\{-1,1\}^n} p(x) = 1$	
$\sum_{x\in\{-1,1\}^n} q(x) = 1$	
$\sum_{x\in\{-1,1\}^n} \chi_S(x)p(x) - \sum_{x\in\{-1,1\}^n} \chi_S(x)q(x) = -\frac{b_S}{w}$	for each $1 \leq \lvert S\rvert \leq k$
$p(x) \geq 0$	for each $x \in \{-1,1\}^n$
$q(x) \geq 0$	for each $x \in \{-1,1\}^n$

By Farkas' Lemma, to prove that it is feasible is equivalent to prove that the following LP is infeasible, with unconstrained variables $y_\varnothing$, $y'_\varnothing$, and y_S for $1 \leq |S| \leq k$.

$y_\varnothing + \sum_{1\leq \lvert S\rvert\leq k} y_S\chi_S(x) \geq 0$	for each $x \in \{-1,1\}^n$
$y'_\varnothing - \sum_{1\leq \lvert S\rvert\leq k} y_S\chi_S(x) \geq 0$	for each $x \in \{-1,1\}^n$
$y_\varnothing + y'_\varnothing - \frac{1}{w}\sum_{1\leq \lvert S\rvert\leq k} y_S b_S < 0$	

To prove that it is infeasible, it suffices to prove that any assignments satisfying the first two sets of constraints must violate the third one. Summing up the first set of constraints for all $x \in \{-1,1\}^n$, we get $y_\varnothing \geq 0$, and similarly we have $y'_\varnothing \geq 0$. Now define a polynomial

$\phi = y_\varnothing + \sum_{1\le|S|\le k} y_S\chi_S$, then the degree of ϕ is at most k and the first two sets of constraints become

$$\phi(x) \ge 0 \text{ for each } x \in \{-1,1\}^n,$$
$$\phi(x) \le y_\varnothing + y'_\varnothing \text{ for each } x \in \{-1,1\}^n.$$

By Lemma 3.1 we have $\sqrt{\sum_{|S|\le k} y_S^2} \le e^k y_\varnothing$, and we set $w = e^k\sqrt{\sum_{|S|\le k} b_S^2}$. Note that we have $w = O(e^k n^{k/2}\delta)$ when P and Q are (k,δ)-indistinguishable. Therefore

$$\frac{1}{w}\sum_{1\le|S|\le k} y_S b_S \le \frac{1}{w}\sum_{1\le|S|\le k} |y_S||b_S| \le \frac{1}{w}\sqrt{\sum_{|S|\le k} y_S^2}\sqrt{\sum_{|S|\le k} b_S^2} \le \frac{e^k y_\varnothing}{w}\sqrt{\sum_{|S|\le k} b_S^2} = y_\varnothing \le y_\varnothing + y'_\varnothing,$$

where the second inequality holds by Cauchy-Schwarz. This violates the third constraint thus completes our proof. $\square$

4 Proofs of Theorem 1.1, 1.3, 1.7, and 1.8

Every symmetric function is a linear combination of the characteristic functions of each level of the Boolean hypercube, given by $\mathsf{EXACT}_{n,r}\colon \{0,1\}^n \to \{0,1\}$ as

$$\mathsf{EXACT}_{n,r}(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i = r, \\ 0 & \text{otherwise.} \end{cases}$$

Note that both $\mathsf{EXACT}_{n,t}$ and $\mathsf{EXACT}_{n,n-t}$ belong to the class $\mathsf{SYM}_{n,t}$ for $0 \le t \le \frac{n}{2}$. The following theorem gives approximating polynomials with degree-weight tradeoff behaviors for $\mathsf{EXACT}_{n,n-t}$ on the Boolean input basis for arbitrary accuracy, thus also on the Fourier input basis by Lemma 2.3. In particular, as $\mathsf{AND}_n = \mathsf{EXACT}_{n,n} \in \mathsf{SYM}_{n,0}$, this theorem gives approximations for AND_n with degree-weight tradeoff behaviors for arbitrary accuracy on both input basis.

Theorem 4.1. *For every ε, k, n, t such that $\sqrt{n(\log(1/\varepsilon)+t)} \le k \le n$ and $0 \le t \le \frac{n}{2}$, there is a polynomial $p\colon \{0,1\}^n \to \mathbb{R}$ that ε-approximates $\mathsf{EXACT}_{n,n-t}$ with degree $O(k)$ and weight $2^{O(n(\log(1/\varepsilon)+t)/k)}$.*

Overview of the proof of Theorem 4.1.

Our goal is to construct a univariate polynomial p^* to ε-approximate $\mathsf{EXACT}_{n,n-t}$ such that $|p^*(\frac{i}{n}) - 1| \le \varepsilon$ for $i = n-t$ and $|p^*(\frac{i}{n})| \le \varepsilon$ for all $i \in [n] \setminus \{n-t\}$. Besides, p^* must have the desired degree and weight.

We can already get a good approximation using Chebyshev polynomials. Let

$$\ell = \log\frac{2}{\varepsilon} + t, \quad d = \sqrt{n\ell}.$$

Without loss of generality assume $d \in \mathbb{N}$, define univariate polynomials q_0 and p_0 by

$$q_0(z) = T_d\left(\frac{n}{n-\ell}\cdot z\right), \quad p_0(z) = q_0(z)/q_0\left(\frac{n-t}{n}\right).$$

For $z \in [0, \frac{n-\ell}{n}]$, we have $\frac{n}{n-\ell}\cdot z \le 1$ thus by Claim 2.4 (iii) we have $\left|q_0(\frac{i}{n})\right| \le 1$ for all $i = 0, 1, \ldots, n-\ell$. The value of $q_0(\frac{n-t}{n})$ is also large enough, as by Claim 2.4 (iii) we have

$$q_0\left(\frac{n-t}{n}\right) = T_d\left(1 + \frac{\ell-t}{n-\ell}\right) \ge 2^{\sqrt{n\ell}\sqrt{\frac{\ell-t}{n-\ell}}-1} \ge 2^{(\ell-t)-1} = \frac{1}{\varepsilon},$$

where the third step uses $\frac{n}{n-\ell} \geq 1$ and $\ell \geq \ell - t$. Therefore $\left|p_0\left(\frac{i}{n}\right)\right| \leq \varepsilon$ for all $i = 0, 1, \ldots, n-\ell$ and $p_0\left(\frac{i}{n}\right) = 1$ for $i = n - t$, thus p_0 is a good approximation for these i's. We have $\deg(p_0) = d$ and $\|p_0\| = 2^{O(d)}$, which are fixed by n, t, and ε.

To get approximations that have degree-weight tradeoff, we would decrease d and increase the power of $\frac{n}{n-\ell} \cdot z$ inside T_d accordingly. We use the same ℓ, and for any $k \geq \sqrt{n\ell}$, let

$$d = \frac{n\ell}{k},$$

thus d decreases when k increases. Without loss of generality assume $d, \frac{k}{d} \in \mathbb{N}$, define

$$q_1(z) = T_d\left(\left(\frac{n}{n-\ell} \cdot z\right)^{\frac{k}{d}}\right), \quad p_1(z) = q_1(z)/q_1\left(\frac{n-t}{n}\right).$$

Similarly we have $\left|q_1\left(\frac{i}{n}\right)\right| \leq 1$ for all $i = 0, 1, \ldots, n - \ell$. We also get

$$q_1\left(\frac{n-t}{n}\right) = T_d\left(\left(1 + \frac{\ell - t}{n - \ell}\right)^{\frac{k}{d}}\right) \geq T_d\left(1 + \frac{\ell - t}{n - \ell} \cdot \frac{k}{d}\right) \geq 2^{d\sqrt{\frac{\ell-t}{n-\ell}\frac{k}{d}}-1} \geq 2^{(\ell-t)-1} = \frac{1}{\varepsilon},$$

where the second step uses Bernoulli's inequality and Claim 2.4 (iv), the third step uses Claim 2.4 (v), the fourth step uses $\sqrt{kd} = \sqrt{n\ell} \geq \sqrt{n(\ell - t)}$ and $\frac{n}{n-\ell} \geq 1$. Therefore p_1 is a good approximation for some i, namely

$$\left|p_1\left(\frac{i}{n}\right)\right| \leq \varepsilon \qquad \text{for all } i = 0, 1, \ldots, n - \ell, \tag{1}$$

$$p_1\left(\frac{i}{n}\right) = 1 \qquad \text{for } i = n - t. \tag{2}$$

The degree of q_1 and p_1 is $d \cdot \frac{k}{d} = k$. Now we are going to bound their weights. We can write $q_1(z) = T_d(a \cdot z^{k/d})$ with $a = \left(\frac{n}{n-\ell}\right)^{k/d}$, so by Claim 2.4 (vi) and Claim 2.2 (i),

$$\|q_1\| \leq \|T_d\| \cdot a^d = 2^{O(d)} \cdot \left(1 + \frac{\ell}{n - \ell}\right)^k \leq 2^{O(d)} \cdot e^{\frac{k\ell}{n-\ell}},$$

where the last step uses $1 + x \leq e^x$ for all $x \in \mathbb{R}$. We can assume $\ell \leq \frac{3}{4}n$, otherwise $\sqrt{n\ell} = \Omega(n)$ thus $d = \Omega(n)$ so we can simply use p_0. Thus we have $n(n - \ell) \geq \frac{1}{4}n^2 \geq \frac{1}{4}k^2$, so $\frac{k\ell}{n-\ell} = O\left(\frac{n\ell}{k}\right) = O(d)$, therefore $\|q_1\| \leq 2^{O(d)}$, which is the same as $\|T_d\|$ up to the $O(\cdot)$ in the exponent. In other words, we can ignore effect of the scaling term $\frac{n}{n-\ell}$ to the weight.

Consequently,

$$\|p_1\| \leq \|q_1\| \cdot \varepsilon \leq 2^{O(d)} = 2^{O(n(\log(1/\varepsilon)+t)/k)}. \tag{3}$$

Therefore p_1 has the degree-weight tradeoff we need in Theorem 4.1.

The problem of p_1 is that for $i = n - \ell + 1, \ldots, n - t - 1$ and $n - t + 1, \ldots, n$, we have no bound of its value. We need the following construction of auxiliary polynomials $T_{n,m}^{(k)}$ with degree-weight tradeoff that can be made zero on some specific points. Multiplying p_1 by such $T_{n,m}^{(k)}$, we can zero out the value on those i's and get the desired approximations.

Lemma 4.2. *For every n, m, k such that $0 \leq m < n$ and $\sqrt{\frac{n}{n-m}} \leq k \leq \frac{n}{n-m}$, there is a univariate polynomial $T_{n,m}^{(k)}$ of degree $O(k)$ and weight $2^{O\left(\frac{n}{k(n-m)}\right)}$ such that*

$$T_{n,m}^{(k)}(1) = 1, \tag{4}$$

$$T_{n,m}^{(k)}\left(\frac{m}{n}\right) = 0, \tag{5}$$

$$\left|T_{n,m}^{(k)}(z)\right| \le 1, \text{ for any } z \in [0,1]. \tag{6}$$

Building on the overview and assuming the lemma, we now present the proof of Theorem 4.1.

Proof of Theorem 4.1. Now we use the same d and ℓ as discussed above, and define the following univariate polynomial p^* by

$$p^*(z) = p_1(z) \prod_{m=n-\ell}^{n-t-1} T_{n-t,m}^{(k'_m)}\left(\frac{n}{n-t} \cdot z\right) \prod_{m=n-t+1}^{n} \left(1 - \left(T_{m,n-t}^{(k''_m)}\left(\frac{n}{m} \cdot z\right)\right)^2\right),$$

using the auxiliary polynomials from Lemma 4.2, where

$$k'_m = k\sqrt{\frac{n-t}{n\ell(n-t-m)}} \qquad \text{for } m = n-\ell, \ldots, n-t-1,$$

$$k''_m = k\sqrt{\frac{m}{nt(m-n+t)}} \qquad \text{for } m = n-t+1, \ldots, n.$$

First, we need to show that our applications of Lemma 4.2 are legitimate.

- $\sqrt{\frac{n-t}{n-t-m}} \le k'_m \le \frac{n-t}{n-t-m}$ for $m \in [n-t-1] \setminus [n-\ell-1]$: on one hand, we have $k \ge \sqrt{n\ell}$ so $k'_m \ge \sqrt{\frac{n-t}{n-t-m}}$; on the other hand, we have $n - m \le \ell$ and $t < \ell \le n$, so $\frac{n}{\ell} \le \frac{n-t}{\ell-t} \le \frac{n-t}{n-t-m}$, thus by $k \le n$ we have $k'_m \le \sqrt{\frac{n(n-t)}{\ell(n-t-m)}} \le \frac{n-t}{n-t-m}$.
- $\sqrt{\frac{m}{m-n+t}} \le k''_m \le \frac{m}{m-n+t}$ for $m \in [n] \setminus [n-t]$: on one hand, we have $k \ge \sqrt{n\ell}$ and $\ell \ge t$ so $k''_m \ge \sqrt{\frac{\ell m}{t(m-n+t)}} \ge \sqrt{\frac{m}{m-n+t}}$; on the other hand, we have $n - m < t \le n$, so $\frac{n}{t} \le \frac{n-(n-m)}{t-(n-m)} = \frac{m}{m-n+t}$, thus by $k \le n$ we have $k''_m \le \sqrt{\frac{nm}{t(m-n+t)}} \le \frac{m}{m-n+t}$.

Second, we are going to show that p^* is a good approximation. By Lemma 4.2 (6) and Equation (1) we have

$$\left|p^*\left(\frac{i}{n}\right)\right| \le \varepsilon \cdot \prod_{m=n-\ell}^{n-t-1} 1 \cdot \prod_{m=n-t+1}^{n} (1-0^2) = \varepsilon \qquad \text{for all } i = 0, 1, \ldots, n-\ell-1. \tag{7}$$

For all $i = n-\ell, \ldots, n-t-1$, when $m = i$ we have $T_{n-t,m}^{(k'_m)}\left(\frac{n}{n-t} \cdot \frac{i}{n}\right) = T_{n-t,m}^{(k'_m)}\left(\frac{m}{n-t}\right) = 0$. For all $i = n-t+1, \ldots, n$, when $m = i$ we have $T_{m,n-t}^{(k''_m)}\left(\frac{n}{m} \cdot \frac{i}{n}\right) = T_{m,n-t}^{(k''_m)}(1) = 1$ thus $1 - \left(T_{m,n-t}^{(k''_m)}\left(\frac{n}{m} \cdot \frac{i}{n}\right)\right)^2 = 0$. We also have $T_{n-t,m}^{(k'_m)}\left(\frac{n}{n-t} \cdot \frac{i}{n}\right) = 1$ and $T_{m,n-t}^{(k''_m)}\left(\frac{n}{m} \cdot \frac{i}{n}\right) = 0$ when $i = n-t$. Therefore

$$p^*\left(\frac{i}{n}\right) = \begin{cases} 0 & \text{for all } i = n-\ell, \ldots, n-t-1, \\ 1 \cdot \prod_{m=n-\ell}^{n-t-1} 1 \cdot \prod_{m=n-t+1}^{n} (1-0^2) = 1, & \text{for } i = n-t, \\ 0 & \text{for all } i = n-t+1, \ldots, n. \end{cases} \tag{8}$$

Therefore $p^*\left(\frac{i}{n}\right) = 1$ for $i = n-t$ and $\left|p^*\left(\frac{i}{n}\right)\right| \le \varepsilon$ otherwise.

Now we are going to bound the degree and weight of p^*. We have $k'_m = k\sqrt{\frac{n-t}{n\ell(n-t-m)}} \le k\frac{1}{\sqrt{\ell(n-t-m)}}$ and $k''_m = k\sqrt{\frac{m}{nt(m-n+t)}} \le k\frac{1}{\sqrt{t(m-n+t)}}$. By Lemma 4.2,

$$\begin{aligned}
\deg(p^*) &\le k + \sum_{m=n-\ell}^{n-t-1} O\big(k'_m\big) + \sum_{m=n-t+1}^{n} O\big(k''_m\big) \\
&\le O\left(k + \frac{k}{\sqrt{\ell}}\sum_{m=n-\ell}^{n-t-1}\frac{1}{\sqrt{n-t-m}} + \frac{k}{\sqrt{t}}\sum_{m=n-t+1}^{n}\frac{1}{\sqrt{m-n+t}}\right) \\
&= O\left(k + \frac{k}{\sqrt{\ell}}\sum_{i=1}^{\ell-t}\frac{1}{\sqrt{i}} + \frac{k}{\sqrt{t}}\sum_{i=1}^{t}\frac{1}{\sqrt{i}}\right) \\
&= O(k), \qquad (9)
\end{aligned}$$

where in the last step we use $\sum_{i=1}^{n}\frac{1}{\sqrt{i}} = O(\sqrt{n})$ for any $n \in \mathbb{N}$. Similar to the argument in the calculation of $\|q_1\|$, we can safely ignore effects of the scaling terms $\frac{n}{n-t}$ and $\frac{n}{m}$ to the weight of p^*. By (3) and Lemma 4.2 we have

$$\begin{aligned}
\log\|p^*\| &\le O(d) + \sum_{m=n-\ell}^{n-t-1} O\left(\frac{n-t}{k'_m(n-t-m)}\right) + \sum_{m=n-t+1}^{n} O\left(\frac{m}{k''_m(m-n+t)}\right) \\
&= O\left(\frac{n\ell}{k} + \sum_{m=n-\ell}^{n-t-1}\frac{\sqrt{n\ell(n-t)}}{k}\frac{1}{\sqrt{n-t-m}} + \sum_{m=n-t+1}^{n}\frac{\sqrt{ntm}}{k}\frac{1}{\sqrt{m-n+t}}\right) \\
&\le O\left(\frac{n\ell}{k} + \frac{n\sqrt{\ell}}{k}\sum_{i=1}^{\ell-t}\frac{1}{\sqrt{i}} + \frac{n\sqrt{t}}{k}\sum_{i=1}^{t}\frac{1}{\sqrt{i}}\right) \\
&= O\Big(\frac{n}{k}(\log(1/\varepsilon)+t)\Big). \qquad (10)
\end{aligned}$$

Finally, define $p\colon \{0,1\}^n \to \mathbb{R}$ by $p(x) = p^*\Big(\frac{\sum_{i=1}^n x_i}{n}\Big)$. The theorem follows from (7)-(10) and Claim 2.2 (ii). $\square$

Now we prove Theorem 1.3 and Theorem 1.8. We are working on the Fourier basis since we need to negate our input variables. Since $\mathsf{OR}_n \in \mathsf{SYM}_{n,0}$, we get Theorem 1.1 from Theorem 1.3, and Theorem 1.7 from Theorem 1.8.

Proof of Theorem 1.3. By Lemma 2.3 we can get the same results for EXACT on the Fourier basis as in Theorem 4.1. Then we can write f as

$$\begin{aligned}
f(x) &= c + \sum_{i=0}^{t} c'_i \cdot \mathsf{EXACT}_{n,i}(x) + \sum_{i=0}^{t} c''_i \cdot \mathsf{EXACT}_{n,n-i}(x) \\
&= c + \sum_{i=0}^{t} c'_i \cdot \mathsf{EXACT}_{n,n-i}(\overline{x}) + \sum_{i=0}^{t} c''_i \cdot \mathsf{EXACT}_{n,n-i}(x), \qquad (11)
\end{aligned}$$

where c, c'_i's, and c''_i's are fixed reals, and $\overline{x} = (-x_1, \ldots, -x_n)$. Now let $\varepsilon' = \frac{\varepsilon}{2t+2}$, then for $0 \le i \le t$, $\sqrt{n(\log(1/\varepsilon')+i)} = O\Big(\sqrt{n(\log(1/\varepsilon)+t)}\Big)$ so we can ignore the constant factor difference and apply Theorem 4.1 with $\varepsilon = \varepsilon'$ and the same k for each $\mathsf{EXACT}_{n,n-i}$. The degrees of the approximations for EXACT are $O(k)$, so the total degree is also $O(k)$. The weights of the approximations for EXACT are $2^{O(n(\log(1/\varepsilon')+i)/k)} = 2^{O(n(\log(1/\varepsilon)+t)/k)}$. By Claim 2.1 the total weight is $O(t)2^{O(n(\log(1/\varepsilon)+t)/k)} = 2^{O(n(\log(1/\varepsilon)+t)/k)}$. $\square$

Proof of Theorem 1.8. Apply Theorem 1.3, set $\delta = \frac{\varepsilon}{2w} = 2^{-O\left(\frac{n}{k}(\log(1/\varepsilon)+t)\right)}$. Then it follows from Theorem 2.5. $\square$

4.1 Proof of Lemma 4.2

Let $d = \frac{2n}{k(n-m)}$ so $\frac{k}{d} = \frac{k^2(n-m)}{2n}$. As $m < n$ we have $d > 0$ and $\frac{k}{d} > 0$. Without loss of generality assume $d, \frac{k}{d} \in \mathbb{N}$, we can define

$$T_{n,m}^{(k)}(z) = T_d\Big(a \cdot z^{k/d} + b\Big),$$

where $a, b \in \mathbb{R}$ are parameters to be set and T_d is the degree-d Chebyshev polynomial. We have $\deg\Big(T_{n,m}^{(k)}\Big) \le d \cdot \frac{k}{d} = k$.

We set a, b such that

$$a + b = 1, \tag{12}$$

$$a\Big(\frac{m}{n}\Big)^{k/d} + b = \cos\Big(\frac{\pi}{2d}\Big), \tag{13}$$

then Property (4) follows from (12) and Claim 2.4 (i), and Property (5) follows from (13) and Claim 2.4 (ii) with $i = 1$.

Our goal is to prove $0 \le a \le 1$. Assume this is true. From (12) we have $b \in [0, 1]$. Hence $a \cdot z^{k/d} + b$ is increasing in $z \in [0, 1]$, mapping $[0, 1]$ to $[b, 1] \subseteq [0, 1]$. Therefore Property (6) follows from Claim 2.4 (iii). Besides, we have $|a| + |b| = a + b = 1$, so by Claim 2.4 (vi) and Claim 2.2 (i) we have $\|T_{n,m}^{(k)}\| = 1 \cdot 2^{O(d)} = 2^{O\left(\frac{n}{k(n-m)}\right)}$. Therefore $T_{n,m}^{(k)}(z)$ is the desired polynomial.

To see that $a \in [0, 1]$, we solve the linear equations (12) and (13) to get

$$a = \frac{1 - \cos\left(\frac{\pi}{2d}\right)}{1 - \left(\frac{m}{n}\right)^{k/d}}. \tag{14}$$

Because $m < n$, we have $\frac{m}{n} < 1$, thus $1 - \left(\frac{m}{n}\right)^{k/d} > 0$. We always have $1 - \cos\left(\frac{\pi}{2d}\right) \ge 0$. Therefore from (14) we have $a \ge 0$. On the other hand, let $u = \frac{n}{n-m}$, from (14) we can get

$$a \le \frac{\frac{1}{2}\left(\frac{\pi}{2d}\right)^2}{1 - \left(1 - \frac{n-m}{n}\right)^{k/d}} \le \frac{\frac{\pi^2}{16}\frac{k^2}{2u^2}}{1 - e^{-\frac{k^2}{2u^2}}}, \tag{15}$$

where the first step uses $\cos x \ge 1 - \frac{1}{2}x^2$ for $x \in \mathbb{R}$ in the numerator, and the second step uses the value of d in the numerator and $1 + x \le e^x$ for $x \in \mathbb{R}$ in the denominator. Since $\sqrt{\frac{n}{n-m}} \le k \le \frac{n}{n-m}$ and $m < n$, we have $0 < \frac{k^2}{2u^2} \le \frac{1}{2}$. Consider the function $f(z) = \frac{z}{1-e^{-z}}$ on $z \in (0, \frac{1}{2}]$. Its derivative $f'(z) = \frac{e^z(e^z-z-1)}{(e^z-1)^2} > 0$ for $z \in (0, 1]$, hence $f(z)$ is increasing in $(0, \frac{1}{2}]$, thus $f(z) \le f\left(\frac{1}{2}\right)$. From (15) we have $a \le \frac{\pi^2}{16} f\left(\frac{1}{2}\right) = \frac{\pi^2\sqrt{e}}{32(\sqrt{e}-1)} < 1$, finishing our proof. $\square$

5 Proofs of Theorem 1.2, 1.4, 1.10, and 1.11

First we generalize the proof in [BW17], reducing the problem into fooling by k-wise indistinguishability. We use ℓ to deal with non-constant ε and symmetric functions with non-constant t.

Lemma 5.1. *Let c'' be any constant. For every n, k and ℓ satisfying $\frac{c''}{16}\sqrt{n\ell} \le k \le \frac{c''}{16}n$, there exists n' with $\ell \le n' \le n$ such that if there exist k'-wise indistinguishable distributions P',Q' over $\{0,1\}^{n'}$ for $k' = c''\sqrt{n'\ell}$, then there exist distributions P, Q over $\{0,1\}^n$ such that the Hamming weight distribution $|P| \equiv |P'|$, $|Q| \equiv |Q'|$, and P and Q are $\left(k, 2^{-\Omega(n\ell/k)}\right)$-indistinguishable.*

Proof. To sample from P, and Q respectively, we select n'-many indices from $[n]$ uniformly at random as "active" indices and then fill in these n' indices using a random sample from P', and Q' respectively; for other indices we simply set them to be 0. Obviously this process keeps the Hamming weight of the samples.

Suppose P' and Q' are k'-wise indistinguishable with $k' = c''\sqrt{n'\ell}$ for some constant $0 < c'' < 1$. For any k indices S of P and Q, if there are at most k' active indices in S, then their projections on S are identical by k'-wise indistinguishability. Therefore the probability that statistical test on k bits can distinguish P from Q is bounded by the probability that such event doesn't happen. By tail bounds of hypergeometric distribution [Hoe63], we have

$$\Pr[\text{more than } k' \text{ active indices in } S] \le e^{-kD\left(\frac{k'+1}{k}||\frac{n'}{n}\right)} = 2^{-\Omega\left(kD\left(\frac{k'}{k}||\frac{n'}{n}\right)\right)}, \tag{16}$$

where $D(a||b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$ is the Kullback-Leibler divergence. By a lower bound from Hellinger distance H, for any p and any $a \ge 16$, we have

$$D(ap||p) \ge 2H^2(ap||p) \ge (\sqrt{ap} - \sqrt{p})^2 = (\sqrt{a}-1)^2 p \ge \frac{1}{2}ap,$$

where the last step comes from the fact that $2\sqrt{a} \le \frac{a}{2}$ for $a \ge 16$. Now we set $n' = \frac{c''^2}{16^2}\frac{n^2}{k^2}\ell$, then we have $k' = \frac{c''^2}{16}\frac{n}{k}\ell$, thus $\frac{k'}{k}/\frac{n'}{n} = 16$, therefore we have

$$2^{-\Omega\left(kD\left(\frac{k'}{k}||\frac{n'}{n}\right)\right)} \le 2^{-\Omega\left(k\frac{k'}{k}\right)} = 2^{-\Omega(k')} = 2^{-\Omega(n\ell/k)}.$$

For $\ell \le n'$, we need $\frac{c''^2}{16^2}\frac{n^2}{k^2} \ge 1$ thus $k \le \frac{c''}{16}n$. For $n' \le n$, we need $\frac{c''^2}{16^2}\frac{n}{k^2}\ell \le 1$ thus $k \ge \frac{c''}{16}\sqrt{n\ell}$. □

Combining the following equivalence between approximate degrees and bounded indistinguishability, with the ε-approximate degree lower bound of symmetric functions due to de Wolf [dW08], improving [Pat92, BT13, She08], we can obtain k-wise indistinguishable distributions that don't fool symmetric functions.

Theorem 5.2 ([BIVW16, Theorem 1.1])**.** *For every ε, n, k, and $f\colon \{0,1\}^n \to \{0,1\}$ the following are equivalent:*

1. *f is not ε-fooled by k-wise indistinguishability;*
2. *The $\varepsilon/2$-approximate degree of f is bigger than k.*

Theorem 5.3 ([dW08])**.** *For any $f \in \mathsf{SYM}_{n,t}$, its ε-approximate degree is $\Omega\left(\sqrt{n(\log(1/\varepsilon)+t)}\right)$.*

Now we can prove Theorem 1.11 with $f = \mathsf{EXACT}_{n,t} \in \mathsf{SYM}_{n,t}$ for $0 \le t \le \frac{n}{2}$, and similarly for $f = \mathsf{THR}_{n,t+1}$ where $\mathsf{THR}_{n,r}$ is True iff the Hamming weight of its input is at least r.[1] In particular for $\mathsf{OR}_n = \mathsf{THR}_{n,1}$ we get Theorem 1.10 by applying $t = 0$. Note that for $t = O(1)$ (in particular for OR), this theorem works for all k as $(n, \varepsilon^{0.999})$-indistinguishability[2] doesn't ε-fool any non-constant function.

[1]Note that $\mathsf{THR}_{n,r} \in \mathsf{SYM}_{n,r-1}$ for $r \le \frac{n}{2}+1$, equivalently $\mathsf{THR}_{n,t+1} \in \mathsf{SYM}_{n,t}$ for $t \le \frac{n}{2}$.
[2]Indeed (n,δ)-indistinguishability for any $\delta > \varepsilon$.

Proof of Theorem 1.11. We consider function $\mathsf{EXACT}_{n,t}$. Let $\ell = \log(2/\varepsilon) + t$, c'' be the constant in $\Omega(\cdot)$ in Theorem 5.3. Set $c' = \frac{c''}{16}$.

For $k \le c''\sqrt{n\ell}$, by Theorem 5.3 and Theorem 5.2 k-wise indistinguishability does not ε-fool $\mathsf{EXACT}_{n,t}$, hence the theorem for $k \le c''\sqrt{n\ell}$ as $2^{-c''^2 n\ell/k} \le 2^{-k}$.

For $c''\sqrt{n\ell} \le k \le \frac{c''}{16}n$, apply Lemma 5.1 to get n' then Theorem 5.3 and Theorem 5.2 give us k'-wise indistinguishable distributions P' and Q' over $\{0,1\}^{n'}$ that don't ε-fool $\mathsf{EXACT}_{n',t}$. The theorem follows by applying Lemma 5.1 again to get distributions P,Q. $\square$

Theorem 2.5 shows that a polynomial p that $(\varepsilon/2)$-approximates f in degree k must have weight at least ε/δ if (k,δ)-indistinguishability does not ε-fool f. Let $c < 1$ be the constant in the $\Omega(\cdot)$ in Theorem 1.11. Then we get Theorem 1.2 from Theorem 1.10, and Theorem 1.4 from Theorem 1.11.

6 Proofs of Theorem 1.5 and 1.9

Proof of Theorem 1.5 with $t = 2$. Given a 2-CNF F, we can first transform F by the following procedure. For each $i \in [n]$, from F we pick up all the terms that contain x_i unnegated. Let m_i be the number of such terms and $C_1^{(i)}$, $C_2^{(i)}$, …, $C_{m_i}^{(i)}$ be these terms. We remove x_i from them to get $C_1'^{(i)}$, $C_2'^{(i)}$, …, $C_{m_i}'^{(i)}$. If $m_i = 0$, define $f_i'(x) = \neg x_i$, otherwise define it as $\bigwedge_{j=1}^{m_i} C_j'^{(i)}$. Remove all the original terms from F, continue for the next i until $i = n$. Then we do this procedure on the remaining terms of F for each $i \in [n]$ again, but this time we are collecting terms that contain $\neg x_i$ and defining $f_i''(x)$ similarly. At last we define

$$F_1(x) = \bigwedge_{i=1}^{n} x_i \vee f_i'(x), \tag{17}$$

$$F_2(x) = \bigwedge_{i=1}^{n} \neg x_i \vee f_i''(x), \tag{18}$$

and by distributive law we have $F = F_1 \wedge F_2$. Note that all the f_i''s and f_i'''s are 1-CNFs, namely ANDs of literals.

We define $F' \colon \{-1,1\}^n \times \{-1,1\}^{2n} \to \{0,1\}$ by $F'(x,y) = \bigwedge_{i=1}^{2n} y_i \vee f_i(x)$, where $f_i = \begin{cases} f_i' & \text{for } i \in [n] \\ f_{i-n}'' & \text{for } i \in [2n] \setminus [n] \end{cases}$. If we set

$$y_i = \begin{cases} x_i & \text{for } i \in [n] \\ -x_{i-n} & \text{for } i \in [2n] \setminus [n] \end{cases}, \tag{19}$$

we will have $F(x) = F'(x,y)$ for all $x \in \{-1,1\}^n$ by (17) and (18). Therefore it suffices to prove this theorem for $F'(x,y)$. Note here it is important to choose the input basis of F' to be the Fourier basis so we can negate the variables without increasing the weight.

Let $N = 2n$ and $B \in [N]$ be an integer to be determined later. Let S_1, …, $S_{\frac{N}{B}}$ be an even partition of $[N]$ into subsets of size B. For each $i \in [\frac{N}{B}]$ we can define $h_i \colon \{-1,1\}^n \times \{-1,1\}^{S_i} \to \{0,1\}$ by $h_i(x,y) = \bigwedge_{j \in S_i} y_j \vee f_j(x)$ and we have $F'(x,y) = \bigwedge_{i=1}^{\frac{N}{B}} h_i(x,y) = \bigwedge_{\frac{N}{B}}(h_1(x,y), \ldots, h_{\frac{N}{B}}(x,y))$, where $\bigwedge_{\frac{N}{B}} \colon \{0,1\}^{\frac{N}{B}} \to \{0,1\}$ is the AND function on $\frac{N}{B}$ bits on the Boolean basis. Our goal is to approximate the outer AND function and the inner h_i's carefully so that the total degree and weight can be bounded as we want.

For any subset $T \subseteq S \subseteq [N]$, define the indicator function $\mathbb{I}(\cdot; T, S)\colon \{-1,1\}^N \to \{0,1\}$ by

$$\mathbb{I}(y; T, S) = \prod_{j \in T} \frac{y_j + 1}{2} \prod_{j \in S \setminus T} \frac{1 - y_j}{2}, \tag{20}$$

so it is 1 if and only if y represents False on T and True on $S \setminus T$. Thus each h_i can be written using interpolation as

$$\begin{aligned} h_i(x,y) &= \sum_{T \subseteq S_i} \left(\bigwedge_{j \in T} f_j(x) \wedge \bigwedge_{j \in T} \neg y_j \wedge \bigwedge_{j \in S_i \setminus T} y_j \right) \\ &= \sum_{T \subseteq S_i} \left(\bigwedge_{j \in T} f_j(x) \cdot \mathbb{I}(y; T, S_i) \right). \end{aligned} \tag{21}$$

Now suppose we have a polynomial $p\colon \{0,1\}^{\frac{N}{B}} \to \mathbb{R}$ that ε_{out}-approximates the outer $\mathsf{AND}_{\frac{N}{B}}$ function within degree d_{out} and weight w_{out}, where ε_{out}, d_{out}, and w_{out} are parameters to be set later. We called this p the outer approximation. We can write it as

$$p(z) = \sum_{\substack{U \subseteq [\frac{N}{B}] \\ |U| \le d_{out}}} a_U \prod_{i \in U} z_i, \tag{22}$$

where $a_U \in \mathbb{R}$, and $\sum_U |a_U| = w_{out}$ by definition. Define $F''\colon \{-1,1\}^n \times \{-1,1\}^{2n} \to \mathbb{R}$ by substituting the outer $\mathsf{AND}_{\frac{N}{B}}$ function in F' by p: $F''(x,y) = p(h_1(x,y), \ldots, h_{\frac{N}{B}}(x,y))$. Since p is a point-wise approximation, we have

$$\|F'' - F'\|_\infty \le \|p - \mathsf{AND}_{\frac{N}{B}}\|_\infty = \varepsilon_{out}. \tag{23}$$

On the other hand, we can expand F'' as

$$\begin{aligned} F''(x,y) &= \sum_{\substack{U \subseteq [\frac{N}{B}] \\ |U| \le d_{out}}} a_U \prod_{i \in U} h_i(x,y) \\ &= \sum_{\substack{U \subseteq [\frac{N}{B}] \\ |U| \le d_{out}}} a_U \prod_{i \in U} \sum_{T \subseteq S_i} \left(\bigwedge_{j \in T} f_j(x) \cdot \mathbb{I}(y; T, S_i) \right) \\ &= \sum_{\substack{U \subseteq [\frac{N}{B}] \\ |U| \le d_{out}}} a_U \sum_{\substack{T\colon U \to \mathcal{P}([N]) \\ T(i) \subseteq S_i, \forall i \in U}} \prod_{i \in U} \left(\bigwedge_{j \in T(i)} f_j(x) \cdot \mathbb{I}(y; T(i), S_i) \right) \\ &= \sum_{\substack{U \subseteq [\frac{N}{B}] \\ |U| \le d_{out}}} a_U \sum_{\substack{T\colon U \to \mathcal{P}([N]) \\ T(i) \subseteq S_i, \forall i \in U}} \left(\prod_{i \in U} \bigwedge_{j \in T(i)} f_j(x) \right) \left(\prod_{i \in U} \mathbb{I}(y; T(i), S_i) \right) \\ &= \sum_{\substack{U \subseteq [\frac{N}{B}] \\ |U| \le d_{out}}} a_U \sum_{\substack{T\colon U \to \mathcal{P}([N]) \\ T(i) \subseteq S_i, \forall i \in U}} \left(\bigwedge_{j \in \mathsf{img}(T)} f_j(x) \cdot \mathbb{I}(y; \mathsf{img}(T), \cup_{i \in U} S_i) \right), \end{aligned} \tag{24}$$

where $\mathsf{img}(T)$ denotes the image of function $T\colon U \to \mathcal{P}([N])$ and $\mathcal{P}([N])$ is the powerset of $[N]$, the first step uses (22), the second step uses (21), the third step exchanges the product with the sum, the fourth step uses properties of multiplication, and the last step uses the fact that multiplication on the Boolean basis is equivalent to AND. It is important that we set the input basis of the outer approximation p (thus the output basis of $\bigwedge_{j\in\mathsf{img}(T)} f_j(x) \cdot \mathbb{I}(y; \mathsf{img}(T), \cup_{i\in U} S_i)$) to be the Boolean basis even though the input basis of the whole function is the Fourier basis; otherwise the last step doesn't hold.

Each $\bigwedge_{j\in\mathsf{img}(T)} f_j(x)$ is a 1-CNF (i.e. AND) since $f_j(x)$ is 1-CNF. Suppose we can approximate them by $\widetilde{\bigwedge_{j\in\mathsf{img}(T)}} f_j(x)$'s within error ε_{in}, degree d_{in}, and weight w_{in}, where ε_{in}, d_{in}, and w_{in} are parameters to be set later. We called these $\widetilde{\bigwedge_{j\in\mathsf{img}(T)}} f_j(x)$'s the inner approximations. Then we can define $\widetilde{F''}\colon \{-1,1\}^n \times \{-1,1\}^{2n} \to \mathbb{R}$ by

$$\widetilde{F''}(x,y) = \sum_{\substack{U\subseteq\left[\frac{N}{B}\right] \\ |U|\le d_{out}}} a_U \sum_{\substack{T\colon U\to\mathcal{P}([N]) \\ T(i)\subseteq S_i, \forall i\in U}} \left(\widetilde{\bigwedge_{j\in\mathsf{img}(T)}} f_j(x) \cdot \mathbb{I}(y; \mathsf{img}(T), \cup_{i\in U} S_i) \right). \tag{25}$$

Observe that for any $y \in \{-1,1\}^N$, for any $U \subseteq \left[\frac{N}{B}\right]$ with $|U| \le d_{out}$, there is only one $T\colon U \to [N]$ with $T(i) \subseteq S_i, \forall i \in U$ such that $\mathbb{I}(y; \mathsf{img}(T), \cup_{i\in U} S_i) = 1$: it is uniquely determined by the value of y on $\cup_{i\in U} S_i$; all other summands will varnish. Therefore we have

$$\|\widetilde{F''} - F''\|_\infty \le \sum_{\substack{U\subseteq\left[\frac{N}{B}\right] \\ |U|\le d_{out}}} a_U \varepsilon_{in} \le \sum_{\substack{U\subseteq\left[\frac{N}{B}\right] \\ |U|\le d_{out}}} |a_U| \varepsilon_{in} = w_{out}\varepsilon_{in}. \tag{26}$$

Hence if we set

$$\varepsilon_{out} = \frac{\varepsilon}{2}, \tag{27}$$

$$\varepsilon_{in} = \frac{\varepsilon}{2w_{out}}, \tag{28}$$

then $\widetilde{F''}$ ε-approximates F' since $\|\widetilde{F''} - F'\|_\infty \le \|\widetilde{F''} - F''\|_\infty + \|F'' - F'\|_\infty \le w_{out}\varepsilon_{in} + \varepsilon_{out} = \varepsilon$, where the first steps uses the triangular inequality, the second step uses (23) and (26), and the third step uses (27),(28). What remains to bound is the degree and weight of $\widetilde{F''}$.

Denote the degree of $\widetilde{F''}$ as d, and the weight of $\widetilde{F''}$ as w. Note that by (20) we have

$$\deg(\mathbb{I}(\cdot; \mathsf{img}(T), \cup_{i\in U} S_i)) = |\cup_{i\in U} S_i| = |U|B \le d_{out}B, \tag{29}$$

$$\|\!|\mathbb{I}(\cdot; \mathsf{img}(T), \cup_{i\in U} S_i)\|\!| \le 1. \tag{30}$$

For each U, there are at most $2^{|\cup_{i\in U} S_i|} \le 2^{d_{out}B}$ T's, since T satisfies that $T(i) \subseteq S_i, \forall i \in U$. Therefore we have

$$d \le d_{out}B + d_{in} \tag{31}$$

$$\log w \le \log \left(\sum_{\substack{U\subseteq\left[\frac{N}{B}\right] \\ |U|\le d_{out}}} a_U 2^{d_{out}B} w_{in} \right) \le \log w_{out} + d_{out}B + \log w_{in}, \tag{32}$$

where the first inequality comes from (25) and (29), and the second inequality follows from (25), (30), Claim 2.1, and the observation above. What remains is to set d_{in}, w_{in}, d_{out}, and w_{out} to get the desired bounds on d and w.

By assumption k satisfies $n^{\frac{2}{3}}(\log(1/\varepsilon))^{\frac{1}{3}} \leq k \leq n$. For convenience we will ignore the difference between N and n, and use them interchangeably, as they are the same up to a multiplicative factor of 2. Set

$$B = \frac{N}{k}, \tag{33}$$

$$k_{in} = k, \tag{34}$$

$$k_{out} = \sqrt{k \log(1/\varepsilon)}, \tag{35}$$

where k_{in} and k_{out} are numbers to be used later as the k's for the inner approximations and the outer approximation, respectively.

For the outer approximation, from (33) and (35) we have

$$\sqrt{\frac{N}{B} \log \frac{1}{\varepsilon_{out}}} = \sqrt{k \log(1/\varepsilon)} = k_{out} \leq k = \frac{N}{B},$$

where the first equality follows from (27) and ignoring the constant factor, the inequality comes from $k \geq N^{2/3}(\log(1/\varepsilon))^{1/3}$ and the fact that $\log(1/\varepsilon) \leq N$ (otherwise we can get all the bounds trivially). This means we can apply Theorem 4.1 with $\varepsilon = \varepsilon_{out}$, $k = k_{out}$, $n = \frac{N}{B}$, and $t = 0$ to get the outer approximating polynomial p with the following parameters:

$$d_{out} = O(k_{out}) = O\Big(\sqrt{k \log(1/\varepsilon)}\Big), \tag{36}$$

$$\log w_{out} = O\left(\frac{N}{Bk_{out}} \log \frac{1}{\varepsilon_{out}}\right) = O\Big(\sqrt{k \log(1/\varepsilon)}\Big), \tag{37}$$

using (27), (33), and (35).

For the inner approximations, we have $k_{in} = k \leq N$, and we also have

$$\sqrt{N \log \frac{1}{\varepsilon_{in}}} = O\left(\sqrt{N \log \frac{w_{out}}{\varepsilon}}\right) = O\left(\sqrt{N \sqrt{k \log(1/\varepsilon)}}\right) \leq O(k) = O(k_{in}), \tag{38}$$

where the first step uses (28), the second step uses (37), and the third step uses $k \geq N^{2/3}(\log(1/\varepsilon))^{1/3}$. Therefore we can invoke Theorem 4.1 with $\varepsilon = \varepsilon_{in}$, $k = k_{in}$, $n = N$, and $t = 0$ to get the inner approximations $\bigwedge_{j \in \mathsf{img}(T)} f_j(x)$ with the following parameters:

$$d_{in} = O(k_{in}) = O(k), \tag{39}$$

$$\log w_{in} = O\left(\frac{N}{k_{in}} \log \frac{1}{\varepsilon_{in}}\right) = O\left(\frac{N}{k} \log w_{out}\right) = O\left(\frac{N}{\sqrt{k}} \sqrt{\log(1/\varepsilon)}\right), \tag{40}$$

using (28), (34), and (37). Note that the input basis of $\bigwedge_{j \in \mathsf{img}(T)} f_j(x)$'s should be the Fourier basis, while the input basis of the polynomials given by Theorem 4.1 is the Boolean basis. We are implicitly using Lemma 2.3 here to change the input basis to the Fourier basis while keeping its degree and weight.

Finally, combining (31), (33), (36) and (39), we get

$$d = O\left(\frac{N}{k} \sqrt{k \log(1/\varepsilon)} + k\right) = O\left(\frac{N}{\sqrt{k}} \sqrt{\log(1/\varepsilon)} + k\right) = O(k),$$

where the last step follows from $k \geq N^{2/3}(\log(1/\varepsilon))^{1/3}$. Combing (32), (33), (36), (37), and (40), we get

$$\log w = O\left(\sqrt{k \log(1/\varepsilon)} + \frac{N}{k} \sqrt{k \log(1/\varepsilon)} + \frac{N}{\sqrt{k}} \sqrt{\log(1/\varepsilon)}\right) = O\left(\frac{N}{\sqrt{k}} \sqrt{\log(1/\varepsilon)}\right),$$

since $k \leq N$ implies $\frac{N}{\sqrt{k}} \sqrt{\log(1/\varepsilon)} \geq \sqrt{k \log(1/\varepsilon)}$. $\square$

Proof of Theorem 1.5. By induction on t: $t = 1$ is Theorem 4.1 for AND. Now assume the theorem holds for $(t-1)$-CNF, and we want to prove it for t-CNF. Similarly to the proof for $t = 2$, Equations (17)-(32) remain the same. The outer function is still $\mathsf{AND}_{\frac{N}{B}}$, while the inner functions $\bigwedge_{j\in\mathsf{img}(T)} f_j(x)$'s become $(t-1)$-CNFs. What remains is to set d_{in}, w_{in}, d_{out}, and w_{out} to get the desired bounds on d and w.

By assumption k satisfies $n^{\frac{t}{t+1}}(\log(1/\varepsilon))^{\frac{1}{t+1}} \le k \le n$. Set

$$B = \frac{N}{k^{2/t}},$$
$$k_{in} = k,$$
$$k_{out} = k^{\frac{1}{t}}(\log(1/\varepsilon))^{\frac{1}{t}}.$$

For the outer approximation, it's not hard to verify that $\sqrt{\frac{N}{B}\log\frac{1}{\varepsilon_{out}}} = k_{out} \le \frac{N}{B}$, so we can apply Theorem 4.1 to get

$$d_{out} = O(k_{out}) = O\Big(k^{\frac{1}{t}}(\log(1/\varepsilon))^{\frac{1}{t}}\Big),$$
$$\log w_{out} = O\bigg(\frac{N}{Bk_{out}}\log\frac{1}{\varepsilon_{out}}\bigg) = O\Big(k^{\frac{1}{t}}(\log(1/\varepsilon))^{\frac{t-1}{t}}\Big).$$

For the inner approximation, it's not hard to verify that $N^{\frac{t-1}{t}}\left(\log\frac{1}{\varepsilon_{in}}\right)^{\frac{1}{t}} \le k_{in} \le N$, so we can use the induction hypothesis for $(t-1)$-CNF to get

$$d_{in} \le c_{t-1}\cdot k_{in} = c_{t-1}\cdot k,$$
$$\log w_{in} \le c_{t-1}\cdot\frac{N}{k_{in}^{1/(t-1)}}\left(\log\frac{1}{\varepsilon_{in}}\right)^{\frac{1}{t-1}} \le c'_{t-1}\cdot\frac{N}{k^{1/t}}(\log(1/\varepsilon))^{\frac{1}{t}},$$

where c'_{t-1} is some constant depending on c_{t-1}.

Combining all these bounds, for some constant c_t (depending only on t) we get

$$d = O\bigg(\frac{N}{k^{1/t}}(\log(1/\varepsilon))^{\frac{1}{t}}\bigg) + c_{t-1}\cdot k \le c_t\cdot k,$$
$$\log w = O\Big(k^{\frac{1}{t}}(\log(1/\varepsilon))^{\frac{t-1}{t}}\Big) + c_{t-1}\cdot\frac{N}{k^{1/t}}(\log(1/\varepsilon))^{\frac{1}{t}} \le c_t\cdot\frac{N}{k^{1/t}}(\log(1/\varepsilon))^{\frac{1}{t}}.$$ □

Proof of Theorem 1.9. Use Theorem 1.5 and Theorem 2.5. □

7 PARITY and OR Have Large Weights over $\{0,1\}^n$

Claim 7.1. *For any $\varepsilon \in (0,1)$, the weight of polynomial f that ε-approximates $\mathsf{PARITY}\colon \{0,1\}^n \to \{-1,1\}$ is at least $(1-\varepsilon)2^n$.*

Proof. Suppose $f\colon \{0,1\}^n \to \mathbb{R}$ ε-approximates PARITY and minimizes the weight. On one hand, we have $\frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)\mathsf{PARITY}(x) \in [1-\varepsilon, 1+\varepsilon]$. On the other hand, write f as $f = \sum_{S\subseteq[n]} c_S \prod_{i\in S} x_i$, then

$$\sum_{x\in\{0,1\}^n} f(x)\mathsf{PARITY}(x) = \sum_{T\subseteq[n]}(-1)^{|T|}\sum_{S\subseteq T} c_S = \sum_{S\subseteq[n]} c_S \sum_{T\supseteq S}(-1)^{|T|} = (-1)^n c_{[n]},$$

where the last step comes from the fact that whenever $S \subsetneq [n]$ we have $\sum_{T\supseteq S}(-1)^{|T|} = 0$ as we can arrange such T's into matching pairs that has exactly opposite value of $(-1)^{|T|}$.

Therefore we have $\|f\| \ge |c_{[n]}| \ge (1-\varepsilon)2^n$. □

Claim 7.2. *For any fixed $\varepsilon \le \frac{1}{3}$, the weight of polynomial f that ε-approximates* $\mathsf{OR}\colon \{0,1\}^n \to \{0,1\}$ *is* $2^{\Omega(\sqrt{n})}$.

Proof. Let f be the polynomial that ε-approximates OR and minimizes the weight. Let $w = \|f\|$. By a sampling argument ([Gro97], c.f. [Zha14, CMS18]) we can get a polynomial $g\colon \{0,1\}^n \to \mathbb{R}$ such that g $\frac{1}{3}$-approximates OR and g has $O(w^2 n)$ monomials. Now define $h\colon \{0,1\}^n \times \{0,1\}^n \to \mathbb{R}$ by $h(x_1 \cdots x_n, y_1 \cdots y_n) = 1 - g(x_1 y_1, \ldots, x_n y_n)$, then h $\frac{1}{3}$-approximates the set disjointness function DISJ, where $x_1 \cdots x_n$ and $y_1 \cdots y_n$ are interpreted as indicators of two sets $X, Y \subseteq [n]$ respectively and $\mathsf{DISJ}(X,Y) = 1$ iff $X \cap Y = \varnothing$. Now h also has $O(w^2 n)$ monomials. By Theorem 8 in [BdW01] we get $w = 2^{\Omega(\sqrt{n})}$. $\square$

8 Discussion and Open Problems

An obvious open question is to prove such degree-weight tradeoffs for more functions. A central function in the area is Surjectivity. For this it would suffice to have a polynomial approximating OR over the domain $\{-1,1\}^m_{\le n}$, in which the Hamming of the input is restricted to be at most n. [She18] showed that the degree of such a polynomial depends on n instead of m as if we are working over $\{-1,1\}^n$. It is natural to ask if the same holds for weight. The answer is negative. To show this, note that the proof of Theorem 1.11 actually gives us (k,δ)-indistinguishable distributions with bounded Hamming weight that don't fool OR. In particular, for any ε, n, m, k satisfying $\frac{c'}{16}\sqrt{m \log(1/\varepsilon)} \le k \le \frac{c'}{16} m$ and $n = \frac{c'^2}{16^2}\frac{m^2}{k^2}\log(1/\varepsilon)$, we have distributions that are $\left(k, 2^{-\Omega(m\log(1/\varepsilon)/k)}\right)$-indistinguishable on $\{-1,1\}^m_{\le n}$ but cannot ε-fool OR. For $m > n$, we have $2^{-\Omega(m\log(1/\varepsilon)/k)} < 2^{-\Omega(n\log(1/\varepsilon)/k)}$ for fixed k and ε. This also means that we need other methods for Surjectivity.

Another open problem is to show tight degree-weight tradeoffs for OR on $\{-1,1\}^m_{\le n}$. Chandrasekaran et al. [CTUW14, Corollary 5.2] proved that it requires weight roughly at least $\left(\frac{m}{k\sqrt{n}}\right)^{\sqrt{n}}$ for constant ε, so when $k\sqrt{n} \le (1-\Omega(1))m$ it requires $2^{\Omega(\sqrt{n})}$.

Another open problem is the approximate weight of parity when k is very close to n. We know that when $k = n$ it is 1, and we show that when when $k = 0.001n$ it requires $2^{\Omega(n)}$. Our upper bound in Theorem 1.3 only tells us that $w = 2^{O(n)}$ for such k. What happens in between? Can we get a better upper bound?

We also lack a matching "does not fool" result for t-CNF as tight approximate degree and weight are not known even for 2-CNF (without promise on the input). The open problem here is to prove lower bounds matching our results for t-CNF.

Acknowledgment

The authors thank Lijie Chen for showing the idea for proving Claim 7.2, Chin Ho Lee for pointing out the paper [OZ18], and Justin Thaler for pointing out the line of works [STT12, CTUW14, BT15]. The authors also thank Mark Bun and Avishay Tal for useful discussions. Part of this work was done during the authors' visit to the Simons Institute for the Theory of Computing.

References

[AGM03] Noga Alon, Oded Goldreich, and Yishay Mansour. Almost k-wise independence versus k-wise independence. *Inf. Process. Lett.*, 88(3):107–110, 2003.

[BdW01] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual Conference on Computational Complexity*, CCC '01, pages 120–, Washington, DC, USA, 2001. IEEE Computer Society.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *Int. Cryptology Conf. (CRYPTO)*, 2016.

[BMTW19] Andrej Bogdanov, Nikhil S. Mande, Justin Thaler, and Christopher Williamson. Approximate degree, secret sharings, and concentration phenomena. Manuscript, 2019.

[BT13] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. In *Proceedings of the 40th International Conference on Automata, Languages, and Programming - Volume Part I*, ICALP'13, pages 303–314, Berlin, Heidelberg, 2013. Springer-Verlag.

[BT15] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In Magnús M. Halldórsson, Kazuo Iwama, Naoki Kobayashi, and Bettina Speckmann, editors, *Automata, Languages, and Programming*, pages 268–280, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.

[BW17] Andrej Bogdanov and Christopher Williamson. Approximate bounded indistinguishability. In Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl, editors, *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, volume 80 of *LIPIcs*, pages 53:1–53:11. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

[Che98] E.W. Cheney. *Introduction to Approximation Theory*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 1998.

[CMS18] Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:176, 2018.

[CTUW14] Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. In *Innovations in Theoretical Computer Science, ITCS'14, Princeton, NJ, USA, January 12-14, 2014*, pages 387–402, 2014.

[dW08] Ronald de Wolf. A note on quantum algorithms and the minimal degree of ε-error polynomials for symmetric functions. *Quantum Info. Comput.*, 8(10):943–950, November 2008.

[Gro97] Vince Grolmusz. On the power of circuits with gates of low l 1 norms. *Theoretical Computer Science*, 188(1-2):117–128, 1997.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[KS06] Adam R. Klivans and Rocco A. Servedio. Toward attribute efficient learning of decision lists and parities. *Journal of Machine Learning Research*, 7:587–602, 2006.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993.

[NS94] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OZ18] Ryan O'Donnell and Yu Zhao. On closeness to k-wise uniformity. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2018, August 20-22, 2018 - Princeton, NJ, USA*, pages 54:1–54:19, 2018.

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing*, STOC '92, pages 468–474, New York, NY, USA, 1992. ACM.

[She08] Alexander A. Sherstov. Approximate inclusion-exclusion for arbitrary symmetric functions. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 112–123, June 2008.

[She13] Alexander A Sherstov. Approximating the and-or tree. *Theory of Computing*, 9(20):653–663, 2013.

[She18] Alexander A. Sherstov. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 311–324, New York, NY, USA, 2018. ACM.

[STT12] Rocco A. Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland*, pages 14.1–14.19, 2012.

[Vio17] Emanuele Viola. Special topics in complexity theory. Lecture notes of the class taught at Northeastern University. Available at http://www.ccs.neu.edu/home/viola/classes/spepf17.html, 2017.

[Zha14] Shengyu Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1878–1885, 2014.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il