View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Dagstuhl Research Online Publication Server

Approximate Bounded Indistinguishability

Andrej Bogdanov[1] and Christopher Williamson[2]

1 **Department of Computer Science and Engineering and Institute for Theoretical Computer Science and Communications, Chinese University of Hong Kong, Hong Kong, China**
andrejb@cse.cuhk.edu.hk

2 **Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong, China**
chris@cse.cuhk.edu.hk

Abstract

Two distributions over n-bit strings are (k, δ)-wise indistinguishable if no statistical test that observes k of the n bits can tell the two distributions apart with advantage better than δ. Motivated by secret sharing and cryptographic leakage resilience, we study the existence of pairs of distributions that are (k, δ)-wise indistinguishable, but can be distinguished by some function f of suitably low complexity. We prove bounds tight up to constants when f is the OR function, and tight up to logarithmic factors when f is a read-once uniform AND $\circ$ OR formula, extending previous works that address the perfect indistinguishability case $\delta = 0$.

We also give an elementary proof of the following result in approximation theory: If p is a univariate degree-k polynomial such that $|p(x)| \le 1$ for all $|x| \le 1$ and $p(1) = 1$, then $\hat{\ell}_1(p) \ge 2^{\Omega(p'(1)/k)}$, where $\hat{\ell}_1(p)$ is the sum of the absolute values of p's coefficients. A more general statement was proved by Servedio, Tan, and Thaler (2012) using complex-analytic methods.

As a secondary contribution, we derive new threshold weight lower bounds for bounded depth AND-OR formulas.

1998 ACM Subject Classification F.0 [Theory of Computation] General

Keywords and phrases pseudorandomness, polynomial approximation, secret sharing

Digital Object Identifier 10.4230/LIPIcs.ICALP.2017.53

1 Introduction

Two random variables X and Y over $\{0,1\}^n$ are (locally) (k, δ)*-wise indistinguishable* if for all subsets $S \subseteq \{1, \dots, n\}$ of size k, the induced marginal distributions $(X_i \colon i \in S)$ and $(Y_i \colon i \in S)$ are within statistical distance δ. We say function $f \colon \{0,1\}^n \to \{0,1\}$ *reconstructs from* the pair (X, Y) with error at most ε if $\mathbb{E}[f(X)] - \mathbb{E}[f(Y)] \ge 1 - \varepsilon$. In this work we investigate conditions under which a suitable f can reconstruct from some pair of locally indistinguishable distributions.

The parity function on n bits provides an extreme example of this phenomenon: The uniform distributions over $\{0,1\}^n$ conditioned on the parity of all the bits taking value zero and one, respectively, are $(n-1, 0)$-wise indistinguishable, but parity reconstructs from them perfectly. Our focus will be on reconstruction functions that have representations of constant depth and size polynomial in n (i.e., in the class AC^0). Functions in this class exclude large parities [6, 18, 7], and are in fact strongly uncorrelated with them under the uniform distribution [8].

© Andrej Bogdanov and Christopher Williamson;
licensed under Creative Commons License CC-BY

44th International Colloquium on Automata, Languages, and Programming (ICALP 2017).
Editors: Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl; Article No. 53; pp. 53:1–53:11
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The OR function

The OR function on n inputs is arguably the most elementary example of this type. We prove matching upper and lower bounds reconstruction by the OR function with respect to (k, δ)-wise indistinguishable distributions. In the range where the reconstruction error is constant, we obtain the following results.

▶ **Theorem 1.** *For every n and $k \geq 5\sqrt{n}$ and every pair of $(k, 2^{-O(n/k)})$-wise indistinguishable distributions X, Y over $\{0,1\}^n$, $\Pr[\mathrm{OR}(X) = 1] - \Pr[\mathrm{OR}(Y) = 1] \leq 1/3$.*

▶ **Theorem 2.** *For every n and $k \leq n/2$ there exists a pair of $(k, 2^{-\Omega(n/k)})$-wise indistinguishable distributions X, Y over $\{0,1\}^n$ such that $\Pr[\mathrm{OR}(X) = 1] - \Pr[\mathrm{OR}(Y) = 1] \geq 2/3$.*

The distributions X and Y in Theorem 2 are uniformly sampleable by circuit families of constant depth and size polynomial in n.

These results extend our previous work with Ishai and Viola [3], which only considered perfect bounded indistinguishability, i.e., the case $\delta = 0$. It is shown there that $(2\sqrt{n}, 0)$-wise and $(\sqrt{n}/2, 0)$-wise indistinguishability yield the conclusions of Theorems 1 and 2, respectively. Those proofs make use of results about the approximability of the OR function by real-valued polynomials of Linial and Nisan [10] and Nisan and Szegedy [13].

Our work with Ishai and Viola also explains the relevance of bounded indistinguishability to the computational complexity of secret sharing and cryptographic leakage resilience. In the context of secret sharing, (k, δ)-wise indistinguishability postulates that the joint view of any k parties can predict the secret with advantage at most δ. In a *visual secret sharing* scheme [12], the secret to be shared is a pixel, each of n parties receives a transparency and the pixel is recovered by superimposing the transparencies. The procedure can be applied independently to every pixel in an image. The *contrast* of the scheme is the fraction of pixels that are reconstructed correctly.

Theorems 1 and 2 describe the best possible tradeoff between the size and the reconstruction advantage of the adversarial coalition for visual secret sharing schemes [12] with constant contrast. Theorem 2' in Section 3 is a refinement of Theorem 2 that specifies the dependence of the indistinguishability parameters on the contrast.

Linial and Nisan [10] studied an incomparable notion of δ-indistinguishability in which the family of statistical tests consists of ANDs over arbitrary subsets of the input bits. They proved analogues of Theorems 1 and 2 for $\delta \geq 2^{-c\sqrt{n}}$ and $\delta \leq 2^{-c\sqrt{n}\log n}$, respectively (for suitable constants c and C).

A consequence in approximation theory

A. A. Markov showed that among all real-valued univariate degree k polynomials p such that $|p(x)| \leq 1$ for all $|x| \leq 1$, the derivative $p'(1)$ is maximized by the Chebyshev polynomial T_k of degree k with $T_k'(1) = k^2$. The weight $\hat{\ell}_1(T_k)$ of the Chebyshev polynomial, defined as the sum of the absolute values of its coefficients, is exponential in k. At the other end of the spectrum, the polynomial x^k has weight 1 and derivative k at 1. Interpolating between the two, the degree-k polynomial $T_r(x^{k/r})$ has weight exponential in r and derivative rk at 1 whenever r divides k. With one small additional hypothesis, we prove that this is the best possible up to the constant in the exponent:

▶ **Theorem 3.** *There exists a constant $C > 0$ such that if $p \colon \mathbb{R} \to \mathbb{R}$ is a degree-k polynomial with $|p(x)| \leq 1$ for all $|x| \leq 1$ and $p(1) = 1$ then $p'(1) \leq Ck \log \hat{\ell}_1(p)$.*

The polynomials $p(x) = T_r(x^{\lfloor k/r \rfloor})$ certify that the bound is tight up to the constant factor.

Servedio, Tan, and Thaler [14] proved a more general form of Theorem 3: Under the same assumptions, they showed that $\max_{x\in[-1,1]}|p'(x)| \leq Ck \log \hat{\ell}_1(p)$. (Their bound is stated in a slightly weaker form.) Our proof is based on elementary counting, a large deviation bound, and some basic calculus, while Servedio et al.'s makes use of Hadamard's Three Circle Theorem from complex analysis.

Perfect and almost-perfect reconstruction

In the setting of secret sharing, reconstruction errors are undesirable as they entail possible loss of information in the sharing phase. There, perfect reconstruction is a desirable feature. The OR function is not up to the task (for non-trivial parameter settings): It is easily observed that perfect reconstruction by OR requires $(1, 1/n)$-wise indistinguishability of the underlying distributions X and Y. We show, however, that read-once CNFs and higher depth AND-OR trees can perfectly reconstruct from distributions of approximate bounded indistinguishability:

▶ **Theorem 4.** *For any fixed d, and for all n and k, there exists a pair of $(k, 2^{-\Omega((n/k)^{1-1/d})})$-wise indistinguishable distributions that can be perfectly reconstructed by the depth-d AND-OR tree with top fan-in $(n/k)^{1/d}$, middle fan-ins $(n/k)^{2/d}$, and bottom fan-in $\frac{k^{(2d-3)/d}}{n^{(d-3)/d}}$.*

Setting $d = 2$ results in the following corollary:

▶ **Corollary 5.** *For all n and k there exists a pair of $(k, 2^{-\Omega((n/k)^{1/2})})$-wise indistinguishable distributions that can be perfectly reconstructed by the function $\text{AND}_{(n/k)^{1/2}} \circ \text{OR}_{(nk)^{1/2}}$.*

Here, $\text{AND}_{n/m} \circ \text{OR}_m$ is a read-once monotone CNF with fan-in m at the bottom OR gates and fan-in n/m at the top AND gate. We prove that Corollary 5 is essentially tight for read-once CNFs. In Proposition 13, however, we show that *almost* perfect reconstruction by functions of this type is possible with substantially better parameters.

Threshold weight

The degree-k threshold weight of a function $f : \{0, 1\}^n \to \{0, 1\}$ is the minimum weight of a nonzero degree-k polynomial p with integer coefficients such that $p(x)f(x) \geq 0$ for all x. Beigel [1] and Servedio et al. [14] construct a length-n decision list that requires degree-k threshold weight $2^{\Omega(\sqrt{n/k})}$. Bun and Thaler [5] give a read-once DNF over n variables that requires the same degree-k threshold weight, and construct a polynomial size AND-OR circuit of depth 3 that requires degree-k threshold weight $2^{\Omega(n/k^{3/2})}$. Sherstov [16] constructed a depth 4 circuit of polynomial size that requires threshold weight $2^{\Omega(\sqrt{n})}$ for all k. Our methods yield the following incomparable bound:

▶ **Corollary 6.** *The depth-d AND-OR tree given in the statement of Theorem 4 requires degree-k threshold weight $2^{\Omega((n/k)^{1-1/d})}$.*

In Section 6 we show that this result implies the degree-independent threshold weight lower bound for formulas of Sherstov [15].

Our proofs

Impossibility of reconstruction by OR from $(2\sqrt{n}, 0)$-wise indistinguishable distributions follows easily from the existence of a degree $2\sqrt{n}$ polynomial p that approximates the OR function pointwise: The distinguishing advantage of the OR function can be at most the pointwise approximation error. The relaxed assumption of approximate local indistinguishability introduces an additional error term proportional to the weight of the approximating polynomial. Lemma 7 recasts the problem in the univariate setting, following previous works. To prove Theorem 1 we instantiate Lemma 7 with a polynomial of the form $T_r(x^{k/r})$ for a suitable choice of r. The relevant properties of this polynomial are that it is bounded on $[-1, 1]$, has weight $2^{O(r)}$ and has value $\Omega(kr/n)$ at $x = 1 + 1/2n$.

In the setting of perfect bounded indistinguishability, the maximum distinguishing advantage of OR with respect to $(k, 0)$-wise indistinguishable distributions *equals* the minimum error that a k-approximating polynomial for OR must have by linear programming duality. Thus, high approximate degree readily implies the existence of locally indistinguishable distributions that can be told apart by the OR function. In the approximate setting this duality is not preserved: While existence of low-degree, low-weight approximate polynomials implies hardness of reconstruction, it is not at all clear that the converse should hold. To prove a converse to Theorem 1 we instead resort to combinatorial means.

We prove Theorem 2 by a reduction to the case of perfect bounded indistinguishability. Previous works give the existence of $(\Omega(\sqrt{\varepsilon n}), 0)$-wise indistinguishable distributions X and Y such that $\Pr[\mathrm{OR}(X) = 1] - \Pr[\mathrm{OR}(Y) = 1] \geq 1 - \varepsilon$. (As observed in [3], the dual polynomials of Špalek [17] and Bun and Thaler [4] show that such distributions can be sampled by uniform constant-depth, polynomial-size circuit families.) We consider the following distributions X' and Y' over $\{0, 1\}^N$, where $N \geq n$: Sample n coordinates of $\{1, \ldots, N\}$ uniformly at random, embed a sample of X and Y, respectively, in these coordinates, and set all the other entries of X and Y to zero. The distinguishing advantage of OR is not affected by this transformation. On the other hand, any "local view" of size K in $\{1, \ldots, N\}$ expects to observe $K \cdot n/N$ of the embedded samples. Provided this number is sufficiently smaller than the indistinguishability parameter k, by a large deviation bound this local view can reconstruct from the distributions X and Y only with small probability.

Since Theorem 2 proves the optimality of Theorem 1, it follows that no choice of univariate polynomial p that is bounded on $[-1, 1]$ can have significantly larger value at $x = 1 + 1/2n$ than the polynomial $T_r(x^{k/r})$ among all those of weight $2^{O(r)}$. By the mean value theorem, there must exist some x in $[1, 1 + 1/2n]$ such that $p'(x)$ is at most $O(kr)$. Proving Theorem 3 requires showing that $p'(1) = O(kr)$. Our proof of Theorem 3 in Section 4 bounds the rate of change of $p'(x)$ around $x = 1$ as a function of the weight of p. Since this norm is small, for a suitable choice of n we can conclude from a somewhat delicate calculation that $p'(1) = O(kr)$ and prove Theorem 3.

The proof of Theorem 4 extends the reduction to perfect bounded indistinguishability to the setting of higher depth trees. This was studied in [3] and is a consequence of the seminal lower bound of Minsky and Papert [11]. Optimality is proved by an explicit construction of low-weight approximating polynomials as in Theorem 1. In the near-perfect setting, the requisite lower bound on approximate degree was obtained by Bun and Thaler [5] (extending Beigel [2]).

2 Optimality of distributions: Proof of Theorem 1

We first present a limitation of the OR function's ability to reconstruct from distributions that are almost k-wise indistinguishable. This is completed in two steps. First, we reduce the question to one of polynomials. Specifically, we demonstrate that the existence of certain polynomials that approximate OR allow us to upper bound the ability of OR to reconstruct from these distributions, where the degree and size of coefficients of the polynomial relate to the indistinguishability parameters. The second step is to construct these polynomials and calculate the indistinguishability parameters as a function of their degree and $\hat{\ell}_1$ norm.

2.1 Indistinguishability from approximating polynomials

▶ **Lemma 7.** *Assume $p\colon \mathbb{R} \to \mathbb{R}$ is a degree-k polynomial such that $|p(x)| \le 1$ for all x such that $|x| \le 1$ and $j(p,n) = p(1+2/n) - 1 \ge 0$. For all pairs of (k,δ)-wise indistinguishable distributions X and Y over $\{-1,1\}^n$,*

$$\mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] \le \frac{2}{2+j(p,n)} + e^2 \cdot \delta \cdot \hat{\ell}_1(p).$$

Here, we assume a representation in which OR evaluates to zero if any of its inputs are -1 and to 1 otherwise.

In the case of perfect indistinguishability (i.e., $\delta = 0$), such approximations of the OR function by polynomials already appear in the work of Linial and Nisan. Lemma 7 shows that the presence of local error δ introduces an additional term proportional to the weight of the approximation polynomial. The term $j(p,n)$ can be improved slightly to $p(1+2/(n-1))-1$.

In the proof of Lemma 7 we will use the following facts, which themselves are proven after the lemma.

▶ **Fact 8.** *Let p be a degree-k univariate polynomial and $p^*(x) = p(ax+b)$, where $|a|+|b| \ge 1$. Then $\hat{\ell}_1(p^*) \le (|a|+|b|)^k \hat{\ell}_1(p)$.*

▶ **Fact 9.** *If p is a univariate polynomial and $\mathbf{p}(x_1,\dots,x_n) = p((x_1+\dots+x_n)/n)$ then $\hat{\ell}_1(\mathbf{p}) \le \hat{\ell}_1(p)$.*

Proof of Lemma 7. Let

$$p^*(x) = \gamma \cdot p\Big(x + \frac{2}{n}\Big) \quad \text{where} \quad \gamma = \frac{1}{2+j(p,n)}.$$

Then $|p^*(x)| \le \gamma$ for $x \in [-1, 1-\frac{2}{n}]$ and $p^*(1) = 1-\gamma$. The multivariate polynomial

$$\mathbf{p}^*(x_1,\dots,x_n) = p^*\Big(\frac{x_1+\dots+x_n}{n}\Big)$$

satisfies $|\mathbf{p}^*(\mathbf{x}) - \mathrm{OR}(\mathbf{x})| \le \gamma$ for all $\mathbf{x} \in \{-1,1\}^n$. Expanding $\mathbf{p}^*$ in the Fourier basis, we can write $\mathbf{p}^*(\mathbf{x}) = \sum_{|S|\le k} \hat{\mathbf{p}}^*_S \chi_S(\mathbf{x})$, where $\chi_S(\mathbf{x}) = \prod_{i\in S} x_i$. Then

$$\begin{aligned}\mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] &\le \mathbb{E}[\mathbf{p}^*(X)+\gamma] - \mathbb{E}[\mathbf{p}^*(Y)-\gamma]\\ &\le 2\gamma + \mathbb{E}[\mathbf{p}^*(X)] - \mathbb{E}[\mathbf{p}^*(Y)]\\ &= 2\gamma + \sum\nolimits_{|S|\le k} \hat{\mathbf{p}}^*_S \cdot (\mathbb{E}[\chi_S(X)] - \mathbb{E}[\chi_S(Y)])\\ &\le 2\gamma + \sum\nolimits_S |\hat{\mathbf{p}}^*_S| \cdot 2\delta\\ &= 2\gamma + 2\delta \cdot \hat{\ell}_1(\mathbf{p}^*).\end{aligned}$$

The second to last step holds because χ_S is a k-local distinguisher with range $\{-1,1\}$, so its distinguishing advantage is at most 2δ. From Facts 8 and 9 it follows that

$$\hat{\ell}_1(\mathbf{p}^*) \le \frac{1}{2+j(p,n)} \cdot \Big(1+\frac{2}{n}\Big)^k \cdot \hat{\ell}_1(p) \le \frac{1}{2} \cdot \Big(\frac{n+2}{n}\Big)^k \cdot \hat{\ell}_1(p) \le \frac{e^2}{2} \cdot \hat{\ell}_1(p).$$

for $k < n$ as desired. ◀

Proof of Fact 8. If all the coefficients of a polynomial q are positive then $\hat{\ell}_1(q) = q(1)$. For $p(x) = \sum_{i=0}^{k} c_i x^i$ let $\tilde{p}(x) = \sum_{i=0}^{k} |c_i| x^i$. Then

$$\hat{\ell}_1(\tilde{p}(|a|x + |b|)) = \tilde{p}(|a| + |b|) \leq (|a| + |b|)^k \tilde{p}(1) = (|a| + |b|)^k \hat{\ell}_1(p).$$

The coefficients of $\tilde{p}(|a|x + |b|)$ dominate those of $p(ax + b)$, so we can conclude that $\hat{\ell}_1(p(ax + b)) \leq \hat{\ell}_1(\tilde{p}(|a|x + |b|)) \leq (|a| + |b|)^k \hat{\ell}_1(p)$. ◀

Proof of Fact 9. If $p(x) = \sum_{i=0}^{k} c_i x^i$ then

$$\hat{\ell}_1(\mathbf{p}) \leq \sum_{i=0}^{d} \frac{|c_i|}{n^i} \hat{\ell}_1\big((x_1 + \ldots + x_n)^i\big) \leq \sum_{i=0}^{d} \frac{|c_i|}{n^i} \cdot n^i = \hat{\ell}_1(p),$$

where the second to last step holds because all of the coefficients in $(x_1 + \cdots + x_n)^i$ are nonnegative, so the weight is $(1 + \cdots + 1)^i = n^i$. ◀

2.2 Construction of approximating polynomials

Our approximating polynomials will take the form of a Chebyshev polynomial evaluated at an appropriately chosen monomial. We note that Servedio, Tan, and Thaler in [14] also use polynomials of this form to give a degree-weight tradeoff of polynomials approximating the OR function.

For the proof of Theorem 1, we set $p(x) = T_r(x^d)$, where $d = \lfloor k^2/20n \rfloor \geq 1$, $r = \lfloor k/d \rfloor$, and T_r is the Chebyshev polynomial of degree r.

The Chebyshev polynomials satisfy (1) $|T_r(x)| \leq 1$ for all $x \in [-1, 1]$, (2) $T_r(1) = 1$, (3) $T_r'(1) = r^2$, (4) $T_r''(x) \geq 0$ for all $x \geq 1$, and (5) $\hat{\ell}_1(T_r) \leq 2^{2r}$. The first four properties are well-known; we provide a short proof of the fifth.

Proof of Property 5. We use an alternate definition of the Chebyshev polynomial: $T_r(x) = \frac{r}{2} \sum_{i=0}^{r/2} \frac{(-1)^i}{r-i} \binom{r-i}{i} 2^{r-2i} x^{r-2i}$. Thus, we have:

$$\hat{\ell}_1(T_r) = \frac{r}{2} \sum_{i=0}^{r/2} \frac{1}{r-i} \binom{r-i}{i} 2^{r-2i} \leq \sum_{i=0}^{r/2} \binom{r-i}{i} 2^{r-2i} \leq 2^r \sum_{i=0}^{r/2} \binom{r}{i} \leq 2^{2r}.$$ ◀

From properties (2), (3), and (4) it follows that

$$j(p, n) = p(1 + 2/n) - 1 \geq p'(1) \cdot \frac{2}{n} = \frac{2dr^2}{n} \geq \frac{k^2}{2dn} \geq 10.$$

The second to last inequality holds since for our choice of parameters $d \leq k$, so $r \geq 1$, and therefore $r \geq k/2d$.

By property (5), $\hat{\ell}_1(p) = 2^{2r} \leq 2^{2k/d}$. We now show this is at most $2^{80n/k}$. When $k^2 \geq 20n$, $d \geq k^2/40n$ and so $\hat{\ell}_1(p) \leq 2^{80n/k}$. Otherwise, $d = 1$, and $\hat{\ell}_1(p)$ is at most $2^{2k} \leq 2^{40n/k}$.

Finally, by property (1) $|p(x)| \leq 1$ for $|x| \leq 1$. By Lemma 7,

$$\mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] \leq \frac{2}{2 + 10} + e^2 \delta \cdot 2^{80n/k} \leq \frac{1}{3}$$

as long as $\delta \leq 2^{-80n/k}/6e^2$, proving Theorem 1.

3 Construction of distributions: Proof of Theorem 2

We reduce the existence of (k, ϵ)-indistinguishable distributions that can be reconstructed by the OR function to the analogous question for distributions of perfect bounded indistinguishability:

▶ **Lemma 10.** *For every $\varepsilon, N, K \leq N/2$ and $n \leq \epsilon N^2/121K^2$ the following holds. Assume there exist $(\sqrt{\epsilon n}, 0)$-wise indistinguishable distributions X, Y over $\{0,1\}^n$. Then there exist distributions X', Y' over $\{0,1\}^N$ such that $\mathbb{E}[\mathrm{OR}(X')] = \mathbb{E}[\mathrm{OR}(X)]$, $\mathbb{E}[\mathrm{OR}(Y')] = \mathbb{E}[\mathrm{OR}(Y)]$, and X', Y' are $(K, 2^{-\Omega(\epsilon N/K)})$-wise indistinguishable.*

Proof. To sample from X' (resp. Y'), first select a random set of n "active" indices among the N choices. Then, sample a string from X (resp. Y) and fill in the n indices with the sampled bits. Fill in the remaining $N - n$ places with 0s. This process does not change the chance that OR accepts a string, so the reconstruction error remains the same.

We now need to check that X', Y' are $(K, 2^{-\Omega(\epsilon N/K)})$-wise indistinguishable. Let S be any subset of $\{1, \dots, N\}$ of size K and E be the event that at most k of the active indices fall in S. Conditioned on E, the projections of X' and Y' on S contain at most k bits from X and Y, respectively, and are therefore perfectly indistinguishable. Therefore the distinguishing advantage of any test $T\colon \{0,1\}^S \to \{0,1\}$ can be at most the probability that E does not occur.

To upper bound this probability, we take a union bound over all possible $\binom{K}{k}$ subsets of k active indices in S. For each such set, there is a probability of n/N that the first index is active, a probability of $(n-1)/(N-1)$ that the second index is active conditioned on the first one, and so on, obtaining:

$$\begin{aligned}\Pr[\text{there are at least } k \text{ active indices in } S] &\leq \binom{K}{k} \cdot \frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \ldots \cdot \frac{n-k+1}{N-k+1} \\ &\leq \left(\frac{eK}{k} \cdot \frac{n}{N-k+1}\right)^k.\end{aligned}$$

Since $K \leq N/2$, we have that $k \leq N/2$ and $1/(N-k+1) \leq 2/N$. Plugging these estimates in, we conclude that the distinguishing advantage is at most $(2eKn/kN)^k$. We now set our parameters so that $\sqrt{\epsilon n} = k = 11nK/N$, implying that $(2eKn/kN)^k$ is upper bounded by $2^{-k} = 2^{-\Omega(nK/N)} = 2^{-\Omega(\epsilon N/K)}$. ◀

Now that we have reduced the problem of finding $(K, 2^{-\Omega(\epsilon N/K)})$-wise indistinguishable distributions to the one of finding $(\sqrt{\epsilon n}, 0)$-wise indistinguishable ones, for some specific n, we are ready to prove the following refinement of Theorem 2, which will be needed for the proof of Theorem 3.

▶ **Theorem 2'.** *For every ε, N, and $K \leq N/2$ there exists a pair of $(K, 2^{-\Omega(\epsilon N/K)})$-wise indistinguishable distributions X', Y' over $\{0,1\}^N$ such that $\mathbb{E}[\mathrm{OR}(X')] - \mathbb{E}[\mathrm{OR}(Y')] \geq 1 - \epsilon$.*

Proof. Corollary 2.2 in [3] shows the existence of $(\sqrt{\epsilon n}, 0)$-wise indistinguishable distributions X, Y over $\{0,1\}^n$ such that $\mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] = 1 - O(\epsilon)$. If $K \leq \sqrt{\epsilon N}/11$ the theorem follows directly from this Corollary. Otherwise, we apply Lemma 10 with $n = \lfloor \epsilon N^2/121K^2 \rfloor$ to X, Y and obtain the desired conclusion. ◀

4 Proof of Theorem 3

To prove Theorem 3 we reason as follows. Suppose there is a polynomial p of degree k such that $|p(x)| \leq 1$ for $|x| \leq 1$ and $p(1) = 1$. For $\varepsilon = p(1 + 2/n) - 1$, Theorem 2' and Lemma 7 together imply that

$$1 - \frac{\varepsilon}{6} \leq \mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] \leq \frac{2}{2+\varepsilon} + 2^{-\Omega(\varepsilon n/k)} \cdot \hat{\ell}_1(p),$$

from where we can conclude that $\hat{\ell}_1(p) \geq \Omega(\varepsilon) \cdot 2^{\Omega(\varepsilon n/k)}$, provided $\varepsilon \leq 1$. If the leading $\Omega(\varepsilon)$ term could be ignored, we would obtain Theorem 3 by taking the limit of the right-hand side as n goes to infinity and $\varepsilon n/2$ approaches $p'(1)$.

To account for the $\Omega(\varepsilon)$ term, we work with a carefully chosen, finite value of n. Our choice of n is sufficiently large so that the term $2^{\Omega(\varepsilon n/k)}$ dominates the term $\Omega(\varepsilon)$ in the expression lower bounding $\hat{\ell}_1(p)$, but sufficiently small so that $\varepsilon n/2$ is still lower bounded by $\Omega(p'(1))$. If n was a function of k only, this would be impossible as the value $\varepsilon = p(1+2/n)-1$ could even be negative. Our choice of n depends on the polynomial p itself via the parameters $\hat{\ell}_1(p)$ and $p'(1)$.

This description assumed that ε was at most one (or bounded by some fixed constant). The case of large ε can be handled along the same lines and is in fact technically easier.

Proof of Theorem 3. Let $p\colon \mathbb{R} \to \mathbb{R}$ be a degree-k polynomial such that $|p(x)| \leq 1$ for all $|x| \leq 1$ and $p(1) = 1$. Let $\varepsilon = p(1 + 2/n) - 1$ for $n = 4k^2\hat{\ell}_1(p)/p'(1)$. Expanding $p(1 + 2/n)$ around 1 we obtain

$$p(1 + 2/n) = p(1) + \frac{p'(1)}{n/2} + \sum_{i \geq 2} \frac{1}{(n/2)^i} \cdot \frac{p^{(i)}(1)}{i!}$$

where $p^{(i)}(1)$ is the i-th derivative of p at 1. Since $p(1) = 1$ it follows that

$$\varepsilon = p(1 + 2/n) - p(1) \geq \frac{p'(1)}{n/2} - \sum_{i \geq 2} \frac{1}{(n/2)^i} \cdot \frac{|p^{(i)}(1)|}{i!}.$$

A calculation of the derivatives shows that $|p^{(i)}|/i! \leq \binom{k}{i} \cdot \hat{\ell}_1(p)$ and so

$$\varepsilon \geq \frac{p'(1)}{n/2} - \sum_{i \geq 2} \frac{\binom{k}{i}}{(n/2)^i} \cdot \hat{\ell}_1(p) = \frac{p'(1)}{n/2} - \hat{\ell}_1(p) \cdot \sum_{i \geq 2} \left(\frac{k}{n/2}\right)^i \cdot \frac{1}{i!}. \tag{1}$$

Since $p'(1) \leq \hat{\ell}_1(p') \leq k\hat{\ell}_1(p)$, n is at least $4k$ and

$$\sum_{i \geq 2} \left(\frac{k}{n/2}\right)^i \cdot \frac{1}{i!} \leq \left(\frac{k}{n/2}\right)^2 \cdot \sum_{i \geq 2} \frac{1}{i!} \leq \left(\frac{k}{n/2}\right)^2.$$

From (1) we obtain that

$$\varepsilon \geq \frac{p'(1)}{n/2} - \hat{\ell}_1(p) \cdot \frac{k^2}{(n/2)^2} \geq \frac{p'(1)}{n}. \tag{2}$$

where the second inequality follows from our choice of n.

By Lemma 7 for every pair of (k, δ)-wise indistinguishable distributions X and Y over $\{0, 1\}^n$,

$$\mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] \leq \frac{2}{2+\varepsilon} + e^2\delta\hat{\ell}_1(p).$$

If $\varepsilon \leq 1$, by Theorem 2' there exist $(k, 2^{-c\varepsilon n/k})$-wise indistinguishable distributions X and Y such that $\mathbb{E}[OR(X)] - \mathbb{E}[OR(Y)] \geq 1 - \varepsilon/6$. Setting $\delta = 2^{-c\varepsilon n/k}$ we obtain

$$e^2\delta\hat{\ell}_1(p) \geq 1 - \frac{\varepsilon}{6} - \frac{2}{2+\varepsilon} \geq \left(1 - \frac{\varepsilon}{6}\right) - \left(1 - \frac{\varepsilon}{3}\right) = \frac{\varepsilon}{6}.$$

Using inequality (2) and the definition of n we have

$$6e^2\hat{\ell}_1(p) \geq \varepsilon \cdot 2^{c\varepsilon n/k} \geq \frac{p'(1)}{n} \cdot 2^{cp'(1)/2k} = \frac{1}{\hat{\ell}_1(p)} \cdot \left(\frac{p'(1)}{2k}\right)^2 \cdot 2^{cp'(1)/2k}.$$

After rearranging terms we obtain

$$\hat{\ell}_1(p) \geq \frac{1}{\sqrt{6}e} \cdot \frac{p'(1)}{2k} \cdot 2^{cp'(1)/4k}.$$

Since $\hat{\ell}_1(p)$ is also at least $p(1) = 1$, it follows that $2^{\Omega(p'(1)/4k)}$, proving the theorem when $\varepsilon \leq 1$.

If $\varepsilon > 1$, by Theorem 2 there exist $(k, 2^{-cn/k})$-wise indistinguishable distributions X and Y such that $\mathbb{E}[\mathrm{OR}(X)] - \mathbb{E}[\mathrm{OR}(Y)] \geq 5/6$ and $e^2\delta\hat{\ell}_1(p) \geq 5/6 - 2/3 \geq 1/6$. Setting $\delta = 2^{-cn/k}$,

$$\hat{\ell}_1(p) \geq \frac{1}{6e^2} \cdot 2^{cn/k} \geq 2^{cp'(1)/2k}$$

using (2) and the assumption $\varepsilon > 1$. ◀

5 AND-OR formulas and perfect reconstruction

In this section we prove Theorem 4, give a variant with better parameters that provides almost-perfect reconstruction, and show that Theorem 4 is tight in the depth-2 case with respect to all uniform read-once AND ∘ OR formulas.

We first extend Lemma 10 to AND ∘ OR formulas of depth d:

▶ **Lemma 11.** *Assume there exist $(k, 0)$-wise indistinguishable distributions X and Y over $\{0,1\}^n$ for a regular depth-d AND ∘ OR tree f over n variables and with lowest-level fan-in m. Then there exist $(K, 2^{-\Omega(k)})$-wise indistinguishable distributions X' and Y' over $\{0,1\}^N$, $N = nM/m$, such that $\mathbb{E}[f'(X')] = \mathbb{E}[f(X)]$ and $\mathbb{E}[f'(Y')] = \mathbb{E}[f(Y)]$, where f' is a function taking the same form as f, except for that the lowest-level fan-in is M, provided $m \leq M/2$, $k/K \geq 4m/M$ and $n \leq m \cdot 2^{m-1}$.*

Proof. To sample from X' (resp. Y'), first sample a string from X (resp. Y), then extend each of the blocks with $M - m$ zeros positioned uniformly at random. Call the indices i in which X'_i inherits some bit of X *active*.

The distribution on active indices of X' can be described in the following alternative manner: First, choose each index i to be potentially active independently at random with probability $p = 2m/M$. If any block of X' has fewer than m potentially active indices, declare failure (F). Conditioned on not failing $(\overline{F})$, choose the active indices in each block uniformly at random among the potentially active ones.

Now let S be any set consisting of at most K inputs of f'. Let B be the event that S contains more than $2pK$ active indices. By Chernoff and union bounds,

$$\Pr[B|\overline{F}] \leq \frac{\Pr[B]}{1 - \Pr[F]} \leq \frac{2^{-pK}}{1 - (n/m)2^{-m}} \leq 2^{-k+1}$$

by our choice of parameters. As in the proof of Lemma 10 we conclude that X' and Y' must satisfy the conclusion. ◀

Proof of Theorem 4. We will use N and K to denote the quantities n and k from the statement of the theorem. By [15] and [3] there exist X and Y that are $(k, 0)$-wise indistinguishable but perfectly reconstructible by $\mathrm{AND}_{n^{1/2d-1}} \circ \mathrm{OR}_{n^{2/2d-1}} \circ \ldots \circ \mathrm{OR}_m$ for $m = n^{2/2d-1}$ and $k = \Omega(n^{\frac{d-1}{2d-1}})$. We will assume $K \geq N^{\frac{d-1}{2d-1}}$, for otherwise there is nothing left to prove. Set $k = (N/CK)^{1-1/d}$ for a sufficiently large constant C and $M = 4Km/k$. If $m > M/2$ then the conclusion follows directly from [15, 3] (as K will be at most a constant in terms of d times $N^{\frac{d-1}{2d-1}}$). If $n > m \cdot 2^{m-1}$, $(N/K)^{1-1/d}$ is upper bounded by a constant so the conclusion holds trivially. Otherwise, the statement of the theorem follows from Lemma 11. ◀

In the case $d = 2$, the parameters in Theorem 4 are the best possible, up to logarithmic terms, for all read-once CNFs. (The regime $k < n^{1/3}$ is resolved in [11, 3].)

▶ **Theorem 12.** *There exists a constant c such that for any n and $k \geq n^{1/3}$, no read-once CNF over n variables can perfectly reconstruct from any pair of $(k, 2^{-\Omega((n/k)^{1/2}\log^2 n)})$-wise indistinguishable distributions.*

The proof is omitted from this version owing to space limitations. If an exponentially small reconstruction error is acceptable, much better parameters for the underlying distributions are achievable:

▶ **Proposition 13.** *For all n, m, and k, the function $\mathrm{AND}_{n/m} \circ \mathrm{OR}_m$ can reconstruct from some pair of $(k, 2^{-\Omega(m/k)})$-wise indistinguishable distributions with error at most $2^{-\Omega(n/m)}$.*

Proof. Bun and Thaler [5] (improving work of Beigel) showed, via the connection in [3], that $\mathrm{AND}_{n/m} \circ \mathrm{OR}_m$ can reconstruct from some pair of $(\sqrt{m}, 0)$-wise indistinguishable distributions with error at most $2^{-\Omega(n/m)}$. We apply Lemma 11 with $k = \sqrt{m}$ and $M = mK/k$. ◀

6 A threshold weight lower bound

Proof of Corollary 6. Let X and Y be $(k, 2^{-\Omega((n/k)^{1-1/d})})$-wise indistinguishable distributions that the function f from Theorem 4 can perfectly reconstruct from. If p is a degree-k polynomial such that $|f(x) - p(x)| \leq 1/2 - 2^{-t}$ then by Lemma 7,

$$1 = \mathbb{E}[f(X)] - \mathbb{E}[f(Y)] \leq (1 - 2^{1-t}) + 2^{1-\Omega((n/k)^{1-1/d})} \cdot \hat{\ell}_1(p),$$

from where $\hat{\ell}_1(p) \geq 2^{\Omega((n/k)^{1-1/d})-t}$. Setting $t = c \cdot (n/k)^{1-1/d}$ gives the desired lower bound on $\hat{\ell}_1(p)$. ◀

The *threshold weight* of a function is its minimum degree-k threshold weight over all k. A result of Krause [9] (see also Lemma 27 in [5]) can be used to convert lower bounds on degree-k threshold weight into ones independent of degree.

▶ **Fact 14** (Krause, 2005). *For $f\colon \{0,1\}^n \to \{0,1\}$, let $F\colon \{0,1\}^{3n} \to \{0,1\}$ be given by $F(x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n) = f(\ldots, (\overline{z_i} \text{ AND } x_i) \text{ OR } (z_i \text{ AND } y_i), \ldots)$. For any k, if f requires degree-k threshold weight w then F requires threshold weight $\sqrt{\min\{w/2n, 2^k\}}$.*

Applying Fact 14 to Corollary 6, we obtain a linear-size depth-d family of formulas that requires threshold weight $2^{\Omega(n^{1/2-1/(4d-6)})}$ on inputs of length n, matching Sherstov's bound for formulas [15].

Acknowledgments. We thank Mark Bun for telling us about the work of Servedio, Tan, and Thaler and for his advice on polynomial approximations, and the ICALP 2017 reviewers for several helpful suggestions.

References

1 Richard Beigel. The polynomial method in circuit complexity. In *8th Structure in Complexity Theory Conference*, pages 82–95. IEEE, 1993.

2 Richard Beigel. Perceptrons, PP, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994. doi:10.1007/BF01263422.

3 Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *CRYPTO*, 2016.

4 Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. In *Coll. on Automata, Languages and Programming (ICALP)*, pages 303–314, 2013.

5 Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 2013.

6 Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

7 Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing, May 28-30, 1986, Berkeley, California, USA*, pages 6–20, 1986. doi:10.1145/12130.12132.

8 Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. Comput.*, 43(5):1699–1708, 2014. doi:10.1137/120897432.

9 M. Krause. On the computational power of boolean decision lists. In *Computational Complexity*, pages 362–375, 2005.

10 Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

11 Marvin Minsky and Seymour Papert. *Perceptrons*. MIT Press, Cambridge, MA, 1969.

12 Moni Naor and Adi Shamir. Visual cryptography. In *Advances in Cryptology – EUROCRYPT'94*, volume 950 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 1994.

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

14 R. Servedio, L-Y. Tan, and J. Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In *COLT*, 2012.

15 Alexander A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *ACM Symp. on the Theory of Computing (STOC)*, pages 223–232, 2014.

16 Alexander A. Sherstov. The power of asymmetry in constant-depth circuits. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2015.

17 Robert Špalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008. URL: http://arxiv.org/abs/0803.4516.

18 Andrew Yao. Separating the polynomial-time hierarchy by oracles. In *26th IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 1–10, 1985.