1 Lecture 10, Guest lecture by Justin Thaler, Scribe: Biswaroop Maiti

This is a guest lecture by Justin Thaler regarding lower bounds on approximate degree [BKT17, BT15, BT17]. Thanks to Justin for giving this lecture and for his help with the write-up. We will sketch some details of the lower bound on the approximate degree of $\mathsf{AND} \circ \mathsf{OR}$, SURJ and some intuition about the techniques used. Recall the definition of SURJ from the previous lecture as below:

Definition 1. The surjectivity function $\mathsf{SURJ}\colon \left(\{-1,1\}^{\log R}\right)^N \to \{-1,1\}$, takes input $x = (x_1, \ldots, x_N)$ where each $x_i \in \{-1,1\}^{\log R}$ is interpreted as an element of $[R]$. $\mathsf{SURJ}(x)$ has value -1 if and only if $\forall j \in [R], \exists i\colon x_i = j$.

Recall from the last lecture that $\mathsf{AND}_R \circ \mathsf{OR}_N\colon \{-1,1\}^{R\times N} \to \{-1,1\}$ is the block-wise composition of the AND function on R bits and the OR function on N bits. In general, we will denote the block-wise composition of two functions f, and g, where f is defined on R bits and g is defined on N bits, by $f_R \circ g_N$. Here, the outputs of R copies of g are fed into f (with the inputs to each copy of g being pairwise disjoint). The total number of inputs to $f_R \circ g_N$ is $R \cdot N$.

1.1 Lower Bound of $d_{1/3}(\mathsf{SURJ})$ via lower bound of $d_{1/3}(\mathbf{AND\text{-}OR})$

Claim 2. $d_{1/3}(\mathsf{SURJ}) = \widetilde{\Theta}(n^{3/4})$.

We will look at only the lower bound in the claim. We interpret the input as a list of N numbers from $[R] := \{1, 2, \cdots R\}$. As presented in [BKT17], the proof for the lower bound proceeds in the following steps.

1. Show that to approximate SURJ, it is necessary to approximate the block-composition $\mathsf{AND}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight at most N. i.e., show that $d_{1/3}(\mathsf{surj}) \geq d^{\leq N}_{1/3}(\mathsf{AND}_R \circ \mathsf{OR}_N)$.

   Step 1 was covered in the previous lecture, but we briefly recall a bit of intuition for why the claim in this step is reasonable. The intuition

comes from the fact that the *converse* of the claim is easy to establish, i.e., it is easy to show that in order to approximate SURJ, it is *sufficient* to approximate $\mathsf{AND}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight exactly N.

This is because SURJ can be expressed as an AND_R (over all range items $r \in [R]$) of the OR_N (over all inputs $i \in [N]$) of "Is input x_i equal to r"? Each predicate of the form in quotes is computed exactly by a polynomial of degree $\log R$, since it depends on only $\log R$ of the input bits, and exactly N of the predicates (one for each $i \in [N]$) evaluates to TRUE.

Step 1 of the lower bound proof for SURJ in [BKT17] shows a converse, namely that the *only* way to approximate SURJ is to approximate $\mathsf{AND}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight at most N.

2. Show that $d_{1/3}^{\leq N}(\mathsf{AND}_R \circ \mathsf{OR}_N) = \widetilde{\Omega}(n^{3/4})$, i.e., the degree required to approximate $\mathsf{AND}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight at most N is at least $D = \widetilde{\Omega}(n^{3/4})$.

   In the previous lecture we also sketched this Step 2. In this lecture we give additional details of this step. As in the papers, we use the concept of a "dual witness." The latter can be shown to be equivalent to bounded indistinguishability.

   Step 2 itself proceeds via two substeps:

   (a) Give a dual witness Φ for $\mathsf{AND}_R \cdot \mathsf{OR}_N$ that places little mass (namely, total mass less then $(R \cdot N \cdot D)^{-D}$) on inputs of hamming weight $\geq N$.

   (b) By modifying Φ, give a dual witness Φ' for $\mathsf{AND}_R \cdot \mathsf{OR}_N$ that places zero mass on inputs of Hamming weight $\geq N$.

In [BKT17], both Substeps 2a and 2b proceed entirely in the dual world (i.e., they explicitly manipulate dual witnesses Φ and Φ'). The main goal of this section of the lecture notes is to explain how to replace Step 2b of the argument of [BKT17] with a wholly "primal" argument.

The intuition of the primal version of Step 2b that we'll cover is as follows. First, we will show that a polynomial $p\colon \{-1,1\}^{R \cdot N} \to \{-1,1\}$ of degree D

that is bounded on the low Hamming Weight inputs, cannot be too big on the high Hamming weight inputs. In particular, we will prove the following claim.

Claim 3. If $p\colon \{-1,1\}^M \to \mathbb{R}$ is a degree D polynomial that satisfies $|p(x)| \leq 4/3$ on all inputs of x of Hamming weight at most D, then $|p(x)| \leq (4/3) \cdot D \cdot M^D$ for *all* inputs x.

Second, we will explain that the dual witness Φ constructed in Step 2a has the following "primal" implication:

Claim 4. For $D \approx N^{3/4}$, any polynomial p of degree D satisfying $|p(x) - (\mathsf{AND}_R \circ \mathsf{OR}_N)(x)| \leq 1/3$ for all inputs x of Hamming weight at most N must satisfy $|p(x)| > (4/3) \cdot D \cdot (R \cdot N)^D$ for some input $x \in \{-1,1\}^{R \cdot N}$.

Combining Claims 3 and 4, we conclude that no polynomial p of degree $D \approx N^{3/4}$ can satisfy

$$|p(x) - (\mathsf{AND}_R \circ \mathsf{OR}_N)(x)| \leq 1/3 \text{ for all inputs } x \text{ of Hamming weight at most } N, \tag{1}$$

which is exactly the desired conclusion of Step 2. This is because any polynomial p satisfying Equation 1 also satisfies $|p(x)| \leq 4/3$ for all x of Hamming weight of most N, and hence Claim 3 implies that

$$|p(x)| \leq \frac{4}{3} \cdot D \cdot (R \cdot N)^D \text{ for } \textit{all} \text{ inputs } x \in \{-1,1\}^{R \cdot N}. \tag{2}$$

But Claim 4 states that any polynomial satisfying both Equations 1 and 2 requires degree strictly larger than D.

In the remainder of this section, we prove Claims 3 and 4.

1.2 Proof of Claim 3

Proof of Claim 3. For notational simplicity, let us prove this claim for polynomials on domain $\{0,1\}^M$, rather than $\{-1,1\}^M$.

Proof in the case that p is symmetric. Let us assume first that p is symmetric, i.e., p is only a function of the Hamming weight $|x|$ of its input x. Then $p(x) = g(|x|)$ for some degree D univariate polynomial g (this is a direct consequence of Minsky-Papert symmetrization, which we have seen in the lectures before). We can express g as below in the same spirit of Lagrange interpolation.

$$g(t) = \sum_{k=0}^{D-1} g(k) \cdot \prod_{i=0}^{D-1} \frac{t-i}{k-i}.$$

Here, the first term, $g(k)$,is bounded in magnitude by $|g(k)| \leq 4/3$, and $|\prod_{i=0}^{D-1} \frac{t-i}{k-i}| \leq M^D$. Therefore, we get the final bound:

$$|g(t)| \leq (4/3) \cdot D \cdot M^D.$$

Proof for general p. Let us now consider the case of general (not necessarily symmetric) polynomials p. Fix any input $x \in \{0,1\}^M$. The goal is to show that $|p(x)| \leq \frac{4}{3} D \cdot M^D$.

Let us consider a polynomial $\hat{p}_x \colon \{0,1\}^{|x|} \to \{0,1\}$ of degree D obtained from p by restricting each input i such that $x_i = 0$ to have the value 0. For example, if $M = 4$ and $x = (0,1,1,0)$, then $\hat{p}_x(y_2, y_3) = p(0, y_2, y_3, 0)$. We will exploit three properties of $\hat{p}_x$:

Property 1. $\deg(\hat{p}_x) \leq \deg(p) \leq D$.

Property 2. Since $|p(x)| \leq 4/3$ for all inputs with $|x| \leq D$, $\hat{p}_x(y)$ satisfies the analogous property: $|\hat{p}_x(y)| \leq 4/3$ for all inputs with $|y| \leq D$.

Property 3. If $\mathbf{1}_{|x|}$ denotes the all-1s vector of length $|x|$, then $\hat{p}_x(\mathbf{1}_x) = p(x)$.

Property 3 means that our goal is to show that $|\widehat{p}(\mathbf{1}_x)| \leq \frac{4}{3} \cdot D \cdot M^D$.

Let $p_x^{\text{symm}} \colon \{0,1\}^M \to \mathbb{R}$ denote the symmetrized version of $\hat{p}_x$, i.e., $p_x^{\text{symm}}(y) = \mathbb{E}_\sigma[\hat{p}_x(\sigma(y))]$, where the expectation is over a random permutation σ of $\{1, \ldots, |x|\}$, and $\sigma(y) = (y_{\sigma(1)}, \ldots, y_{\sigma(|x|)})$. Since $\sigma(\mathbf{1}_{|x|}) = \mathbf{1}_{|x|}$ for all permutations σ, $p_x^{\text{symm}}(\mathbf{1}_{|x|}) = \hat{p}_x(\mathbf{1}_{|x|}) = p(x)$. But p_x^{symm} is symmetric, so Properties 1 and 2 together mean that the analysis from the first part of the proof implies that $|p_x^{\text{symm}}(y)| \leq \frac{4}{3} \cdot D \cdot M^D$ for all inputs y. In particular, letting $y = \mathbf{1}_{|x|}$, we conclude that $|p(x)| \leq \frac{4}{3} \cdot D \cdot M^D$ as desired. □

Discussion. One may try to simplify the analysis of the general case in the proof Claim 3 by considering the polynomial $p^{\text{symm}} \colon \{0,1\}^M \to \mathbb{R}$ defined via $p^{\text{symm}}(x) = \mathbb{E}_\sigma[p(\sigma(x))]$, where the expectation is over permutations σ of $\{1, \ldots, M\}$. p^{symm} is a symmetric polynomial, so the analysis for symmetric polynomials immediately implies that $|p^{\text{symm}}(x)| \leq \frac{4}{3} \cdot D \cdot M^D$. Unfortunately, this does *not* mean that $|p(x)| \leq \frac{4}{3} \cdot D \cdot M^D$.

This is because the symmetrized polynomial p^{symm} is averaging the values of p over all those inputs of a given Hamming weight. So, a bound on this averaging polynomial does not preclude the case where p is massively positive on some inputs of a given Hamming weight, and massively negative on other inputs of the same Hamming weight, and these values cancel out to obtain a small average value. That is, it is not enough to conclude that on the average over inputs of any given Hamming weight, the magnitude of p is not too big.

Thus, we needed to make sure that when we symmetrize $\hat{p}_x$ to p_x^{sym}, such large cancellations don't happen, and a bound of the average value of $\hat{p}$ on a given Hamming weight really gives us a bound on p on the input x itself. We defined $\hat{p}_x$ so that $\hat{p}_x(\mathbf{1}_M) = p(x)$. Since there is only *one* input in $\{0,1\}^M$ of Hamming weight M, $p_x^{\mathrm{symm}}(\mathbf{1}_M)$ does not average $\hat{p}_x$'s values on many inputs, meaning we don't need to worry about massive cancellations.

A note on the history of Claim 3. Claim 3 was implicit in [RS10]. They explicitly showed a similar bound for symmetric polynomials using primal view and (implicitly) gave a different (dual) proof of the case for general polynomials.

1.3 Proof of Claim 4

1.3.1 Interlude Part 1: Method of Dual Polynomials [BT17]

A dual polynomial is a dual solution to a certain linear program that captures the approximate degree of any given function $f\colon \{-1,1\}^n \to \{-1,1\}$. These polynomials act as certificates of the high approximate degree of f. The notion of strong LP duality implies that the technique is lossless, in comparison to symmetrization techniques which we saw before. For any function f and any ε, there is always some dual polynomial Ψ that witnesses a tight ε-approximate degree lower bound for f. A dual polynomial that witnesses the fact that $\mathsf{d}_\varepsilon(f) \geq d$ is a function $\Psi\colon \{-1,1\}^n \to \{-1,1\}$ satisfying three properties:

- **Correlation analysis:**

$$\sum_{x\in\{-1,1\}^n} \Psi(x)\cdot f(x) > \varepsilon.$$

If Ψ satisfies this condition, it is said to be well-correlated with f.

- **Pure high degree:** For all polynomials $p\colon \{-1,1\}^n \to \mathbb{R}$ of degree less than d, we have

$$\sum_{x \in \{-1,1\}^n} p(x) \cdot \Psi(x) = 0.$$

  If Ψ satisfies this condition, it is said to have *pure high degree* at least d.

- ℓ_1 **norm:**

$$\sum_{x \in \{-1,1\}^n} |\Psi(x)| = 1.$$

1.3.2 Interlude Part 2: Applying The Method of Dual Polynomials To Block-Composed Functions

For any function $f\colon \{-1,1\}^n \to \{-1,1\}$, we can write an LP capturing the approximate degree of f. We can prove lower bounds on the approximate degree of f by proving lower bounds on the value of feasible solution of this LP. One way to do this is by writing down the Dual of the LP, and exhibiting a feasible solution to the dual, thereby giving an upper bound on the value of the Dual. By the principle of LP duality, an upper bound on the Dual LP will be a lower bound of the Primal LP. Therefore, exhibiting such a feasible solution, which we call a dual witness, suffices to prove an approximate degree lower bound for f.

However, for any given dual witness, some work will be required to verify that the witness indeed meets the criteria imposed by the Dual constraints.

When the function f is a block-wise composition of two functions, say h and g, then we can try to construct a good dual witness for f by looking at dual witnesses for each of h and g, and combining them carefully, to get the dual witness for $h \circ g$.

The dual witness Φ constructed in Step 2a for $\mathsf{AND} \circ \mathsf{OR}$ is expressed below in terms of the dual witness of the inner OR function viz. Ψ_{OR} and the dual witness of the outer AND, viz. Ψ_{AND}.

$$\Phi(x_1 \dots x_R) = \Psi_{\mathsf{AND}}\left(\cdots, \mathsf{sgn}(\Psi_{\mathsf{OR}}(x_i)), \cdots\right) \cdot \prod_{i=1}^{R} |\Psi_{\mathsf{OR}}(x_i)|. \tag{3}$$

This method of combining dual witnesses Ψ_{AND} for the "outer" function AND and Ψ_{OR} for the "inner function" Ψ_{OR} is referred to in [BKT17, BT17] as *dual block composition.*

1.3.3 Interlude Part 3: Hamming Weight Decay Conditions

Step 2a of the proof of the SURJ lower bound from [BKT17] gave a dual witness Φ for $\mathsf{AND}_R \circ \mathsf{OR}_N$ (with $R = \Theta(N)$) that had pure high degree $\tilde{\Omega}(N^{3/4})$, and also satisfies Equations 4 and 5 below.

$$\sum_{|x|>N} |\Phi(x)| \ll (R \cdot N \cdot D)^{-D} \tag{4}$$

$$\text{For all } t = 0, \ldots, N, \sum_{|x|=t} |\Phi(x)| \leq \frac{1}{15 \cdot (1+t)^2}. \tag{5}$$

Equation 4 is a very strong "Hamming weight decay" condition: it shows that the total mass that Ψ places on inputs of high Hamming weight is very small. Hamming weight decay conditions play an essential role in the lower bound analysis for SURJ from [BKT17]. In addition to Equations 4 and 5 themselves being Hamming weight decay conditions, [BKT17]'s proof that Φ satisfies Equations 4 and 5 exploits the fact that the dual witness Ψ_{OR} for OR can be chosen to simultaneously have pure high degree $N^{1/4}$, and to satisfy the following weaker Hamming weight decay condition:

Claim 5. There exist constants c_1, c_2 such that for all $t = 0, \cdots N$,

$$\sum_{|x|=t} \Psi_{\mathsf{OR}}(x) \leq c_1 \cdot \frac{1}{(1+t)^2} \cdot \exp(-c_2 \cdot t/N^{1/4}). \tag{6}$$

(We will not prove Claim 5 in these notes, we simply state it to highlight the importance of dual decay to the analysis of SURJ).

Dual witnesses satisfying various notions of Hamming weight decay have a natural primal interpretation: they witness approximate degree lower bounds for the target function ($\mathsf{AND}_R \circ \mathsf{OR}_N$ in the case of Equation 4, and OR_N in the case of Equation 6) *even when the approximation is allowed to be exponentially large on inputs of high Hamming weight.* This primal interpretation of dual decay is formalized in the following claim.

Claim 6. Let $L(t)$ be any function mapping $\{0, 1, \ldots, N\}$ to $\mathbb{R}_+$. Suppose Ψ is a dual witness for f satisfying the following properties:

- (Correlation): $\sum_{x \in \{-1,1\}^n} \Psi(x) \cdot f(x) > 1/3$.
- (Pure high degree): Ψ has pure high degree D.
- (Dual decay): $\sum_{|x|=t} |\Psi(x)| \leq \frac{1}{5 \cdot (1+t)^2 \cdot L(t)}$ for all $t = 0, 1, \ldots, N$.

Then there is no degree D polynomial p such that

$$|p(x) - f(x)| \leq L(t) \text{ for all } t = 0, 1, \ldots, N. \tag{7}$$

Proof. Let p be any degree D polynomial. Since Ψ has pure high degree D, $\sum_{x \in \{-1,1\}^N} p(x) \cdot \Psi(x) = 0$.

We will now show that if p satisfies Equation 7, then the other two properties satisfied by Ψ (correlation and dual decay) together imply that $\sum_{x \in \{-1,1\}^N} p(x) \cdot \Psi(x) > 0$, a contradiction.

$$\begin{aligned}
\sum_{x \in \{-1,1\}^N} \Psi(x) \cdot p(x) &= \sum_{x \in \{-1,1\}^N} \Psi(x) \cdot f(x) - \sum_{x \in \{-1,1\}^N} \Psi(x) \cdot (p(x) - f(x)) \\
&\geq 1/3 - \sum_{x \in \{-1,1\}^N} |\Psi(x)| \cdot |p(x) - f(x)| \\
&\geq 1/3 - \sum_{t=0}^{N} \sum_{|x|=t} |\Psi(x)| \cdot L(t) \\
&\geq 1/3 - \sum_{t=0}^{N} \frac{1}{5 \cdot (1+t)^2 \cdot L(t)} \cdot L(t) \\
&= 1/3 - \sum_{t=0}^{N} \frac{1}{5 \cdot (1+t)^2} > 0.
\end{aligned}$$

Here, Line 2 exploited that Ψ has correlation at least $1/3$ with f, Line 3 exploited the assumption that p satisfies Equation 7, and Line 4 exploited the dual decay condition that Ψ is assumed to satisfy. □

1.3.4 Proof of Claim 4

Proof. Claim 4 follows from Equations 4 and 5, combined with Claim 6. Specifically, apply Claim 6 with $f = \mathsf{AND}_R \circ \mathsf{OR}_N$, and

$$L(t) = \begin{cases} 1/3 \text{ if } t \leq N \\ (R \cdot N \cdot D)^D \text{ if } t > N. \end{cases}$$

□

2 Generalizing the analysis for SURJ to prove a nearly linear approximate degree lower bound for AC^0

Now we take a look at how to extend this kind of analysis for SURJ to obtain even stronger approximate degree lower bounds for other functions in AC^0. Recall that SURJ can be expressed as an AND_R (over all range items $r \in [R]$) of the OR_N (over all inputs $i \in [N]$) of "Is input x_i equal to r"? That is, SURJ simply evaluates $\mathsf{AND}_R \circ \mathsf{OR}_N$ on the inputs $(\dots, y_{j,i}, \dots)$ where $y_{j,i}$ indicates whether or not input x_i is equal to range item $j \in [R]$.

Our analysis for SURJ can be viewed as follows: It is a way to turn the AND function on R bits (which has approximate degree $\Theta\left(\sqrt{R}\right)$) into a function on close to R bits, with polynomially larger approximate degree (i.e. SURJ is defined on $N \log R$ bits where, say, the value of N is $100R$, i.e., it is a function on $100R \log R$ bits). So, this function is on not much more than R bits, but has approximate degree $\tilde{\Omega}(R^{3/4})$, polynomially larger than the approximate degree of AND_R.

Hence, the lower bound for SURJ can be seen as a hardness amplification result. We turn the AND function on R bits to a function on slightly more bits, but the approximate degree of the new function is significantly larger.

From this perspective, the lower bound proof for SURJ showed that in order to approximate SURJ, we need to not only approximate the AND_R function, but, additionally, instead of feeding the inputs directly to AND gate itself, we are further driving up the degree by feeding the input through OR_N gates. The intuition is that we cannot do much better than merely approximate the AND function and then approximating the block composed

OR_N gates. This additional approximation of the OR gates give us the extra exponent in the approximate degree expression.

We will see two issues that come in the way of naive attempts at generalizing our hardness amplification technique from AND_R to more general functions.

2.1 Interlude: Grover's Algorithm

Grover's algorithm [Gro96] is a quantum algorithm that finds with high probability the unique input to a black box function that produces a given output, using $O(\sqrt{N})$ queries on the function, where N is the size of the the domain of the function. It is originally devised as a database search algorithm that searches an unsorted database of size N and determines whether or not there is a record in the database that satisfies a given property in $O(\sqrt{N})$ queries. This is strictly better compared to deterministic and randomized query algorithms because they will take $\Omega(N)$ queries in the worst case and in expectation respectively. Grover's algorithm is optimal up to a constant factor, for the quantum world.

2.2 Issues: Why a dummy range item is necessary

In general, let us consider the problem of taking any function f that does not have maximal approximate degree (say, with approximate degree $n^{1-\Omega(1)}$), and turning it into a function on roughly the same number of bits, but with polynomially larger approximate degree.

In analogy with how $\mathsf{SURJ}(x_1, \ldots, x_N)$ equals $\mathsf{AND}_R \circ \mathsf{OR}_N$ evaluated on inputs $(\ldots, y_{ji}, \ldots)$, where y_{ji} indicates whether or not $x_i = j$, we can consider the block composition $f_R \circ \mathsf{OR}_N$ evaluated on $(\ldots, y_{ji}, \ldots)$, and hope that this function has polynomially larger approximate degree than f_R itself.

Unfortunately, this does not work. Consider for example the case $f_R = \mathsf{OR}_R$. The function $\mathsf{OR}_R \circ \mathsf{OR}_N = \mathsf{OR}_{R \cdot N}$ evaluates to 1 on all possible vectors $(\ldots, y_{ji}, \ldots,)$, since all such vectors of Hamming weight exactly $N > 0$.

One way to try to address this is to introduce a dummy range item, all occurrences of which are simply ignored by the function. That is, we can consider the (hopefully harder) function G to interpret its input as a list of N numbers from the range $[R]_0 := \{0, 1, \ldots, R\}$, rather than range $[R]$, and define $G = f_R \circ \mathsf{OR}_N(y_{1,1}, \ldots, y_{R,N})$ (note that variables $y_{0,1}, \ldots, y_{0,N}$,

which indicate whether or not each input x_i equals range item 0, are simply ignored).

In fact, in the previous lecture we already used this technique of introducing a "dummy" range item, to ease the lower bound analysis for SURJ itself. Last lecture we covered Step 1 of the lower bound proof for SURJ, and we let $z_0 = \sum_{i=1}^{N} y_{0,i}$ denote the frequency of the dummy range item, 0. The introduction of this dummy range item let us replace the condition $\sum_{j=0}^{R} z_j = N$ (i.e., the sum of the frequencies of all the range items is *exactly* N) by the condition $\sum_{j=1}^{R} z_j \leq N$ (i.e., the sum of the frequencies of all the range items is *at most* N).

2.3 A dummy range item is not sufficient on its own

Unfortunately, introducing a dummy range item is not sufficient on its own. That is, even when the range is is $[R]_0$ rather than $[R]$, the function $G = f_R \circ \mathsf{OR}_N(y_{1,1}, \ldots, y_{R,N})$ may have approximate degree that is *not* polynomially larger than that of f_R itself. An example of this is (once again) $f_R = \mathsf{OR}_R$. With a dummy range item, $\mathsf{OR}_R \circ \mathsf{OR}_N(y_{1,1}, \ldots, y_{R,N})$ evaluates to TRUE if and only if at least one of the N inputs is *not* equal to the dummy range item 0. This problem has approximate degree $O(N^{1/2})$ (it can be solved using Grover search).

Therefore, the most naive approach at general hardness amplification, even with a dummy range item, does not work.

2.4 The approach that works

The approach that succeeds is to consider the block composition $f \circ \mathsf{AND}_{\log R} \circ \mathsf{OR}_N$ (i.e., apply the naive approach with a dummy range item not to f_R itself, but to $f_R \circ \mathsf{AND}_{\log R}$). As pointed out in Section 2.3, the $\mathsf{AND}_{\log R}$ gates are crucial here for the analysis to go through.

It is instructive to look at where exactly the lower bound proof for SURJ breaks down if we try to adapt it to the function $\mathsf{OR}_R \circ \mathsf{OR}_N = \mathsf{OR}_{R \cdot N}$ (rather than the function $\mathsf{AND}_R \circ \mathsf{OR}_N$ which we analyzed to prove the lower bound for SURJ). Then we can see why the introduction of the $\mathsf{AND}_{\log R}$ gates fixes the issue.

When analyzing the more naively defined function $G = (\mathsf{OR}_R \circ \mathsf{OR}_N)(y_{1,1}, \ldots, y_{R,N})$ (with a dummy range item), Step 1 of the lower bound analysis for SURJ

does work unmodified to imply that in order to approximate G, it is necessary to approximate block composition of $\mathsf{OR}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight at most N. But Step 2 of the analysis breaks down: one can approximate $\mathsf{OR}_R \circ \mathsf{OR}_N$ on inputs of Hamming weight at most N using degree just $O(N^{1/2})$.

Why does the Step 2 analysis break down for $\mathsf{OR}_R \circ \mathsf{OR}_N$? If one tries to construct a dual witness Φ for $\mathsf{OR}_R \circ \mathsf{OR}_N$ by applying dual block composition (cf. Equation 3, but with the dual witness Ψ_{AND} for AND_R replaced by a dual witness for OR_R), Φ will not be well-correlated with $\mathsf{OR}_R \circ \mathsf{OR}_N$.

Roughly speaking, the correlation analysis thinks of each copy of the inner dual witness $\Psi_{\mathsf{OR}}(x_i)$ as consisting of a sign, $\mathsf{sgn}(\Psi_{\mathsf{OR}})(x_i)$, and a magnitude $|\Psi_{\mathsf{OR}}(x_i)|$, and the inner dual witness "makes an error" on x_i if it outputs the wrong sign, i.e., if $\mathsf{sgn}(\Psi_{\mathsf{OR}})(x_i) \neq \mathsf{OR}(x_i)$. The correlation analysis winds up performing a union bound over the probability (under the product distribution $\prod_{i=1}^{R} |\Psi_{\mathsf{OR}}(x_i)|$) that *any* of the R copies of the inner dual witness makes an error. Unfortunately, each copy of the inner dual witness makes an error with constant probability under the distribution $|\Psi_{\mathsf{OR}}|$. So at least one of them makes an error under the product distribution with probability very close to 1. This means that the correlation of the dual-block-composed dual witness Φ with $\mathsf{OR}_R \circ \mathsf{OR}_N$ is poor.

But if we look at $\mathsf{OR}_R \circ (\mathsf{AND}_{\log R} \circ \mathsf{OR}_N)$, the correlation analysis *does* go through. That is, we can give a dual witness Ψ_{in} for $\mathsf{AND}_{\log R} \circ \mathsf{OR}_N$ and a dual witness Ψ_{out} for OR_R such that the the dual-block-composition of Ψ_{out} and Ψ_{in} is well-correlated with $\mathsf{OR}_R \circ (\mathsf{AND}_{\log R} \circ \mathsf{OR}_N)$.

This is because [BT15] showed that for $\epsilon = 1 - 1/(3R)$, $d_\epsilon\left(\mathsf{AND}_{\log R} \circ \mathsf{OR}_N\right) = \Omega(N^{1/2})$. This means that $(\mathsf{AND}_{\log R} \circ \mathsf{OR}_N)$ has a dual witness Ψ_{in} that "makes an error" with probability just $1/(3R)$. This probability of making an error is so low that a union bound over all R copies of Ψ_{in} appearing in the dual-block-composition of Ψ_{out} and Ψ_{in} implies that with probability at least $1/3$, *none* of the copies of Ψ_{in} make an error.

In summary, the key difference between OR_N and $\mathsf{AND}_{\log R} \circ \mathsf{OR}_N$ that allows the lower bound analysis to go through for the latter but not the former is that the latter has ϵ-approximate degree $\Omega(N^{1/2})$ for $\epsilon = 1 - 1/(3R)$, while the former only has ϵ-approximate degree $\Omega(N^{1/2})$ if ϵ is a constant bounded away from 1.

To summarize, the SURJ lower bound can be seen as a way to turn the function $f_R = \mathsf{AND}_R$ into a harder function $G = \mathsf{SURJ}$, meaning that G has polynomially larger approximate degree than f_R. The right approach to

generalize the technique for arbitrary f_R is to (a) introduce a dummy range item, all occurrences of which are effectively ignored by the harder function G, *and* (b) rather than considering the "inner" function OR_N, consider the inner function $\mathsf{AND}_{\log R} \circ \mathsf{OR}_N$, i.e., let $G = f_R \circ \mathsf{AND}_{\log R} \circ \mathsf{OR}_N(y_{1,1} \ldots, y_{R \log R,N})$. The $\mathsf{AND}_{\log R}$ gates are essential to make sure that the error in the correlation of the inner dual witness is very small, and hence the correlation analysis for the dual-block-composed dual witness goes through. Note that G can be interpreted as follows: it breaks the range $[R \log R]_0$ up into R blocks, each of length $\log R$, (the dummy range item is excluded from all of the blocks), and for each block it computes a bit indicating whether or not every range item in the block has frequency at least 1. It then feeds these bits into f_R.

By recursively applying this construction, starting with $f_R = \mathsf{AND}_R$, we get a function in AC^0 with approximate degree $\Omega(n^{1-\delta})$ for any desired constant $\delta > 0$.

2.5 $k-$distinctness

The above mentioned very same issue also arises in [BKT17]'s proof of a lower bound on the approximate degree of the k-distinctness function. Step 1 of the lower bound analysis for SURJ reduced analyzing k-distinctness to analyzing $\mathsf{OR} \circ \mathsf{TH}^k_N$ (restricted to inputs of Hamming weight at most N), where TH^k_N is the function that evaluates to TRUE if and only if its input has Hamming weight at least k. The lower bound proved in [BKT17] for k-distinctness is $\Omega(n^{3/4-1/(2k)})$. OR is the TH^1 function. So, $\mathsf{OR}_R \circ \mathsf{TH}^k$ is "close" to $\mathsf{OR}_R \circ \mathsf{OR}_N$. And we've seen that the correlation analysis of the dual witness obtained via dual-block-composition breaks down for $\mathsf{OR}_R \circ \mathsf{OR}_N$.

To overcome this issue, we have to show that TH^k_N is harder to approximate than OR_N itself, but we have to give up some small factor in the process. We will lose some quantity compared to the $\Omega(n^{3/4})$ lower bound for SURJ. It may seem that this loss factor is just a technical issue and not intrinsic, but this is not so. In fact, this bound is almost tight. There is an upper bound from a complicated quantum algorithm [BL11, Bel12] for k-distinctness that makes $O(n^{3/4-1/(2^{k+2}-4)}) = n^{3/4-\Omega(1)}$ that we won't elaborate on here.

References

[Bel12] Aleksandrs Belovs. Learning-graph-based quantum algorithm for k-distinctness. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 207–216. IEEE, 2012.

[BKT17] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *arXiv preprint arXiv:1710.09079*, 2017.

[BL11] Aleksandrs Belovs and Troy Lee. Quantum algorithm for k-distinctness with prior knowledge on the input. *arXiv preprint arXiv:1108.3022*, 2011.

[BT15] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *International Colloquium on Automata, Languages, and Programming*, pages 268–280. Springer, 2015.

[BT17] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. *arXiv preprint arXiv:1703.05784*, 2017.

[Gro96] Lov K Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 212–219. ACM, 1996.

[RS10] Alexander A Razborov and Alexander A Sherstov. The sign-rank of AC^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010.