THEORY OF COMPUTING, Volume 16 (10), 2020, pp. 1–71
www.theoryofcomputing.org

The Polynomial Method Strikes Back: Tight Quantum Query Bounds via Dual Polynomials

Mark Bun* Robin Kothari Justin Thaler†

Received February 6, 2019; Revised November 26, 2019; Published October 24, 2020

Abstract. The approximate degree of a Boolean function f is the least degree of a real polynomial that approximates f pointwise to error at most $1/3$. The approximate degree of f is known to be a lower bound on the quantum query complexity of f (Beals et al., FOCS 1998 and J. ACM 2001).

We find tight or nearly tight bounds on the approximate degree and quantum query complexities of several basic functions. Specifically, we show the following.

- k-Distinctness: For any constant k, the approximate degree and quantum query complexity of the k-distinctness function is $\Omega(n^{3/4-1/(2k)})$. This is nearly tight for large k, as Belovs (FOCS 2012) has shown that for any constant k, the approximate degree and quantum query complexity of k-distinctness is $O(n^{3/4-1/(2^{k+2}-4)})$.
- Image size testing: The approximate degree and quantum query complexity of testing the size of the image of a function $[n] \to [n]$ is $\tilde{\Omega}(n^{1/2})$. This proves a conjecture of

An extended abstract of this paper [28] appeared in the Proceedings of the 50th ACM Symposium on Theory of Computing (STOC 2018).

*Supported by NSF grant CCF-1947889.

†Supported by NSF CAREER award CCF-1845125.

ACM Classification: F.2.3

AMS Classification: 68Q17, 81P68

Key words and phrases: polynomial method, quantum query complexity, dual polynomials, surjectivity

 DOI: 10.4086/toc.2020.v016a010

Ambainis et al. (SODA 2016), and it implies tight lower bounds on the approximate degree and quantum query complexity of the following natural problems.

  – k-Junta testing: A tight $\tilde{\Omega}(k^{1/2})$ lower bound for k-junta testing, answering the main open question of Ambainis et al. (SODA 2016).
  – Statistical distance from uniform: A tight $\tilde{\Omega}(n^{1/2})$ lower bound for approximating the statistical distance of a distribution from uniform, answering the main question left open by Bravyi et al. (STACS 2010 and IEEE Trans. Inf. Theory 2011).
  – Shannon entropy: A tight $\tilde{\Omega}(n^{1/2})$ lower bound for approximating Shannon entropy up to a certain additive constant, answering a question of Li and Wu (2017).

- Surjectivity: The approximate degree of the surjectivity function is $\tilde{\Omega}(n^{3/4})$. The best prior lower bound was $\Omega(n^{2/3})$. Our result matches an upper bound of $\tilde{O}(n^{3/4})$ due to Sherstov (STOC 2018), which we reprove using different techniques. The quantum query complexity of this function is known to be $\Theta(n)$ (Beame and Machmouchi, Quantum Inf. Comput. 2012 and Sherstov, FOCS 2015).

Our upper bound for surjectivity introduces new techniques for approximating Boolean functions by low-degree polynomials. Our lower bounds are proved by significantly refining techniques recently introduced by Bun and Thaler (FOCS 2017).

Contents

1 Introduction

Approximate degree. The approximate degree of a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}(f)$, is the least degree of a real polynomial p such that $|p(x) - f(x)| \leq 1/3$ for all $x \in \{-1,1\}^n$. Approximate degree is a basic measure of the complexity of a Boolean function, and has diverse applications throughout theoretical computer science.

Upper bounds on approximate degree are at the heart of the most powerful known learning algorithms in a number of models [10, 46, 47, 49, 50, 63, 68], algorithmic approximations for the inclusion-exclusion principle [45, 70], and algorithms for differentially private data release [35, 87]. A recent line of work [84, 85] has used approximate degree upper bounds to show new lower bounds on the formula and graph complexities of explicit functions.

Lower bounds on approximate degree have enabled progress in several areas of complexity theory, including communication complexity [27, 36, 38, 39, 41, 64, 69, 72, 74, 77], circuit complexity [60, 71], oracle separations [15, 21], and secret-sharing [20]. Most importantly for this paper, approximate degree lower bounds have been critical in shaping our understanding of *quantum query complexity* [1, 3, 13],

In spite of the importance of approximate degree, major gaps remain in our knowledge. In particular, the approximate degrees of many basic functions are still unknown. Our goal in this paper is to determine the approximate degrees of several natural functions for which there were significant gaps between known upper and lower bounds.

Quantum query complexity. While determining the approximate degree of basic functions of interest is a test of our understanding of approximate degree, it is also motivated by the study of quantum algorithms. In the quantum query model, a quantum algorithm is given query access to the bits of an input x, and the goal is to compute some function f of x while minimizing the number of queried bits. Quantum query complexity captures much of the power of quantum computing, and most quantum algorithms were discovered in or can easily be described in the query setting.

Approximate degree was one of the first general lower bound techniques for quantum query complexity. In 1998, Beals et al. [13] observed that the bounded-error quantum query complexity of a function f is lower bounded by (one half times) the approximate degree of f. Since polynomials are sometimes easier to understand than quantum algorithms, this observation led to a number of new lower bounds on quantum query complexity. This method of proving quantum query lower bounds is called the *polynomial method*.

After several significant quantum query lower bounds were proved via the polynomial method (including the work of Aaronson and Shi [3], who proved optimal lower bounds for the collision and element distinctness problems), the polynomial method took a back seat. Since then, the positive-weights adversary method [5, 12, 52, 92] and the newer negative-weights adversary method [44, 55, 66] have become the tools of choice for proving quantum query lower bounds (with some notable exceptions, such as Zhandry's recent tight lower bound for the set equality problem [91]). This leads us to our second goal for this work.

In this article, we seek to resolve several open problems in quantum query complexity using approximate degree as the lower bound technique. A distinct advantage of proving quantum query lower bounds with the polynomial method is that any such bound can be "lifted" via Sherstov's *pattern matrix method* [72] to a quantum *communication* lower bound (even with unlimited shared entanglement [56]);

Problem	Best prior upper bound	Our lower bound	Best prior lower bound
k-Distinctness	$O(n^{3/4-1/(2^{k+2}-4)})$ [16]	$\tilde{\Omega}(n^{3/4-1/(2k)})$	$\tilde{\Omega}(n^{2/3})$ [3]
Image size testing	$O(\sqrt{n}\log n)$ [9]	$\tilde{\Omega}(\sqrt{n})$	$\tilde{\Omega}(n^{1/3})$ [9]
k-Junta testing	$O(\sqrt{k}\log k)$ [9]	$\tilde{\Omega}(\sqrt{k})$	$\tilde{\Omega}(k^{1/3})$ [9]
SDU	$O(\sqrt{n})$ [23]	$\tilde{\Omega}(\sqrt{n})$	$\tilde{\Omega}(n^{1/3})$ [3,23]
Shannon entropy	$\tilde{O}(\sqrt{n})$ [23,58]	$\tilde{\Omega}(\sqrt{n})$	$\tilde{\Omega}(n^{1/3})$ [58]

Table 1: Our lower bounds on quantum query complexity and approximate degree vs. prior work.

such a result is not known for any other quantum query lower bound technique. More generally, using approximate degree as a lower bound technique for quantum query complexity has other advantages, such as the ability to show lower bounds for zero-error and small-error quantum algorithms [24], unbounded-error quantum algorithms [13], and time-space tradeoffs [48].

Quantum query complexity and approximate degree. We illustrate the power of the polynomial method by proving optimal or nearly optimal bounds on several functions studied in the quantum computing literature. These results are summarized in Table 1, and definitions of the problems considered can be found in Section 1.1. Since the upper bounds for these functions were shown using quantum algorithms, our results determine both the quantum query complexity and approximate degree of these functions.

For most of the functions studied in this paper, the positive-weights adversary bound provably cannot show optimal lower bounds due to the certificate complexity barrier [83,92] and the property testing barrier [44]. While these barriers do not apply to the negative-weights variant (which is actually capable of proving tight quantum query lower bounds for *all* functions [55,66]), the negative-weights adversary method is often challenging to apply to specific problems, and tight bounds for the problems we consider have remained unknown for a long time.

For the functions presented in Table 1, the approximate degree and quantum query complexity are essentially the same. This is not the case for the surjectivity function, which has played an important role in the literature on approximate degree and quantum query complexity. Specifically, Beame and Machmouchi [14] showed that surjectivity has quantum query complexity $\tilde{\Theta}(n)$. On the other hand, Sherstov recently showed that surjectivity has approximate degree $\tilde{O}(n^{3/4})$ [78]. Surjectivity is the only known example of a "natural" function separating approximate degree from quantum query complexity; prior examples of such functions [2,7] were contrived, and (unlike surjectivity) specifically constructed to separate the two measures.

Our final result gives a nearly tight estimate of the approximate degree of surjectivity. We prove a new lower bound of $\tilde{\Omega}(n^{3/4})$, which matches Sherstov's upper bound up to logarithmic factors. We also give a new construction of an approximating polynomial of degree $\tilde{O}(n^{3/4})$, using very different techniques than [78]. We believe that our proof of this $\tilde{O}(n^{3/4})$ upper bound is of independent interest. In particular, our lower bound proof for surjectivity is specifically tailored to showing optimality of our upper bound construction, in a sense that can be made formal via complementary slackness. We are optimistic that our

Problem	Prior upper bound	Our upper bound	Our lower bound	Prior lower bound
Surjectivity	$\tilde{O}(n^{3/4})$ [78]	$\tilde{O}(n^{3/4})$	$\tilde{\Omega}(n^{3/4})$	$\tilde{\Omega}(n^{2/3})$ [3]

Table 2: Our bounds on the approximate degree of surjectivity vs. prior work.

approximation techniques will be useful for showing additional tight approximate-degree bounds in the future.

1.1 Our results

We now describe our results and prior work on these functions in more detail.

1.1.1 Functions considered

We now informally describe the functions studied in this paper. These functions are formally defined in Section 2.4.

Let R be a power of two and $N \geq R$, and let $n = N \cdot \log_2 R$. Most of the functions that we consider interpret their inputs in $\{-1,1\}^n$ as a list of N numbers from a range $[R]$, and determine whether this list satisfies various natural conditions. We let the *frequency* f_i of range item $i \in R$ denote the number of times i appears in the input list.

In this paper we study the following functions in which the input is N numbers from a range $[R]$:

- Surjectivity (SURJ): Do all range items appear at least once?
- k-Distinctness: Is there a range item that appears k or more times?
- Image size testing: Decide if all range items appear at least once or if at most $\gamma \cdot R$ range items appear at least once, under the promise that one of these is true.
- Statistical distance from uniform (SDU): Interpret the input as a probability distribution p, where $p_i = f_i/N$. Compute the statistical distance of p from the uniform distribution over R up to some small additive error ε.
- Shannon entropy: Interpret the input as a probability distribution p, where $p_i = f_i/N$. Compute the Shannon entropy $\sum_{i \in R} p_i \cdot \log(1/p_i)$ of p up to additive error ε.

An additional function we consider that does not fit neatly into the framework above is k-junta testing.

- k-Junta testing: Given an input in $\{-1,1\}^n$ representing the truth table of a function $\{-1,1\}^{\log n} \to \{-1,1\}$, determine whether this function depends on at most k of its input bits, or is at least ε-far from any such function.

We give tight or nearly tight bounds on the quantum query complexities and/or approximate degrees of all of the functions above. Our lower bounds for SURJ, k-distinctness, image size testing, SDU, and entropy approximation all require N to be "sufficiently larger" than R, by a certain constant factor. For simplicity, throughout this introduction we do not make this requirement explicit, and for this reason we label the theorems in this introduction informal.

1.1.2 Results in detail

Surjectivity. In the surjectivity problem we are given N numbers from $[R]$ and must decide if every range item appears at least once in the input.

The quantum query complexity of this problem was studied by Beame and Machmouchi [14], who proved a lower bound of $\tilde{\Omega}(n)$, which was later improved by Sherstov to the optimal $\Theta(n)$ [80]. Beame and Machmouchi [14] explicitly leave open the question of determining the approximate degree of surjectivity. Recently, Sherstov [78] showed an upper bound of $\tilde{O}(n^{3/4})$ on the approximate degree of this function. The best prior lower bound was $\tilde{\Omega}(n^{2/3})$ [3, 33].

We give a completely different construction of an approximating polynomial for surjectivity with degree $\tilde{O}(n^{3/4})$. We also prove a matching lower bound, which shows that the approximate degree of the surjectivity function is $\tilde{\Theta}(n^{3/4})$.

Theorem 1.1 (Informal)**.** *The approximate degree of* SURJ *is* $\tilde{\Theta}(n^{3/4})$.

k**-Distinctness.** In this problem, we are given N numbers in $[R]$ and must decide if any range item appears at least k times in the list (i. e., is there an $i \in [R]$ with $f_i \geq k$?). This generalizes the well-studied element distinctness problem, which is the same as 2-distinctness.

Ambainis [8] first used quantum walks to give an $O(n^{k/(k+1)})$ upper bound on the quantum query complexity of any problem with certificates of size k, including k-distinctness and k-sum.[1] Later, Belovs introduced a beautiful new framework for designing quantum algorithms [17] and used it to improve the upper bound for k-distinctness to $O(n^{3/4-1/(2^{k+2}-4)})$ [16]. Several subsequent works have used Belovs' k-distinctness algorithm as a black-box subroutine for solving more complicated problems (e. g., [58, 61]).

As for lower bounds, Aaronson and Shi [3] established an $\tilde{\Omega}(n^{2/3})$ lower bound on the approximate degree of k-distinctness for any $k \geq 2$. Belovs and Špalek used the adversary method to prove a lower bound of $\Omega(n^{k/(k+1)})$ on the quantum query complexity of k-sum, showing that Ambainis' algorithm is tight for k-sum. They asked whether their techniques can prove an $\omega(n^{2/3})$ quantum query lower bound for k-distinctness. We achieve this goal, but using the polynomial method instead of the adversary method. Our main result is the following:

Theorem 1.2 (Informal)**.** *For any $k \geq 2$, the approximate degree and quantum query complexity of k-distinctness is* $\tilde{\Omega}(n^{3/4-1/(2k)})$.

This is nearly tight for large k, as it approaches Belovs' upper bound of $O(n^{3/4-1/(2^{k+2}-4)})$. Note that the exponent in both bounds approaches $3/4$. It remains an intriguing open question to close the gap between $n^{3/4-1/(2^{k+2}-4)}$ and $n^{3/4-1/(2k)}$, especially for small values of $k \geq 3$.

Our k-distinctness lower bound also implies an $\tilde{\Omega}(n^{3/4-1/(2k)})$ lower bound on the quantum query complexity of approximating the maximum frequency, F_∞, of any element up to relative error less than $1/k$ [61], improving over the previous best bound of $\tilde{\Omega}(n^{2/3})$.

[1] In the k-sum problem, we are given N numbers in $[R]$ and asked to decide if any k of them sum to $0 \pmod{R}$.

Image size testing. In this problem, we are given N numbers in $[R]$ and $0 < \gamma < 1$, and must decide if every range item appears at least once or if at most $\gamma \cdot R$ range items appear at least once. We show for any $\gamma > 0$, the problem has approximate degree and quantum query complexity $\tilde{\Omega}(\sqrt{n})$. This holds as long as $N = c \cdot R$ for a certain constant $c > 0$.

Theorem 1.3 (Informal)**.** *The approximate degree and quantum query complexity of image size testing is* $\tilde{\Omega}(\sqrt{n})$.

This lower bound is tight, matching a quantum algorithm of Ambainis, Belovs, Regev, and de Wolf [9], and confirms a conjecture from their work. The previous best lower bound was $\Omega(n^{1/3})$ [9] obtained via reduction to the collision lower bound [3]. The classical query complexity of this problem is $\Theta(n/\log n)$ [89].

The version of image size testing we define is actually a special case of the one studied in [9]. The version we define is solvable via the following simple algorithm making $O(\sqrt{n})$ queries: pick a random range item, and Grover search for an instance of that range item. The fact that our lower bound holds even for this special case of the problem considered in prior work obviously only makes our lower bound stronger.

This lower bound also serves as a starting point to establish the next three lower bounds.

***k*-Junta testing.** In this problem, we are given the truth table of a Boolean function and have to determine if the function depends on at most k variables or if it is ε-far from any such function.

The best classical algorithm for this problem uses $O(k \log k + k/\varepsilon)$ queries [19]. The problem was first studied in the quantum setting by Atıcı and Servedio [11], who gave a quantum algorithm making $O(k/\varepsilon)$ queries. This was later improved by Ambainis et al. [9] to $\tilde{O}(\sqrt{k/\varepsilon})$. They also proved a lower bound of $\Omega(k^{1/3})$. Via a connection established by Ambainis et al., our image size testing lower bound implies a $\tilde{\Omega}(\sqrt{k})$ lower bound on the approximate degree and quantum query complexity of k-junta testing (for some $\varepsilon = \Omega(1)$).

Theorem 1.4 (Informal)**.** *The approximate degree and quantum query complexity of k-junta testing is* $\tilde{\Omega}(\sqrt{k})$.

This matches the upper bound of [9], resolving the main open question from their paper.

Statistical distance from uniform (SDU)**.** In this problem, we are given a list of N numbers in $[R]$, which we interpret as a probability distribution p, where $p_i = f_i/N$, the fraction of times i appears. The goal is to compute the statistical distance between p and the uniform distribution to error ε.

This problem was studied by Bravyi, Harrow, and Hassidim [23], who gave an $O(\sqrt{n})$-query quantum algorithm approximating the statistical distance between two input distributions to additive error $\varepsilon = \Omega(1)$. We show that the approximate degree and quantum query complexity of this task are $\tilde{\Omega}(\sqrt{n})$, even when one of the distributions is known to be the uniform distribution.

Theorem 1.5 (Informal)**.** *There is a constant* $c > 0$ *such that the approximate degree and quantum query complexity of approximating the statistical distribution of a distribution over a range of size n from the uniform distribution over the same range to additive error c is is* $\tilde{\Omega}(\sqrt{n})$.

This matches the upper bound of Bravyi et al. [23] and answers the main question left open from that work. Note that the classical query complexity of this problem is $\Theta(n/\log n)$ [89].

Entropy approximation. As in the previous problem, we interpret the input as a probability distribution, and the goal is to compute its Shannon entropy to additive error ε. The classical query complexity of this problem is $\Theta(n/\log n)$ [89]. We show that, for some $\varepsilon = \Omega(1)$, the approximate degree and quantum query complexity are $\tilde{\Omega}(\sqrt{n})$.

Theorem 1.6 (Informal)**.** *There is a constant $\varepsilon > 0$ such that the approximate degree and quantum query complexity of approximating the Shannon entropy of a distribution over a range of size n to additive error ε is is $\tilde{\Omega}(\sqrt{n})$.*

This too is tight, answering a question of Li and Wu [58].

1.2 Prior work on lower bounds for approximate degree

A relatively new lower-bound technique for approximate degree called the *method of dual polynomials* plays an essential role in our paper. This method of dual polynomials dates back to work of Sherstov [75] and Špalek [82], though dual polynomials had been used earlier to resolve longstanding questions in communication complexity [36, 57, 71, 72, 81]. To prove a lower bound for a function f via this method, one exhibits an explicit *dual polynomial* for f, which is a dual solution to a certain linear program capturing the approximate degree of f.

A notable feature of the method of dual polynomials is that it is lossless, in the sense that it can exhibit a tight lower bound on the approximate degree of any function f (though actually applying the method to specific functions may be highly challenging). Prior to the method of dual polynomials, the primary tool available for proving approximate degree lower bounds was symmetrization, introduced by Minsky and Papert [60] in the 1960s. Although powerful, symmetrization is not a lossless technique.

Most prior work on the method of dual polynomials can be understood as establishing hardness amplification results. Such results show how to take a function f that is "somewhat hard" to approximate by low-degree polynomials, and turn f into a related function g that is much harder to approximate. Here, harder means either that g requires larger degree to approximate to the same error as f, or that approximations to g of a given degree incur much larger error than do approximations to f of the same degree.

Results for block-composed functions. Until very recently, the method of dual polynomials had been used exclusively to prove hardness amplification results for *block-composed* functions. That is, the harder function g would be obtained by block-composing f with another function h, i. e., $g = h \circ f$. Here, a function $g\colon \{-1,1\}^{n\cdot m} \to \{-1,1\}$ is the block-composition of $h\colon \{-1,1\}^n \to \{-1,1\}$ and $f\colon \{-1,1\}^m \to \{-1,1\}$ if g interprets its input as a sequence of n blocks, applies f to each block, and then feeds the n outputs into h.

The method of dual polynomials turns out to be particularly suited to analyzing block-composed functions, as there are sophisticated ways of "combining" dual witnesses for h and f individually to give an effective dual witness for $h \circ f$ [21, 29, 30, 73, 75, 79–81, 86]. Prior work on analyzing block-composed functions has, for example, resulted in the determination of the approximate degree of the function $f(x) = \bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} x_{ij}$, known as the AND-OR tree, which had been open for 19 years [29, 73], established new lower bounds for AC^0 under basic complexity measures including discrepancy [30, 79, 80, 86], sign-rank [32], and threshold degree [79, 80], and resolved a number of open questions about the power of statistical zero knowledge proofs [21].

Beyond block-composed functions. While the aforementioned results led to considerable progress in complexity theory, many basic questions require understanding the approximate degree of *non*-block-composed functions. One prominent example with many applications is to exhibit an AC^0 circuit over n variables with approximate degree $\Omega(n)$. Until very recently, the best result in this direction was Aaronson and Shi's well-known $\tilde{\Omega}(n^{2/3})$ lower bound on the approximate degree of the element distinctness function (which is equivalent to k-distinctness for $k=2$) [3]. However, Bun and Thaler [33] recently achieved a near-resolution of this problem by proving the following theorem.

Theorem 1.7 (Bun and Thaler [33])**.** *For any constant $\delta > 0$, there is an AC^0 circuit with approximate degree $\Omega(n^{1-\delta})$.*

The reason that Theorem 1.7 *required* moving beyond block-composed functions is the following result of Sherstov [76].

Theorem 1.8 (Sherstov)**.** *For any Boolean functions f and h, $\widetilde{\deg}(h \circ f) = O\left(\widetilde{\deg}(h) \cdot \widetilde{\deg}(f)\right)$.*

Theorem 1.8 implies that the approximate degree of $h \circ f$ (viewed as a function of its input size) is never higher than the approximate degree of f or h individually (viewed as a function of their input sizes). For example, if f and h are both functions on n inputs, and both have approximate degree $O(n^{1/2})$, then $h \circ f$ has $N := n^2$ inputs, and by Theorem 1.8, $\widetilde{\deg}(h \circ f) = O(n^{1/2} \cdot n^{1/2}) = O(N^{1/2})$.

This means that block-composing multiple AC^0 functions does not result in a function of higher approximate degree (as a function of its input size) than that of the individual functions. Bun and Thaler [33] overcome this hurdle by introducing a way of analyzing functions that cannot be written as a block-composition of simpler functions.

Bun and Thaler's techniques set the stage to finding the approximate degrees of many basic functions using the method of dual polynomials. However, they were not refined enough to accomplish this on their own. Our lower bounds in this paper are obtained by refining and extending the methods of [33].

1.3 Our techniques

In order to describe our techniques, it is helpful to explain the process by which we discovered the tight $\tilde{\Theta}(n^{3/4})$ lower and upper bounds for surjectivity (cf. Theorem 1.1). It has previously been observed [29, 33, 86] that optimal dual polynomials for a function f tend to be tailored (in a sense that can be made precise via complementary slackness) to showing optimality of some specific approximation technique for f. Hence, constructing a dual polynomial for f can provide a strong hint as to how to construct an optimal approximation for f, and vice versa.

Upper bound for surjectivity. In [33], Bun and Thaler constructed a dual polynomial witnessing a suboptimal bound of $\tilde{\Omega}(n^{2/3})$ for SURJ. Even though this dual polynomial is suboptimal, it still provided a major clue as to what an optimal approximation for SURJ should look like: it curiously ignored all inputs failing to satisfy the following condition.

Condition 1.9. Every range item has frequency at most T, for a specific threshold $T = O(N^{1/3}) \ll N$.

This suggested that an optimal approximation for SURJ should treat inputs satisfying Condition 1.9 differently than other inputs, leading us to the following multi-phase construction (for clarity and brevity, this overview is simplified). The first phase constructs a polynomial p of degree $O(n^{3/4})$ approximating SURJ on all inputs satisfying Condition 1.9. However, p may be *exponentially large* on other inputs. The second phase constructs a polynomial q of degree $O(n^{3/4})$ that is *exponentially small* on inputs x that do *not* satisfy Condition 1.9 (in particular, $q(x) \ll 1/p(x)$ for such x), and is close to 1 otherwise. The product $p \cdot q$ still approximates SURJ on inputs satisfying Condition 1.9, and is *exponentially small* on all other inputs. Notice that $\deg(p \cdot q) \leq \deg(p) + \deg(q) = O(n^{3/4})$. Combining the above with an additional averaging step (the details of which we omit from this introduction) yields an approximation to SURJ that is accurate on *all* inputs.

Lower bound for surjectivity. With the $O(n^{3/4})$ upper bound in hand, we were able to identify the fundamental bottleneck preventing further improvement of the upper bound. This suggested a way to refine the techniques of [33] to prove a matching lower bound. Once the tight lower bound for SURJ was established, we were able to identify additional refinements to analyze the other functions that we consider. We now describe this in more detail.

Bun and Thaler's [33] (suboptimal) lower bound analysis for SURJ proceeds in two stages. In the first stage, proving a lower bound for SURJ (on N input list items and R range items) is reduced to the problem of proving a lower bound for the *block-composed* function $\mathsf{AND}_R \circ \mathsf{OR}_N$,[2] under the promise that the input has Hamming weight at most N.[3] In this paper, we use this stage of their analysis unmodified.

The second stage proves an $\tilde{\Omega}(R^{2/3})$ lower bound for the latter problem by leveraging much of the machinery developed to analyze the approximate degree of block-composed functions [29, 65, 73]. To describe this machinery, we require the following notion. A dual polynomial that witnesses the fact that $\widetilde{\deg}_\varepsilon(f_n) \geq d$ is a function $\psi\colon \{-1,1\}^n \to [-1,1]$ satisfying the following three conditions.

- $\sum_{x \in \{-1,1\}^n} \psi(x) \cdot f(x) > \varepsilon$. If ψ satisfies this condition, it is said to be *well correlated* with f.
- $\sum_{x \in \{-1,1\}^n} |\psi(x)| = 1$. If ψ satisfies this condition, it is said to have ℓ_1-norm equal to 1.
- 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.

In more detail, the second stage of the analysis from [33] itself proceeds in two steps. First, the authors consider a dual witness ψ for the high approximate degree of $\mathsf{AND}_R \circ \mathsf{OR}_N$ that was constructed in prior work [29]. ψ is constructed by taking dual witnesses ϕ and γ for the high approximate degrees of AND_R and OR_N individually, and "combining" them in a specific way [54, 75, 81] to obtain a dual witness for the high approximate degree of their block composition $\mathsf{AND}_R \circ \mathsf{OR}_N$.

[2]When it is not clear from context, we use subscripts to denote the number of variables on which a function is defined.

[3]Note that a reduction the other direction is straightforward: to approximate SURJ, it suffices 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 input bits $i \in [N]$) 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 inputs, and exactly N of these predicates (one for each $i \in [N]$) evaluate to TRUE.

Unfortunately, ψ only witnesses a lower bound for $\mathsf{AND}_R \circ \mathsf{OR}_N$ *without* the promise that the Hamming weight of the input is at most N. To address this issue, it is enough to "post-process" ψ so that it no longer "exploits" any inputs of Hamming weight larger than N (formally, $\psi(x)$ should equal zero for any inputs in $\{-1,1\}^{R\cdot N}$ of Hamming weight more than N). The authors accomplish this by observing that ψ "almost ignores" all such inputs (i. e., it places exponentially little total mass on all such inputs), and hence it is possible to perturb ψ to make it *completely* ignore all such inputs.

Key to this step is the fact that the "inner" dual witness γ for the high approximate degree of the OR_N function satisfies a "Hamming weight decay" condition:

$$|\gamma(x)| \cdot \binom{N}{|x|} \leq \frac{1}{\mathrm{poly}(|x|)}, \tag{1.1}$$

for a suitable polynomial function.

To improve the lower bound for SURJ from $\tilde{\Omega}(n^{2/3})$ to the optimal $\tilde{\Omega}(n^{3/4})$, we observe that γ in fact satisfies a much stronger decay condition: while the inverse-polynomial decay property of Equation (1.1) is tight for small Hamming weights $|x|$, $|\gamma(x)|$ actually decays *exponentially quickly* once $|x|$ is larger than a certain threshold t. This observation is enough to obtain the tight $\tilde{\Omega}(n^{3/4})$ lower bound for SURJ.

For intuition, it is worth mentioning that a primal formulation of the dual decay condition that we exploit shows that any low-degree polynomial p that is an accurate approximation to OR_N on low Hamming-weight inputs requires large degree, *even if* $|p(x)|$ *is allowed to be exponentially large for inputs of Hamming weight more than* t.[4] This is precisely the bottleneck that prevents us from improving our upper bound for SURJ to $o(N^{3/4})$. In this sense, our dual witness is intuitively tailored to showing optimality of the techniques used in our upper bound.

Other lower bounds. To obtain the lower bound for k-distinctness, the first stage of the analysis of [33] reduces to a question about the approximate degree of the block-composed function $\mathsf{OR}_R \circ \mathsf{THR}_N^k$, under the promise that the input has Hamming weight at most N. Here $\mathsf{THR}_N^k \colon \{-1,1\}^N \to \{-1,1\}$ denotes the function that evaluates to -1 if and only if the Hamming weight of its input is at least k. By constructing a suitable dual witness for THR_N^k, and combining it with a dual witness for OR_N via similar techniques as in our construction for SURJ, we are able to prove our $\Omega(n^{3/4-1/(2k)})$ lower bound for k-distinctness. (This description glosses over several significant technical issues that must be dealt with to ensure that the combined dual witness is well-correlated with $\mathsf{OR}_R \circ \mathsf{THR}_N^k$).[5]

Recall that our lower bounds for k-junta testing, SDU, and entropy approximation are derived as consequences of our lower bound for image size testing. The connection between image size testing and

[4] We do not formally describe this primal formulation of the dual decay condition, because it is not necessary to prove any of the results in this paper.

[5] Specifically, our analysis requires the dual witness γ for THR_N^k to be very well-correlated with THR_N^k in a certain one-sided sense. Roughly, we need the probability distribution $|\gamma|$ to have the property that, conditioned on γ outputting a negative value, the input to γ is in $\left(\mathsf{THR}_N^k\right)^{-1}(-1)$ with probability at least $1-1/(3R)$. This property was not required in the analysis for SURJ, which is why our lower bound for SURJ is larger by a factor of $n^{1/(2k)}$ than our lower bound for k-distinctness (recall that $n = N\log_2 R$ and N is a constant factor larger than R). This seemingly technical issue is at least partially intrinsic: a polynomial loss compared to the $\tilde{\Omega}(n^{3/4})$ lower bound for SURJ is unavoidable, owing to Belovs' $n^{3/4-\Omega(1)}$ upper bound [16] for k-distinctness.

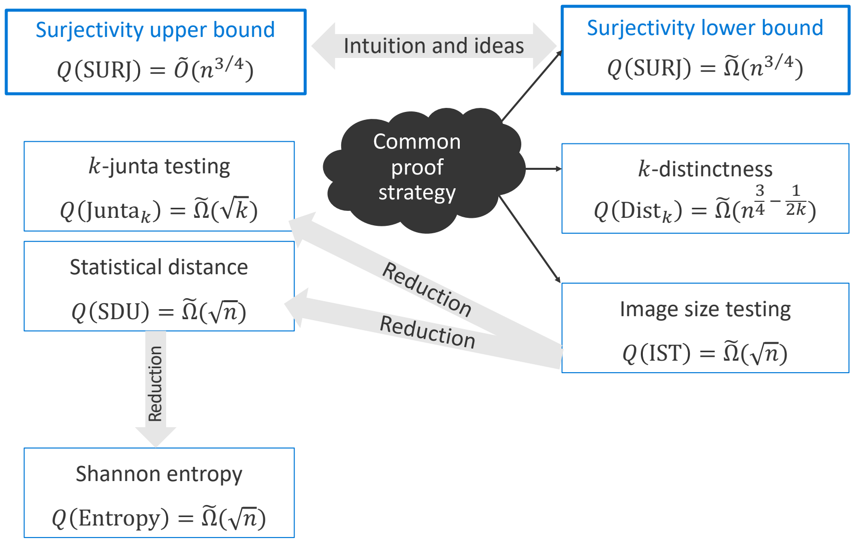

Figure 1: Diagram of reductions and relationships amongst our results.

junta testing was established by Ambainis et al. [9]. The reason that the image testing lower bound implies lower bounds for SDU is the following. Consider any distribution p over $[R]$ such that all probabilities p_i are integer multiples of $1/N$ for some $N = O(R)$. Then if p has full support, p is guaranteed to be somewhat close to uniform, while if p has small support, p must be very far from uniform. We obtain our lower bound for entropy approximation using a simple reduction from SDU due to Vadhan [88].

To obtain our lower bound for image size testing, we observe that the first stage of the analysis of [33] reduces to a question about the approximate degree of the block-composed function $\mathsf{GapAND}_R \circ \mathsf{OR}_N$, under the promise that the input has Hamming weight at most N. Here, GapAND_R is the promise function that outputs -1 if all of its inputs equal -1, outputs $+1$ if fewer than $\gamma \cdot R$ of its inputs are -1, and is undefined otherwise.

Roughly speaking, we obtain the desired $\tilde{\Omega}(n^{1/2})$ lower bound by combining a dual witness for $\mathsf{GapAND}_R \circ \mathsf{OR}_N$ from prior work [30] with the same techniques as in our construction for SURJ. However, additional technical refinements to the analysis of [33] are required to obtain our results. In particular, the analysis of [33] only provides a lower bound for SURJ if $N = \Omega(R \cdot \log^2(R))$. But in order to infer our lower bound for SDU and entropy approximation (as well as k-junta testing for $\varepsilon = \Omega(1)$), it is essential that the lower bound hold for $N = O(R)$. This is because a distribution with full support is guaranteed to be $\Omega(1)$-close to uniform if all probabilities are integer multiples of $1/N$ with $N = O(R)$, but this is not the case otherwise. (Consider, e. g., a distribution that places mass $1 - 1/\log^2(R)$ on a single range item, and spreads out the remaining mass evenly over all other range items). Refining the methods of [33] to yield lower bounds even when $N = O(R)$ requires a significantly more delicate analysis than in [33].

A diagram indicating how we obtain our results and the relationships that we establish between problems is given in Figure 1.

1.4 Outline for the rest of the paper

Section 2 covers preliminary definitions and lemmas. Section 3 presents the $\tilde{O}(n^{3/4})$ upper bound for SURJ, while Section 4 presents the matching $\tilde{\Omega}(n^{3/4})$ lower bound. Section 5 gives the $\tilde{\Omega}(n^{3/4-1/(2k)})$ lower bound for k-distinctness. Section 6 presents the lower bound for image size testing, and its implications for junta testing, SDU, and Shannon entropy approximation. Section 7 concludes by briefly describing some additional consequences of our results, as well as a number of open questions and directions for future work.

2 Preliminaries

2.1 Notation

For a natural number N, let $[N] = \{1,2,\ldots,N\}$ and $[N]_0 = \{0,1,2,\ldots,N\}$. All logarithms are taken in base 2 unless otherwise noted. As is standard, we say that a function $f(n)$ is in $\tilde{O}(h(n))$ if there exists a constant k such that $f(n)$ is in $O(h(n)\cdot\log^k(h(n)))$.

We will frequently work with Boolean functions under the promise that their inputs have low Hamming weight. To this end, we introduce the following notation for the set of low Hamming-weight inputs.

Definition 2.1. For $1 \le T \le n$, let $H^n_{\le T}$ denote the subset of $\{-1,1\}^n$ consisting of all inputs of Hamming weight at most T. We use $|x|$ to denote the Hamming weight of an input $x \in \{-1,1\}^n$, so $H^n_{\le T} = \{x \in \{-1,1\}^n : |x| \le T\}$.

2.2 Two variants of approximate degree and their dual formulations

There are two natural notions of approximate degree for promise problems (i. e., for functions defined on a strict subset $\mathcal{X}$ of $\{-1,1\}^n$). One notion requires an approximating polynomial p to be bounded in absolute value even on inputs in $\{-1,1\}^n \setminus \mathcal{X}$. The other places no restrictions on p outside of the promise $\mathcal{X}$. We make use of *both* notions. Hence, we must introduce some (non-standard) notation to distinguish the two.

Definition 2.2 (Approximate degree with boundedness outside of the promise required)**.** Let $\varepsilon > 0$ and $\mathcal{X} \subseteq \{-1,1\}^n$. The ε*-approximate degree* of a Boolean function $f : \mathcal{X} \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial $p : \mathcal{X} \to \mathbb{R}$ such that $|p(x) - f(x)| \le \varepsilon$ for all $x \in \mathcal{X}$ and $|p(x)| \le 1+\varepsilon$ for all $x \in \{-1,1\}^n \setminus \mathcal{X}$. We use the term *approximate degree* without qualification to refer to $\widetilde{\deg}(f) = \widetilde{\deg}_{1/3}(f)$.

The following standard dual formulation of this first variant of approximate degree can be found in, e. g., [31].

Proposition 2.3. *Let $\mathcal{X} \subseteq \{-1,1\}^n$, and let $f : \mathcal{X} \to \{-1,1\}$. Then $\widetilde{\deg}_\varepsilon(f) \ge d$ if and only if there* *exists a function* $\psi\colon \{-1,1\}^n \to \mathbb{R}$ *with the following properties.*

$$\sum_{x\in\mathcal{X}} \psi(x)\cdot f(x) - \sum_{x\in\{-1,1\}^n\setminus\mathcal{X}} |\psi(x)| > \varepsilon, \tag{2.1}$$

$$\sum_{x\in\{-1,1\}^n} |\psi(x)| = 1, \text{ and} \tag{2.2}$$

$$\text{for every polynomial } p\colon \{-1,1\}^n \to \mathbb{R} \text{ of degree less than } d, \sum_{x\in\{-1,1\}^n} p(x)\cdot\psi(x) = 0. \tag{2.3}$$

We will refer to functions $\psi\colon \{-1,1\}^n \to \mathbb{R}$ as *dual polynomials*. We refer to $\sum_{x\in\{-1,1\}^n} |\psi(x)|$ as the ℓ_1-norm of ψ, and denote this quantity by $\|\psi\|_1$. If ψ satisfies Equation (2.3), it is said to have *pure high degree* at least d.

Given a function $\psi\colon \{-1,1\}^n \to \mathbb{R}$, and a (possibly partial) function $f\colon \mathcal{X} \to \{-1,1\}$, where $\mathcal{X} \subseteq \{-1,1\}^n$, we let $\langle f,\psi\rangle := \sum_{x\in\mathcal{X}} f(x)\cdot\psi(x) - \sum_{x\in\{-1,1\}^n\setminus\mathcal{X}} |\psi(x)|$, and refer to this as the correlation of f and ψ. So Condition (2.1) is equivalent to requiring ψ and f to have correlation great than ε.

Definition 2.4 (Approximate degree with unboundedness permitted outside of the promise)**.** Let $\varepsilon > 0$ and $\mathcal{X}$ be a finite set. The ε-*unbounded approximate degree* of a Boolean function $f : \mathcal{X} \to \{-1,1\}$, denoted $\widetilde{\mathrm{ubdeg}}_\varepsilon(f)$, is the least degree of a real polynomial $p : \mathcal{X} \to \mathbb{R}$ such that $|p(x) - f(x)| \le \varepsilon$ for all $x \in \mathcal{X}$ (if $\mathcal{X}$ is a strict subset of a larger domain, then no constraints are placed on $p(x)$ for $x \notin \mathcal{X}$). We use the term *unbounded approximate degree* without qualification to refer to $\widetilde{\mathrm{ubdeg}}(f) = \widetilde{\mathrm{ubdeg}}_{1/3}(f)$.

The following standard dual formulation of this second variant of approximate degree can be found in, e. g., [72]. A dual polynomial $\psi : \{-1,1\}^n \to [-1,1]$ witnessing the fact that $\widetilde{\mathrm{ubdeg}}_\varepsilon(f) \ge d$ is the same as a dual witness for $\widetilde{\deg}_\varepsilon(f) \ge d$, but with the additional requirement that $\psi(x) = 0$ outside of $\mathcal{X}$.

Proposition 2.5. *Let* $\mathcal{X} \subseteq \{-1,1\}^n$*, and let* $f : \mathcal{X} \to \{-1,1\}$*. Then* $\widetilde{\mathrm{ubdeg}}_\varepsilon(f) \ge d$ *if and only if there exists a function* $\psi\colon \{-1,1\}^n \to \mathbb{R}$ *satisfying the following conditions.*

$$\psi(x) = 0 \text{ for all } x \notin \mathcal{X}, \tag{2.4}$$

$$\sum_{x\in\mathcal{X}} \psi(x)\cdot f(x) > \varepsilon, \tag{2.5}$$

$$\sum_{x\in\{-1,1\}^n} |\psi(x)| = 1, \text{ and} \tag{2.6}$$

$$\text{For every polynomial } p\colon \{-1,1\}^n \to \mathbb{R} \text{ of degree less than } d, \sum_{x\in\{-1,1\}^n} p(x)\cdot\psi(x) = 0. \tag{2.7}$$

Observe that $\widetilde{\deg}(f)$ and $\widetilde{\mathrm{ubdeg}}(f)$ coincide for total functions f. To avoid notational clutter, when referring to the approximate degree of total functions, we will use the shorter notation $\widetilde{\deg}(f)$.

2.3 Basic facts about polynomial approximations

The seminal paper of Nisan and Szegedy [62] gave tight bounds on the approximate degree of the AND_n and OR_n functions.

Lemma 2.6. *For any constant $\varepsilon \in (0,1)$, the functions* AND *and* OR *on n bits have ε-approximate degree $\Theta(n^{1/2})$, and the same holds for their negations.*

Approximate degree is invariant under negating the inputs or output of a function, and hence the result for AND implies the result for NAND, OR, etc.

The following lemma, which forms the basis of the well-known symmetrization argument, is due to Minsky and Papert [60].

Lemma 2.7. *Let $p\colon \{-1,1\}^n \to \mathbb{R}$ be an arbitrary polynomial and let $[n]_0$ denote the set $\{0,1,\ldots,n\}$. Then there is a univariate polynomial $q\colon \mathbb{R} \to \mathbb{R}$ of degree at most* $\deg(p)$ *such that*

$$q(t) = \frac{1}{\binom{n}{t}} \sum_{x \in \{-1,1\}^n \colon |x|=t} p(x)$$

for all $t \in [n]_0$.

2.4 Functions of interest

We give formal definitions of the surjectivity and k-distinctness we consider, as well as several variations that will be helpful in proving our lower bounds. The formal definitions of the other functions we study, including image size testing, SDU, and the Shannon entropy functions, are deferred to Section 6.

2.4.1 Surjectivity

Definition 2.8. For $N \geq R$, we define $\mathsf{SURJ}_{N,R} : [R]^N \to \{-1,1\}$ by $\mathsf{SURJ}_{N,R}(s_1,\ldots,s_N) = -1$ iff for every $j \in [R]$, there exists an i such that $s_i = j$.

When N and R are clear from context, we will often refer to the function SURJ without the explicit dependence on these parameters. It will sometimes be convenient to think of the input to $\mathsf{SURJ}_{N,R}$ as a function mapping $\{-1,1\}^n \to \{-1,1\}$ rather than $[R]^N \to \{-1,1\}$. When needed, we assume that R is a power of 2 and an element of $[R]$ is encoded in binary using $\log R$ bits. In this case we will view surjectivity as a function on $n = N \log R$ bits, i. e., $\mathsf{SURJ} : \{-1,1\}^n \to \{-1,1\}$.

For technical reasons, when proving lower bounds, it will be more convenient to work with a variant of SURJ where the range $[R]$ is augmented by a "dummy element" 0 that is simply ignored by the function. That is, while any of the items $s_1,\ldots,s_N$ may take the dummy value 0, the presence of a 0 in the input is not required for the input to be deemed surjective. We denote this variant of surjectivity by dSURJ. More formally:

Definition 2.9. For $N \geq R$, we define $\mathsf{dSURJ}_{N,R} : [R]_0^N \to \{-1,1\}$ by $\mathsf{dSURJ}_{N,R}(s_1,\ldots,s_N) = -1$ iff for every $j \in [R]$, there exists an i such that $s_i = j$.

The following simple reduction shows that a lower bound on the approximate degree of dSURJ implies a lower bound for SURJ itself.

Proposition 2.10. *Let $\varepsilon > 0$ and $N \geq R$. Then*

$$\widetilde{\deg}_\varepsilon(\mathsf{dSURJ}_{N,R}) \leq \widetilde{\deg}_\varepsilon(\mathsf{SURJ}_{N+1,R+1}) \cdot \log(R+1).$$

Proof. Let $p : \{-1,1\}^{(N+1)\cdot\log(R+1)} \to \{-1,1\}$ be a polynomial of degree d that ε-approximates $\mathsf{SURJ}_{N+1,R+1}$. We will use p to construct a polynomial of degree d that ε-approximates $\mathsf{dSURJ}_{N,R}$. Recall that an input to $\mathsf{dSURJ}_{N,R}$ takes the form $(s_1,\dots,s_N)$ where each s_i is the binary representation of a number in $[R]_0$. Define the transformation $T : [R]_0 \to [R+1]$ by

$$T(s) = \begin{cases} R+1 & \text{if } s = 0, \\ s & \text{otherwise.} \end{cases}$$

Note that as a mapping between binary representations, the function T is exactly computed by a vector of polynomials of degree at most $\log(R+1)$. For every $(s_1,\dots,s_N) \in [R]_0^N$, observe that

$$\mathsf{dSURJ}_{N,R}(s_1,\dots,s_N) = \mathsf{SURJ}_{N+1,R+1}(T(s_1),\dots,T(s_N),R+1).$$

Hence, the polynomial

$$p(T(s_1),\dots,T(s_N),R+1)$$

is a polynomial of degree $d\cdot\log(R+1)$ that ε-approximates $\mathsf{dSURJ}_{N,R}$. □

2.4.2 k-Distinctness

Definition 2.11. For integers k,N,R with $k \le N$, define the function $\mathsf{DIST}^k_{N,R} : [R]^N \to \{-1,1\}$ by $\mathsf{DIST}^k_{N,R}(s_1,\dots,s_N) = -1$ iff there exist $r \in [R]$ and distinct indices $i_1,\dots,i_k$ such that $s_{i_1} = \cdots = s_{i_k} = r$.

As with surjectivity, it will be convenient to work with a variant of k-distinctness where $[R]$ is augmented with a dummy item:

Definition 2.12. For integers k,N,R with $k \le N$, define the function $\mathsf{dDIST}^k_{N,R} : [R]_0^N \to \{-1,1\}$ by $\mathsf{dDIST}^k_{N,R}(s_1,\dots,s_N) = -1$ iff there exist $r \in [R]$ and distinct indices $i_1,\dots,i_k$ such that $s_{i_1} = \cdots = s_{i_k} = r$.

For $k \ge 2$, a lower bound on the approximate degree of dDIST^k implies a lower bound on the approximate degree of DIST^k. The restriction that $k \ge 2$ is essential, because the function DIST^1 is the constant function that evaluates to TRUE on *any* input (since at least one range item must always have frequency at least one), whereas dDIST^1 contains OR_N as a subfunction, and hence has approximate degree at least $\Omega(\sqrt{N})$.

Proposition 2.13. *Let $\varepsilon > 0$, $N,R \in \mathbb{N}$, and $k \ge 2$. Then*

$$\deg_\varepsilon(\mathsf{dDIST}^k_{N,R}) \le \deg_\varepsilon(\mathsf{DIST}^k_{N,R+N})\cdot\log(R+1).$$

Proof. The proof is similar to that of Proposition 2.10, but uses a slightly more involved reduction. Let $p : \{-1,1\}^{2N\cdot\log(R+N)} \to \mathbb{R}$ be a polynomial of degree d that ε-approximates $\mathsf{DIST}^k_{N,R+N}$. We will use p to construct a polynomial of degree d that ε-approximates $\mathsf{dDIST}^k_{N,R}$. For each $i = 1,\dots,R$, define a transformation $T_i : [R]_0 \to [R+N]$ by

$$T_i(s) = \begin{cases} R+i & \text{if } s = 0, \\ s & \text{otherwise.} \end{cases}$$

As a mapping between binary representations, the function $T = (T_1, \ldots, T_N)$ is exactly computed by a vector of polynomials of degree at most $\log(R+1)$. For every $(s_1, \ldots, s_N) \in [R]_0^N$, observe that

$$\mathsf{dDIST}^k_{N,R}(s_1, \ldots, s_N) = \mathsf{DIST}^k_{N,R+N}(T_1(s_1), \ldots, T_N(s_N)).$$

Hence, the polynomial

$$p(T(s_1), \ldots, T(s_N))$$

is a polynomial of degree $d \cdot \log(R+1)$ that ε-approximates $\mathsf{dSURJ}_{N,R}$. $\square$

2.5 Connecting symmetric properties and block-composed functions

An important ingredient in [33] is the relationship between the approximate degree of a property of a list of numbers (such as SURJ) and the approximate degree of a simpler block-composed function, defined as follows.

Definition 2.14. For functions $f : Y^n \to Z$ and $g : X \to Y$, define the *block composition* $f \circ g : X^n \to Z$ by $(f \circ g)(x_1, \ldots, x_n) = f(g(x_1), \ldots, g(x_n))$, for all $x_1, \ldots, x_n \in X$.

Fix $R, N \in \mathbb{N}$, let $f : \{-1,1\}^R \to \{-1,1\}$ and let $g : \{-1,1\}^N \to \{-1,1\}$. Suppose g is a *symmetric* function, in the sense that for any $x \in \{-1,1\}^N$ and any permutation $\sigma : [N] \to [N]$, we have

$$g(x_1, \ldots, x_N) = g(x_{\sigma(1)}, \ldots, x_{\sigma(N)}).$$

Equivalently, the value of g on any input x depends only on its Hamming weight $|x|$.

The functions f and g give rise to two functions. The first, which we denote by $F^{\text{prop}} : [R]_0^N \to \{-1,1\}$, is a certain property of a list of numbers $s_1, \ldots, s_N \in [R]_0$. The second, which we denote by $F^{\leq N} : H^{N \cdot R}_{\leq N} \to \{-1,1\}$, is the block composition of f and g restricted to inputs of Hamming weight at most N. Formally, these functions are defined as

$$F^{\text{prop}}(s_1, \ldots, s_N) = f(g(\mathbb{1}[s_1 = 1], \ldots, \mathbb{1}[s_N = 1]), \ldots, g(\mathbb{1}[s_1 = R], \ldots, \mathbb{1}[s_N = R]))$$
$$F^{\leq N}(x_1, \ldots, x_R) = \begin{cases} f(g(x_1), \ldots, g(x_R)) & \text{if } x_1, \ldots, x_R \in \{-1,1\}^N, |x_1| + \cdots + |x_R| \leq N, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

The following proposition from [33], which in turn relies heavily on a clever symmetrization argument due to due to Ambainis [6], relates the approximate degrees of the two functions F^{prop} and $F^{\leq N}$.

Theorem 2.15. *Let $f : \{-1,1\}^R \to \{-1,1\}$ be any function and let $g : \{-1,1\}^N \to \{-1,1\}$ be a symmetric function. Then for F^{prop} and $F^{\leq N}$ defined above, and for any $\varepsilon > 0$, we have*

$$\deg_\varepsilon(F^{\text{prop}}) \geq \widetilde{\text{ubdeg}}_\varepsilon(F^{\leq N}).$$

In the case where $f = \mathsf{AND}_R$ and $g = \mathsf{OR}_N$, the function $F^{\text{prop}}(s_1, \ldots, s_N)$ is the surjectivity function augmented with a dummy item, $\mathsf{dSURJ}_{N,R}(s_1, \ldots, s_N)$. Hence,

Corollary 2.16. *Let $N, R \in \mathbb{N}$. Then for any $\varepsilon > 0$,*

$$\widetilde{\deg}_\varepsilon(\mathsf{dSURJ}_{N,R}) \geq \widetilde{\mathrm{ubdeg}}_\varepsilon(F^{\leq N})$$

where $F^{\leq N} : H^{N \cdot R}_{\leq N} \to \{-1,1\}$ is the restriction of $\mathsf{AND}_R \circ \mathsf{OR}_N$ to $H^{N \cdot R}_{\leq N}$.

Definition 2.17. For integers k, N with $k \leq N$, define the function $\mathsf{THR}^k_N : \{-1,1\}^N \to \{-1,1\}$ by $\mathsf{THR}^k_N(x) = -1$ iff $|x| \geq k$.

If we let $f = \mathsf{OR}_R$ and $g = \mathsf{THR}^k_N$, then the function F^{prop} is the dummy augmented k-distinctness function $\mathsf{dDIST}^k_{N,R}$.

Corollary 2.18. *Let $N, R \in \mathbb{N}$. Then for any $\varepsilon > 0$,*

$$\widetilde{\deg}_\varepsilon(\mathsf{dDIST}^k_{N,R}) \geq \widetilde{\mathrm{ubdeg}}_\varepsilon(G^{\leq N})$$

where $G^{\leq N} : H^{N \cdot R}_{\leq N} \to \{-1,1\}$ is the restriction of $\mathsf{OR}_R \circ \mathsf{THR}^k_N$ to $H^{N \cdot R}_{\leq N}$.

2.6 The dual block method

This section collects definitions and preliminary results on the dual block method [54, 75, 81] for constructing dual witnesses for a block-composed function $F \circ f$ by combining dual witnesses for F and f.

Definition 2.19. Let $\Psi : \{-1,1\}^M \to \mathbb{R}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ be functions that are not identically zero. Let $x = (x_1, \ldots, x_M) \in (\{-1,1\}^m)^M$. Define the *dual block composition* of Ψ and ψ, denoted $\Psi \star \psi : (\{-1,1\}^m)^M \to \mathbb{R}$, by

$$(\Psi \star \psi)(x_1, \ldots, x_M) = 2^M \cdot \Psi(\ldots, \mathrm{sgn}\,(\psi(x_i)), \ldots) \cdot \prod_{i=1}^{M} |\psi(x_i)|.$$

Proposition 2.20 ([33, 75])**.** *The dual block composition satisfies the following conditions.*

***Preservation of ℓ_1-norm*:** *If $\|\Psi\|_1 = 1$, $\|\psi\|_1 = 1$, and $\langle \psi, \mathbf{1} \rangle = 0$, then*

$$\|\Psi \star \psi\|_1 = 1. \tag{2.8}$$

***Multiplicativity of pure high degree*:** *If $\langle \Psi, P \rangle = 0$ for every polynomial $P \colon \{-1,1\}^M \to \mathbb{R}$ of degree less than D, and $\langle \psi, p \rangle = 0$ for every polynomial $p \colon \{-1,1\}^m \to \mathbb{R}$ of degree less than d, then for every polynomial $q \colon \{-1,1\}^{m \cdot M} \to \mathbb{R}$,*

$$\deg q < D \cdot d \implies \langle \Psi \star \psi, q \rangle = 0. \tag{2.9}$$

***Associativity*:** *For every $\zeta : \{-1,1\}^{m_\zeta} \to \mathbb{R}$, $\varphi : \{-1,1\}^{m_\varphi} \to \mathbb{R}$, and $\psi : \{-1,1\}^{m_\psi} \to \mathbb{R}$, we have*

$$(\zeta \star \varphi) \star \psi = \zeta \star (\varphi \star \psi). \tag{2.10}$$

2.7 A refinement of a technical lemma from prior work

The following technical proposition refines techniques of Bun and Thaler [33]. This proposition is useful for "zeroing out" the mass that a dual polynomial ξ places on inputs of high Hamming weight, if ξ is obtained via the dual-block method.

Definition 2.21. Let $M \in \mathbb{N}$ and $\alpha, \beta > 0$. A function $\omega : [M]_0 \to \mathbb{R}$ satisfies the (α, β)-decay condition if

$$\sum_{t=0}^{M} \omega(t) = 0, \tag{2.11}$$

$$\sum_{t=0}^{M} |\omega(t)| = 1, \tag{2.12}$$

$$|\omega(t)| \leq \alpha \exp(-\beta t)/t^2 \quad \forall t = 1, 2, \ldots, M. \tag{2.13}$$

Proposition 2.22. *Let $R \in \mathbb{N}$ be sufficiently large, and let $\Phi : \{-1,1\}^R \to \mathbb{R}$ with $\|\Phi\|_1 = 1$. For $M \leq R$, let $\omega : [M]_0 \to \mathbb{R}$ satisfy the (α, β)-decay condition with parameters $1 \leq \alpha \leq R^2, \beta \in (4\ln^2 R/(\sqrt{\alpha}R), 1)$.*

Let $N = \lceil 20\sqrt{\alpha} \rceil R$, and define $\psi : \{-1,1\}^N \to \mathbb{R}$ by $\psi(x) = \omega(|x|)/\binom{N}{|x|}$. If $D < N$ is such that

$$\textit{For every polynomial } p \textit{ with } \deg p < D, \textit{ we have } \langle \Phi \star \psi, p \rangle = 0, \tag{2.14}$$

then there exist $\Delta \geq \beta\sqrt{\alpha}R/(4\ln^2 R)$ and a function $\zeta : (\{-1,1\}^N)^R \to \mathbb{R}$ such that

$$\textit{For every polynomial } p \textit{ with } \deg p < \min\{D, \Delta\}, \textit{ we have } \langle \zeta, p \rangle = 0, \tag{2.15}$$

$$\|\zeta - \Phi \star \psi\|_1 \leq \frac{2}{9}, \tag{2.16}$$

$$\|\zeta\|_1 = 1, \tag{2.17}$$

$$\zeta \textit{ is supported on } H_{\leq N}^{N \cdot R}. \tag{2.18}$$

The key refinement of Proposition 2.22 relative to the analysis of Bun and Thaler is that Proposition 2.22 applies when $N = \Theta(R)$ (assuming $\alpha = O(1)$). In contrast, the techniques of Bun and Thaler required $N = \Omega(R \cdot \log^2 R)$. As indicated in Section 1.3, this refinement will be essential in obtaining our lower bounds for SDU, entropy approximation, and junta testing for constant proximity parameter.

The proof of Proposition 2.22 occurs in two steps. First, in Proposition 2.23, we show that $\xi = \Phi \star \psi$ places an exponentially small amount of mass on inputs outside of $H_{\leq N}^{N \cdot R}$. Second, in Proposition 2.25, we construct a correction object ν that zeroes out the mass ξ places outside of $H_{\leq N}^{N \cdot R}$ without decreasing its pure high degree. Combining ξ with ν yields the desired object ζ.

Proposition 2.23. *Let $\Phi : \{-1,1\}^R \to \mathbb{R}$ and $\psi : \{-1,1\}^n \to \mathbb{R}$ satisfy the conditions of Proposition 2.22. Then for sufficiently large R, there exists $\Delta \geq \beta\sqrt{\alpha}R/(4\ln^2 R)$ such that, for $N = \lceil 20\sqrt{\alpha} \rceil R$,*

$$\sum_{x \notin H_{\leq N}^{N \cdot R}} |(\Phi \star \psi)(x)| \leq (2NR)^{-2\Delta}. \tag{2.19}$$

Proof. Recall that $\psi(x) = \omega(|x|)/\binom{N}{|x|}$ where $\omega \colon [M]_0 \to \mathbb{R}$. By Equation (2.11) we may write $\omega = \omega_{+1} - \omega_{-1}$ where ω_{+1} and ω_{-1} are non-negative functions satisfying

$$\sum_{t=0}^{k} \omega_{+1}(t) = \sum_{t=0}^{k} \omega_{-1}(t) = 1/2. \tag{2.20}$$

By the definition of dual block composition, we have

$$(\Phi \star \psi)(x_1, \dots, x_R) = 2^R \cdot \Phi(\dots, \operatorname{sgn}(\psi(x_i)), \dots) \cdot \prod_{i=1}^{R} |\psi(x_i)|.$$

Consequently,

$$\begin{aligned}
\sum_{x \notin H_{\leq N}^{N \cdot R}} |(\Phi \star \psi)(x)| &= 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{\substack{(x_1,\dots,x_R) \notin H_{\leq N}^{N \cdot R} \text{ s.t.} \\ \operatorname{sgn}(\psi(x_1)) = z_1, \dots, \operatorname{sgn}(\psi(x_R)) = z_R}} \prod_{i=1}^{R} |\psi(x_i)| \right) \\
&= 2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{(x_1,\dots,x_R) \notin H_{\leq N}^{N \cdot R}} \prod_{i=1}^{R} \frac{\omega_{z_i}(|x_i|)}{\binom{N}{|x_i|}} \right).
\end{aligned} \tag{2.21}$$

Observe that for any $(t_1, \dots, t_R) \in [M]_0^R$, the number of inputs $(x_1, \dots, x_R) \in \left(\{-1,1\}^N\right)^R$ such that $|x_i| = t_i$ for all $i \in [R]$ is exactly $\prod_{i=1}^{R} \binom{N}{t_i}$. Hence, defining

$$P = \{(t_1, \dots, t_R) \in [M]_0^R : t_1 + \cdots + t_R > N\},$$

we may rewrite Expression (2.21) as

$$2^R \sum_{z \in \{-1,1\}^R} |\Phi(z)| \left(\sum_{(t_1,\dots,t_R) \in P} \prod_{i=1}^{R} \omega_{z_i}(t_i) \right).$$

To control this quantity, we appeal to the following combinatorial lemma, whose proof we delay to Section 2.7.1 (this lemma is a substantial refinement of [33, Lemma 32]).

Lemma 2.24. *Let $\alpha \leq R^2$, let $\beta \in (4\ln^2 R/(\sqrt{\alpha}R), 1)$, and let R be sufficiently large. Let $N = \lceil 20\sqrt{\alpha} \rceil R$. Let $\eta_i : [M]_0 \to \mathbb{R}$, for $i = 1, \dots R$, be a sequence of non-negative functions where for every i,*

$$\sum_{r=0}^{M} \eta_i(r) \leq 1/2, \tag{2.22}$$

$$\eta_i(r) \leq \alpha \exp(-\beta r)/r^2 \qquad \forall r = 1, \dots, M. \tag{2.23}$$

Let $P = \{\vec{t} = (t_1, \dots, t_R) \in [M]_0^R : t_1 + \cdots + t_R > N\}$. Then

$$\sum_{\vec{t} \in P} \prod_{i=1}^{R} \eta_i(t_i) \leq 2^{-R} \cdot (2NR)^{-2\Delta}$$

where $\Delta \geq \beta\sqrt{\alpha}R/(4\ln^2 R)$.

Observe that the functions ω_{z_i} satisfy Condition (2.22) (cf. Equation (2.20)) and Condition (2.23) (cf. Property (2.13)). We complete the proof by bounding

$$\begin{aligned}2^R \sum_{z\in\{-1,1\}^R} |\Phi(z)| \left(\sum_{\vec{t}\in P}\prod_{i=1}^R \omega_{z_i}(t_i)\right) &\le 2^R \sum_{z\in\{-1,1\}^R} |\Phi(z)|\cdot\left(2^{-R}\cdot(2NR)^{-2\Delta}\right)\\ &= (2NR)^{-2\Delta}.\end{aligned}$$

Here, the equality appeals to the condition that $\|\Phi\|_1 = 1$. □

Our final dual witness ζ is obtained by modifying $\xi = \Phi \star \psi$ to zero out all of the mass it places on inputs of total Hamming weight larger than N. The following proposition gives sufficient conditions under which this postprocessing step can be done.

Proposition 2.25 ([33, Proposition 34], building on [65])**.** *Let $N \ge R > D$ and let $\xi : (\{-1,1\}^N)^R \to \mathbb{R}$ be any function such that*

$$\sum_{x\notin H^{N\cdot R}_{\le N}} |\xi(x)| \le (2NR)^{-2D}.$$

Then there exists a function $\nu : (\{-1,1\}^N)^R \to \mathbb{R}$ such that

$$\begin{aligned}&\textit{For all polynomials } p\colon (\{-1,1\}^N)^R \to \mathbb{R},\ \deg p < D \implies \langle \nu, p\rangle = 0,\\ &\|\nu\|_1 \le 1/10,\\ &|x| > N \implies \nu(x) = \xi(x).\end{aligned}$$

Remark 2.26. Proposition 2.25 is framed exclusively in the language of dual polynomials: it states that if a dual polynomial ξ places very little mass on inputs of Hamming weight more than N, then there exists another dual polynomial ν with certain useful properties (we ultimately use ν to zero out the mass that ξ places on inputs of Hamming weight more than N). Proposition 2.25 also has a clean *primal* formulation. Roughly speaking, it is equivalent to a bound on the growth rate of any polynomial of degree at most D that is bounded at all inputs of Hamming weight at most D. We direct the interested reader to [90] for details of this primal formulation.

We are now ready to combine Proposition 2.23 and Proposition 2.25 to complete the proof of Proposition 2.22.

Proof of Proposition 2.22. Let $\xi = \Phi \star \psi$. By Proposition 2.23, we have

$$\sum_{x\notin H^{N\cdot R}_{\le N}} |(\Phi\star\psi)(x)| \le (2NR)^{-2\Delta},$$

where $\Delta \ge \beta\sqrt{\alpha}R/(4\ln^2 R)$. By Proposition 2.25, there exists a function $\nu : (\{-1,1\}^N)^R \to \mathbb{R}$ such that

$$\text{For all polynomials } p\colon (\{-1,1\}^N)^R \to \mathbb{R},\ \deg p < \min\{D,\Delta\} \implies \langle \nu, p\rangle = 0, \tag{2.24}$$

$$\|\nu\|_1 \le 1/10, \tag{2.25}$$

$$|x| > N \implies \nu(x) = \xi(x). \tag{2.26}$$

Observe that $\|\xi-\nu\|_1>0$, as $\|\xi\|_1=1$ (cf. Equation (2.8)) and $\|\nu\|_1\le 1/10$ (cf. Inequality (2.25)). Define the function

$$\zeta(x)=\frac{\xi(x)-\nu(x)}{\|\xi-\nu\|_1}.$$

Since $\nu(x)=\xi(x)$ whenever $|x|>N$ (cf. Equation (2.26)), the function ζ is supported on the set $H^{N\cdot R}_{\le N}$, establishing (2.18). We establish (2.16) by computing

$$\begin{aligned}
\|\zeta-\xi\|_1 &= \sum_{x\in(\{-1,1\}^N)^R}\left|\frac{\xi(x)-\nu(x)}{\|\xi-\nu\|_1}-\xi(x)\right| \\
&\le \sum_{x\in(\{-1,1\}^N)^R}\left(\frac{1}{\|\xi-\nu\|_1}-1\right)|\xi(x)|+\frac{1}{\|\xi-\nu\|_1}|\nu(x)| \\
&\le \left(\frac{1}{\|\xi\|_1-\|\nu\|_1}-1\right)\cdot\|\xi\|_1+\frac{1}{\|\xi\|_1-\|\nu\|_1}\cdot\|\nu\|_1 \\
&\le \left(\frac{1}{1-1/10}-1\right)+\frac{1/10}{1-1/10} \\
&= \frac{2}{9}.
\end{aligned}$$

Equation (2.17) is immediate from the definition of ζ. Finally, (2.15) follows from (2.14), (2.24), and linearity. $\square$

2.7.1 Proof of Lemma 2.24

Our final task is to prove Lemma 2.24, restated here for the reader's convenience.

Lemma 2.24. *Let $\alpha\le R^2$, let $\beta\in(4\ln^2 R/(\sqrt{\alpha}R),1)$, and let R be sufficiently large. Let $N=\lceil 20\sqrt{\alpha}\rceil R$. Let $\eta_i:[M]_0\to\mathbb{R}$, for $i=1,\ldots R$, be a sequence of non-negative functions where for every i,*

$$\sum_{r=0}^{M}\eta_i(r)\le 1/2, \tag{2.22}$$

$$\eta_i(r)\le\alpha\exp(-\beta r)/r^2 \qquad \forall r=1,\ldots,M. \tag{2.23}$$

Let $P=\{\vec{t}=(t_1,\ldots,t_R)\in[M]_0^R : t_1+\cdots+t_R>N\}$. Then

$$\sum_{\vec{t}\in P}\prod_{i=1}^{R}\eta_i(t_i)\le 2^{-R}\cdot(2NR)^{-2\Delta}$$

where $\Delta\ge\beta\sqrt{\alpha}R/(4\ln^2 R)$.

Lemma 2.27. *Let $k,n\in\mathbb{N}$ with $k\le n$. Then $\binom{n}{k}\le\left(\frac{en}{k}\right)^k$.*

Lemma 2.28. *Let $m\in\mathbb{N}$ and $\eta>0$. Then*

$$\sum_{r=m}^{\infty}e^{-\eta r}=\frac{1}{1-e^{-\eta}}\cdot e^{-\eta m}.$$

Proof. We calculate

$$\sum_{r=m}^{\infty} \mathrm{e}^{-\eta r} = \mathrm{e}^{-\eta m} \cdot \sum_{r=0}^{\infty} \mathrm{e}^{-\eta r} = \mathrm{e}^{-\eta m} \cdot \frac{1}{1-\mathrm{e}^{-\eta}}. \qquad \square$$

Lemma 2.29. *Let $m \in N$. Then*

$$\sum_{r=1}^{m} \frac{1}{\sqrt{r}} \leq 2\sqrt{m}.$$

Proof. Using the fact that the function $1/\sqrt{t}$ is decreasing, we may estimate

$$\sum_{r=1}^{m} \frac{1}{\sqrt{r}} \leq \int_0^m \frac{1}{\sqrt{t}}\, dt = 2\sqrt{m}. \qquad \square$$

Proof. Let $T = \lfloor \frac{N}{2M} \rfloor$. For each $s \in \{T, \ldots, R\}$, let

$$C(s) = \max\left\{ \frac{N}{6s \ln R}, \frac{N}{6\sqrt{R \cdot s}} \right\}.$$

We begin with a simple, but important, structural observation about the set P. Let $t = (t_1, \ldots, t_R) \in [M]_0^R$ be a sequence such that $t_1 + \cdots + t_R > N$. Then we claim that there exists an $s \in \{T \ldots, R\}$ such that $t_i \geq C(s)$ for at least s indices $i \in [R]$. To see this, assume without loss of generality that the entries of $\vec{t}$ are sorted so that $t_1 \geq t_2 \geq \cdots \geq t_R$. Then there must exist an $s \geq T$ such that $t_s \geq C(s)$. Otherwise, because no t_i can exceed M, we would have:

$$\begin{aligned} t_1 + \cdots + t_R &< T \cdot M + \sum_{s=1}^{R} C(s) \\ &\leq \frac{N}{2} + \frac{N}{6 \ln R} \sum_{s=1}^{\lfloor R/\ln^2 R \rfloor} \frac{1}{s} + \frac{N}{6\sqrt{R}} \sum_{s=\lceil R/\ln^2 R \rceil}^{R} \frac{1}{\sqrt{s}} \\ &\leq \frac{N}{2} + \frac{N}{6} + \frac{N}{3} \\ &= N. \end{aligned}$$

where the last inequality follows from Lemma 2.29 and the fact that $\sum_{s=1}^{m} s^{-1} \leq \ln m + 1$ (and R is sufficiently large). Since the entries of $\vec{t}$ are sorted, the preceding values $t_1, \ldots, t_{s-1} \geq C(s)$ as well.

For each subset $S \subseteq [R]$, define

$$P_S = \{\vec{t} \in P : t_i \geq C(|S|) \text{ for all indices } i \in S\}.$$

The observations above guarantee that for every $\vec{t} = (t_1, \ldots, t_R) \in P$, there exists some set S of size at least

$s \in \{T,\dots,R\}$ such that $t_i \geq C(s)$ for all $i \in S$. Hence

$$\begin{aligned}
\sum_{\vec{t}\in P}\prod_{i=1}^{R}\eta_i(t_i) &\leq \sum_{s=T}^{R}\sum_{S\subseteq[R]:|S|=s}\sum_{\vec{t}\in P_S}\prod_{i=1}^{R}\eta_i(t_i)\\
&\leq \sum_{s=T}^{R}\binom{R}{s}\left(\max_{i\in\{1,\dots,R\}}\left(\sum_{r=\lceil C(s)\rceil}^{M}\eta_i(r)\right)\right)^{s}\left(\max_{i\in\{1,\dots,R\}}\left(\sum_{r=0}^{M}\eta_i(r)\right)\right)^{R-s}\\
&\leq 2^{-R}\sum_{s=T}^{R}\binom{R}{s}\left(\sum_{r=\lceil C(s)\rceil}^{M}2\alpha\exp(-\beta r)r^{-2}\right)^{s} && \text{by Properties (2.22) and (2.23)}\\
&\leq 2^{-R}\sum_{s=T}^{R}\left(\frac{R\mathrm{e}}{s}\right)^{s}\left(\frac{2\alpha}{(C(s))^2}\right)^{s}\cdot\sum_{r=\lceil C(s)\rceil}^{\infty}\exp\left(-\beta rs\right) && \text{by Lemma 2.27}\\
&\leq 2^{-R}\sum_{s=T}^{R}\left(\frac{R\mathrm{e}}{s}\right)^{s}\left(\frac{2\alpha}{(C(s))^2}\right)^{s}\cdot\frac{1}{1-\mathrm{e}^{-\beta s}}\cdot\exp\left(-\beta sC(s)\right) && \text{by Lemma 2.28}\\
&\leq 2^{-R}\sum_{s=T}^{R}\left(\frac{R\mathrm{e}}{s}\right)^{s}\left(\frac{72\alpha Rs}{N^2}\right)^{s}\cdot\frac{2}{\beta}\cdot\exp\left(-\frac{\beta N}{6\ln R}\right) && \text{by definition of } C(s)\\
&\qquad\text{and since } \frac{1}{1-\mathrm{e}^{-\beta s}}\leq\frac{1}{1-(1-\beta/2)}=\frac{2}{\beta}\text{ for } s\geq 1,\beta\in(0,1)\\
&\leq 2^{-R}\sum_{s=T}^{R}2^{-s}\cdot\exp\left(-\frac{3\beta\sqrt{\alpha}R}{\ln R}+\ln(2/\beta)\right) && \text{setting } N=\lceil 20\sqrt{\alpha}\rceil R\\
&\leq 2^{-R}\cdot(2NR)^{-2\Delta},
\end{aligned}$$

where

$$\Delta=\frac{1}{2\ln(2NR)}\cdot\left(\frac{3\beta\sqrt{\alpha}R}{\ln R}-\ln(2/\beta)\right)\geq\frac{\beta\sqrt{\alpha}R}{4\ln^2 R}$$

holds for sufficiently large R by the restrictions placed on α and β in the statement of the lemma. □

3 Upper bound for surjectivity

The goal of this section is to prove that the approximate degree of the surjectivity function (Definition 2.8) is $\tilde{O}(N^{3/4})$.

Theorem 3.1. *For any $R\in\mathbb{N}$, we have $\widetilde{\deg}(\mathsf{SURJ}_{N,R})=\tilde{O}(N^{3/4})$.*

For the remainder of the section, we focus on proving Theorem 3.1 in the case that $R=\Theta(N)$. This is without loss of generality by the following argument. If $R>N$, then surjectivity is identically false, and hence has (exact) degree 0. And if $R=o(N)$, then we can reduce to the case $R=\Theta(N)$ as follows. Let $N'=2N$ and $R'=R+N$; clearly $R'=\Theta(N)$. Given an input x to $\mathsf{SURJ}_{N,R}$, obtain an input x' to $\mathsf{SURJ}_{N',R'}$ by appending range elements $R+1,\dots,R+N$ to x. This construction guarantees that $\mathsf{SURJ}_{N,R}(x)=\mathsf{SURJ}_{N',R'}(x')$. It follows that an approximation of degree $\tilde{O}(N^{3/4})$ for $\mathsf{SURJ}_{N',R'}$ implies an approximation of the same degree for $\mathsf{SURJ}_{N,R}$.

Section Outline. This section is structured as follows. Section 3.1 introduces some notation that is specific to this section. Section 3.2 provides some intuition for the construction of the polynomial approximation for surjectivity using the simpler function NOR as a warmup example. Section 3.3 introduces some terminology that is useful for providing intuitive descriptions of our final polynomial construction using the language of algorithms. Finally, in Section 3.4 we formally apply our strategy to surjectivity in order to prove Theorem 3.1.

3.1 Notation

In this section, we make a few departures from the notation used in the introduction and later sections in order to more easily convey the algorithmic intuition behind our polynomial constructions. First, we will consider Boolean functions $f : \{0,1\}^n \to \{0,1\}$, where 1 corresponds to logical TRUE and 0 corresponds to logical FALSE. For such a Boolean function, we say that a polynomial $p : \{0,1\}^n \to \mathbb{R}$ is an ε-approximating polynomial for f if $p(x) \in [0,\varepsilon]$ when $f(x) = 0$ and $p(x) \in [1-\varepsilon, 1]$ when $f(x) = 1$. For such a polynomial p, it will be useful to think of $p(x)$ as representing the probability that a randomized or quantum algorithm accepts when run on input x.

By extension, throughout this section we will use $\widetilde{\deg}_\varepsilon(f)$ to denote the least degree of a real polynomial p that ε-approximates f in the sense described above, and we will write $\widetilde{\deg}(f) = \widetilde{\deg}_{1/3}(f)$. Given an input $x \in \{0,1\}^n$, we will use $|x|$ to denote its Hamming weight (i. e., $|x| = \sum_i x_i$).

3.2 Warmup: Approximating NOR

To convey the intuition behind our polynomial construction for surjectivity, we start by considering a much simpler function as an illustrative example. Consider the negation of the OR function on n bits, NOR: $\{0,1\}^n \to \{0,1\}$. We will give a novel construction of an approximating polynomial for NOR of degree $\tilde{O}(\sqrt{n})$. Of course, this is not terribly interesting since it is already known that $\widetilde{\deg}(\mathsf{NOR}) = \Theta(\sqrt{n})$ (cf. Lemma 2.6). But this construction highlights the main idea behind the more involved constructions to follow.

In many models of computation, such as deterministic or randomized query complexity, the NOR function remains just as hard if we restrict to inputs with $|x| = 0$ or $|x| = 1$ (where $|x|$ denotes the Hamming weight of $x \in \{0,1\}^n$). This is fairly intuitive, since distinguishing these two types of inputs essentially requires finding a single 1 among n possible locations. The fact that these inputs represent the "hard case" is true for approximate degree as well: any polynomial that uniformly approximates NOR to error $1/3$ on Hamming weights 0 and 1, and merely remains bounded in $[0,1]$ on the rest of the hypercube, has degree $\Omega(\sqrt{n})$ [62].

However, if we remove the boundedness constraint on inputs of Hamming weight larger than 1, then there is a polynomial of degree *one* that exactly equals the NOR function on inputs of Hamming weight 0 and 1, namely, the polynomial $1 - \sum_i x_i$. That is, if we view NOR as a promise function mapping $H^n_{\leq 1}$ to $\{-1,1\}$, then its approximate degree is $\Theta(n^{1/2})$, but its *unbounded* approximate degree is just 1.

More generally, suppose that we only want the polynomial to approximate the NOR function on all inputs of Hamming weight up to $T \leq n$, and we place no restrictions on the polynomial when evaluated at inputs of Hamming weight larger than T. This can be achieved by a polynomial of degree $O(\sqrt{T})$ (see Lemma 3.3 for a proof).

Let us refer to the set of low Hamming-weight inputs as $\mathcal{P}$, i. e.,

$$\mathcal{P} = H^n_{\leq T} = \{x \in \{0,1\}^n : 0 \leq |x| \leq T\}. \tag{3.1}$$

The above discussion shows that we can construct a polynomial p of degree $O(\sqrt{T})$ that tightly approximates NOR on inputs $x \in \mathcal{P}$, though $|p(x)|$ may be exponentially large for $x \notin \mathcal{P}$. On the other hand, distinguishing inputs with $|x| = 0$ from inputs with $x \notin \mathcal{P}$ also seems "easy". For example, a randomized algorithm that simply samples $\Theta(n/T)$ bits and declares $|x| = 0$ if and only if it does not see a 1 is correct with high probability. Analogously, we can construct a low-degree polynomial $\tilde{q}$ which distinguishes between $|x| = 0$ or $x \notin \mathcal{P}$ (and is bounded in $[0,1]$ for *all* inputs in $\{0,1\}^n$): it is not hard to show (via an explicit construction involving Chebyshev polynomials, or by appeal to a quantum algorithm called quantum counting [22]) that there exists a polynomial $\tilde{q}$ of degree $O(\sqrt{n/T})$ that accomplishes this.

To summarize the above discussion, we can construct polynomials p and $\tilde{q}$, of degree $O(\sqrt{T})$ and $O(\sqrt{n/T})$, respectively, with the following properties.

$$p(x) \in \begin{cases} [9/10, 1] & \text{if } |x| = 0 \\ [0, 1/10] & \text{if } 1 \leq |x| \leq T \\ \mathbb{R} & \text{if } x \notin \mathcal{P} \end{cases} \qquad \tilde{q}(x) \in \begin{cases} [9/10, 1] & \text{if } |x| = 0 \\ [0, 1] & \text{if } 1 \leq |x| \leq T \\ [0, 1/10] & \text{if } x \notin \mathcal{P} \end{cases} \tag{3.2}$$

Now consider the polynomial $p(x) \cdot \tilde{q}(x)$. This polynomial approximates NOR on $|x| = 0$, since its value is in $[0.81, 1]$. It also approximates NOR on $1 \leq |x| \leq T$, since its value is in $[0, 1/10]$. However, when $x \notin \mathcal{P}$, although $\tilde{q}(x)$ is small, we cannot ensure that the product $p(x) \cdot q(x)$ is small, since we have no control over $p(x)$ for such x.

To fix this, we will construct a new polynomial q that behaves like $\tilde{q}$ for $x \in \mathcal{P}$ and is *extremely* small when $x \notin \mathcal{P}$ (in particular $0 \leq q(x) \ll 1/|p(x)|$ for such x). To understand how small we need q to be, we need to determine just how large $p(x)$ can be on inputs with $x \notin \mathcal{P}$.

To understand the behavior of $p(x)$ outside of $\mathcal{P}$, we can either analyze the behavior of an explicit polynomial of our choice for the NOR function or we can appeal to a general result about the growth of polynomials that are bounded in a region (see Lemma 3.3). In either case, we get that there exists an upper bound $M = \exp(O(\sqrt{T} \log n))$ such that for all $x \notin \mathcal{P}$, $|p(x)| \leq M$.

This leads us to design a new polynomial q, which has the same behavior as $\tilde{q}$ for all $x \in \mathcal{P}$, but is at most $1/(3M)$ when $x \notin \mathcal{P}$. We can construct the polynomial q from $\tilde{q}$ by applying standard error reduction to reduce the approximation error of $\tilde{q}$ to $\varepsilon = 1/(3M)$, which increases the degree by a multiplicative factor of $O(\log(3M)) = O(\sqrt{T} \log n)$. Thus, $\deg(q) = O(\sqrt{n/T} \cdot \sqrt{T} \log n) = \tilde{O}(\sqrt{n})$.

In summary, we have now constructed polynomials p and q with the following behavior:

$$p(x) \in \begin{cases} [9/10, 1] & \text{if } |x| = 0 \\ [0, 1/10] & \text{if } 1 \leq |x| \leq T \\ [0, M] & \text{if } x \notin \mathcal{P} \end{cases} \qquad q(x) \in \begin{cases} [1 - 1/(3M), 1] & \text{if } |x| = 0 \\ [0, 1] & \text{if } 1 \leq |x| \leq T \\ [0, 1/(3M)] & \text{if } x \notin \mathcal{P} \end{cases} \tag{3.3}$$

Caricatures of these polynomials are depicted in Figure 2 and Figure 3. It is now easy to see that the product $r(x) = p(x) \cdot q(x)$ is a (1/3)-error approximation to NOR for *all* $x \in \{0,1\}^n$. The degree of the constructed polynomial is $\deg(r) \leq \deg(p) + \deg(q) = \tilde{O}(\sqrt{T} + \sqrt{n}) = \tilde{O}(\sqrt{n})$.

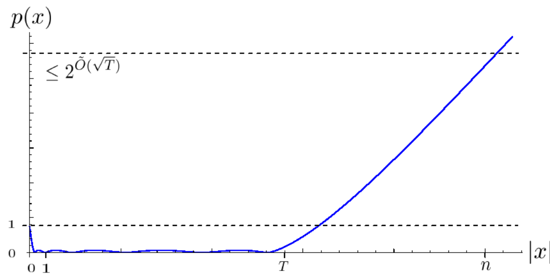

Figure 2: Caricature of the polynomial $p(x)$

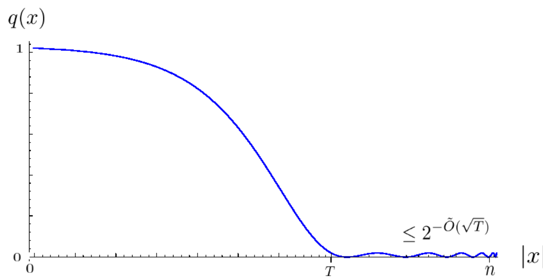

Figure 3: Caricature of the polynomial $q(x)$

Thus our constructed polynomial, r, has degree $\tilde{O}(\sqrt{n})$ and approximates NOR to error $1/3$, which is optimal by Lemma 2.6.

3.3 Informal terminology: polynomials as algorithms

Before moving on to surjectivity, we briefly introduce some terminology that will allow us convey the intuition of more involved constructions by reasoning about polynomials as if they were algorithms.

Consider three Boolean functions $p_1 : \{0,1\}^n \to \{0,1\}$, $p_2 : \{0,1\}^n \to \{0,1\}$ and $p_3 : \{0,1\}^n \to \{0,1\}$, and suppose there are deterministic algorithms A_1, A_2, and A_3 that compute these Boolean functions exactly. Now it makes sense to say "run algorithm A_1 in input x; if it accepts then output $A_2(x)$, and if it rejects, then output $A_3(x)$." The Boolean function computed by this is $A_2(x)$ if $A_1(x) = 1$ and $A_3(x)$ if $A_1(x) = 0$. Observe that the following polynomial composition of p_1, p_2, p_3 computes the same Boolean function: $p_1(x)p_2(x) + (1 - p_1(x))p_3(x)$.

We would like to use this terminology to discuss combining polynomials more generally. For polynomials p that approximate Boolean functions by outputting a value in $[0,1]$ on any input x, we can imagine $p(x)$ as representing the probability that a randomized (or quantum) algorithm accepts on input x, and the same interpretation goes through. We will also use the same terminology for polynomials that output values greater than 1, in which case we cannot interpret the output as a probability, but expressions like $p_1(x)p_2(x) + (1 - p_1(x))p_3(x)$ still make sense.

For example, in the previous section we had two polynomials p and q and we constructed the polynomial $r(x) = p(x) \cdot q(x)$ by multiplying the two polynomials together. Another way to think of this is that r is the polynomial obtained when we "run" the polynomial q and output p if it accepts and output 0 if it rejects. This yields the polynomial $q(x)p(x) + (1 - q(x)) \cdot 0 = q(x)p(x)$. We would like to informally describe polynomial constructions using this language, which will be especially useful when the constructions become more involved. So for example, we would informally describe the polynomial we constructed for the NOR function as follows (Polynomial 1):

Polynomial 1 An informal description of the polynomial approximation for NOR

1: Using q, check if $|x| > T$ (with exponentially small probability of error if indeed $|x| > T$). If so, halt and output 0.

2: Using p, compute NOR under the promise that $0 \leq |x| \leq T$ and output the result.

3.4 Approximating surjectivity

We now construct a polynomial that approximates surjectivity using the strategy described above. Recall that $\mathsf{SURJ} : [R]^N \to \{0,1\}$ is defined by $\mathsf{SURJ}(x_1,\ldots,x_N) = 1$ if and only if for all $r \in [R]$ there exists an $i \in [N]$ such that $x_i = r$.

In this section, we will use the notation

$$\#_r(x) = |\{i \in [N] : x_i = r\}| \tag{3.4}$$

to denote the number of times the range element r appears in the input x. Note that the surjectivity function evaluates to 1 if and only if $\#_r(x) \geq 1$ for all $r \in [R]$.

Finally, we will also need to consider a generalized version of surjectivity, $\mathcal{R}$-surjectivity for some set $\mathcal{R} \subseteq [R]$, which we denote $\mathsf{SURJ}_{\mathcal{R}}$. As with surjectivity,

$$\mathsf{SURJ}_{\mathcal{R}} : [R]^N \to \{0,1\}, \tag{3.5}$$

and $\mathsf{SURJ}_{\mathcal{R}}(x_1,\ldots,x_N) = 1$ if and only if for all $r \in \mathcal{R}$ there exists an $i \in [N]$ such that $x_i = r$. In other words, $\mathsf{SURJ}_{\mathcal{R}}(x) = 1$ if and only if for all $r \in \mathcal{R}$ we have $\#_r(x) \geq 1$. Note that $\mathsf{SURJ}_{[R]} = \mathsf{SURJ}$. Our construction will actually show more generally that $\widetilde{\deg}(\mathsf{SURJ}_{\mathcal{R}}) = \tilde{O}(N^{3/4})$ for all $\mathcal{R} \subseteq [R]$.

3.4.1 Approximating surjectivity on the hard inputs

To implement the strategy described in Section 3.2 in the context of surjectivity, we first choose a set $\mathcal{P}$ of inputs that we consider to be "hard". Since surjectivity can be phrased as asking whether all range items appear at least once in the input, it is natural to consider the hard inputs to be those that have few occurrences of each range item. Intuitively, this is because on such inputs, evidence that any given range item r appears at least once in the input is hard to find.

To this end, we define $\mathcal{P}$ as the set of inputs for which every range item appears at most T times, for a parameter T to be chosen later:

$$\mathcal{P} = \{x : \forall r \in [R],\ \#_r(x) \leq T\}. \tag{3.6}$$

More generally, when we consider $\mathcal{R}$-surjectivity, we denote the set of hard inputs $\mathcal{P}_{\mathcal{R}}$ and define it as

$$\mathcal{P}_{\mathcal{R}} = \{x : \forall r \in \mathcal{R},\ \#_r(x) \leq T\}. \tag{3.7}$$

In this section, we will construct a polynomial $p_{\mathcal{R}}$ that approximates $\mathsf{SURJ}_{\mathcal{R}}$ on $\mathcal{P}_{\mathcal{R}}$ to bounded error, but may be exponentially large outside of $\mathcal{P}_{\mathcal{R}}$. The value of T in the definition of $\mathcal{P}_{\mathcal{R}}$ will be chosen later; for now we only assume that our choice will satisfy $T = N^{\Theta(1)}$, which simplifies some expressions since we have $\log T = \Theta(\log N)$.

Overview of the construction of $p_{\mathcal{R}}$. To construct a polynomial that works on the hard inputs, $\mathcal{P}_{\mathcal{R}}$, we first express $\mathsf{SURJ}_{\mathcal{R}}$ as

$$\mathsf{SURJ}_{\mathcal{R}}(x) = \bigwedge_{r \in \mathcal{R}} \bigvee_{i \in [N]} \mathbb{1}[x_i = r], \tag{3.8}$$

where $\mathbb{1}[x_i = r]$ is the indicator function that takes value 1 when $x_i = r$ and 0 otherwise. Observe that for any fixed $r \in \mathcal{R}$, the function $\mathbb{1}[x_i = r]$ depends on only $\log R$ bits of x and hence can be exactly computed by a polynomial of degree at most $\log R$.

Since our goal in this section is to construct a polynomial $p_{\mathcal{R}}$ that approximates $\mathsf{SURJ}_{\mathcal{R}}$ on all inputs in $\mathcal{P}_{\mathcal{R}}$ and may be exponentially large outside of $\mathcal{P}_{\mathcal{R}}$, we can assume that each inner OR gate in Equation (3.8) is fed an input of Hamming weight at most T. Hence, our approach will be to first construct a low-degree polynomial q that approximates $\mathsf{AND}_{|\mathcal{R}|} \circ \mathsf{OR}_N$ for inputs in $\left(H^N_{\leq T}\right)^{|\mathcal{R}|}$. We then obtain the polynomial $p_{\mathcal{R}}$ that approximates $\mathsf{SURJ}_{\mathcal{R}}$ at all inputs in $\mathcal{P}_{\mathcal{R}}$ by composing q with the indicator functions $\{\mathbb{1}[x_i = r]\}_{i \in [N], r \in \mathcal{R}}$. Notice that the degree of $p_{\mathcal{R}}$ is at most $\deg(q) \cdot \log R$.

To construct q, our approach is as follows. First, we construct a polynomial V_T of degree $O(T^{1/2} \log N)$ that approximates OR_N to error $O(1/N)$ at all inputs of Hamming weight at most T. (However, $V_T(x)$ may be as large as $\exp(T^{1/2} \log N)$ for inputs x of larger Hamming weight). Invoking Lemma 2.6, we let w be a polynomial of degree $\Theta(N^{1/2})$ that approximates $\mathsf{AND}_{|\mathcal{R}|}$ to error $1/20$, and finally we define $q := w \circ V_T$. A simple and elegant analysis of Buhrman et al. [25] (cf. Lemma 3.4) allows us to argue that q indeed approximates $\mathsf{AND}_{|\mathcal{R}|} \circ \mathsf{OR}_N$ on $\left(H^N_{\leq T}\right)^{|\mathcal{R}|}$.

Preliminary lemmas. Before formally defining and analyzing $p_{\mathcal{R}}$, we record a few important facts about Chebyshev polynomials that will be useful throughout the remainder of this section.

Lemma 3.2 (Properties of Chebyshev polynomials)**.** *Let $d \in \mathbb{N}$.*

(1) There exists a polynomial $T_d : \mathbb{R} \to \mathbb{R}$ (the Chebyshev polynomial of degree d) such that $T_d(x) \in [-1,1]$ for all $x \in [-1,1]$ and $T_d(1+\mu) \geq \frac{1}{2}\exp(d\sqrt{\mu})$ for all $\mu \in (0,1)$.

(2) For any polynomial $p : \mathbb{R} \to \mathbb{R}$ of degree d with $|p(x)| \leq 1$ for all $x \in [-1,1]$, we have that for any x with $|x| > 1$,

$$|p(x)| \leq |T_d(x)| \leq (2|x|)^d, \tag{3.9}$$

where T_d is the Chebyshev polynomial of degree d.

Proof. To establish property (1), we use the fact that for $\mu > 0$, the value of the degree-d Chebyshev polynomial can be written [37, §3.2 Problem 8(f)] as

$$\begin{aligned} T_d(1+\mu) &= \cosh(d \operatorname{arcosh}(1+\mu)) \\ &\geq \cosh(d\sqrt{\mu}) && \text{for } \mu \leq 1 \\ &\geq \frac{1}{2} \cdot \exp(d\sqrt{\mu}). \end{aligned}$$

Property (2) appears as [37, 3.2 Problem 19]. □

We are now ready to establish the existence of a low-degree polynomial that approximates OR at all inputs of low Hamming weight.

Lemma 3.3 (Approximating OR on inputs of low Hamming weight)**.** *Let $\varepsilon \in (0,1)$ and $1 \le T \le N$. There is a polynomial $V_{T,\varepsilon} : \{0,1\}^N \to \mathbb{R}$ of degree $O(\sqrt{T}\log(1/\varepsilon))$ such that*

$$V_{T,\varepsilon}(x) \in \begin{cases} [0,\varepsilon] & \text{if } |x| = 0 \\ [1-\varepsilon,1] & \text{if } 1 \le |x| \le T \\ [-a,a] \text{ for some } a \in \exp\left(O\left(\sqrt{T}\cdot\log N\cdot\log(1/\varepsilon)\right)\right) & \text{if } |x| > T \end{cases} . \tag{3.10}$$

Proof. Choose $d = O(\sqrt{T}\log(1/\varepsilon))$ so that $M := T_d(1+1/T)+1 \ge 2/\varepsilon$ (as guaranteed by Property 1 of Lemma 3.2). Define $V_{T,\varepsilon}$ by the following affine transformation of T_d:

$$V_{T,\varepsilon}(x) = \left(1 - \frac{1}{M}\right) - \frac{1}{M}\cdot T_d\left(1 + \frac{1}{T} - \frac{|x|}{T}\right).$$

Then for $|x| = 0$, we have

$$V_{T,\varepsilon}(x) = \left(1 - \frac{1}{M}\right) - \frac{1}{M}\cdot T_d\left(1 + \frac{1}{T}\right) = 0.$$

If $1 \le |x| \le T$, then $1 + 1/T - |x|/T \in [-1,1]$, so $V_{T,\varepsilon(x)} \in [1-\varepsilon,1]$. Finally, if $T+1 \le |x| \le N$, then

$$|V_{T,\varepsilon}(x)| \le 1 + \frac{1}{M}\cdot T_d(N/T) \le \exp(O(\sqrt{T}\cdot\log N\cdot\log(1/\varepsilon)),$$

where the final inequality holds by Equation (3.9). □

The following lemma shows that if p and q are approximating polynomials for Boolean functions f and g, respectively, then the block composition $p \circ q$ approximates $f \circ g$, with a blowup in error that is proportional to the number of variables over which f is defined. The proof is due to Buhrman et al. [25], but our formulation is slightly different so we give the proof for completeness.

Lemma 3.4. *Let $f : \{0,1\}^n \to \{0,1\}$ and $g : X \to \{0,1\}$ be Boolean functions, where $X \subseteq \{0,1\}^m$ for some m. Let $p : \{0,1\}^n \to [0,1]$ be an ε-approximating polynomial for f, and let $q : X \to [0,1]$ be a δ-approximating polynomial for g. Then the block composition $p \circ q : X^n \to \mathbb{R}$ is an $(\varepsilon + \delta n)$-approximating polynomial for $f \circ g : X^n \to \{0,1\}$.*

Proof. Fix any input $x = (x_1,\dots,x_n) \in X^n$, and let $y = (g(x_1),\dots,g(x_n)) \in \{0,1\}^n$. Let $z \in [0,1]^n$ be defined by $z = (q(x_1),\dots,q(x_n))$. Since p is an ε-approximating polynomial for f, by the triangle inequality, it suffices to show that $|p(y) - p(z)| \le \delta n$.

Let Z be a random variable on $\{0,1\}^n$ where each $Z_i = 1$ independently with probability z_i. Then due to the multilinearity of p, we have

$$p(z) = \mathbb{E}[p(Z)] = \Pr[Z = y]\cdot p(y) + \Pr[Z \ne y]\cdot \mathbb{E}[p(Z)|Z \ne y].$$

Since q is a δ-approximating polynomial for g, we have $|y_i - z_i| \le \delta$ for every $i \in [n]$. Hence,

$$\Pr[Z = y] \ge (1-\delta)^n \ge 1 - \delta n.$$

Because p is bounded in $[0,1]$, this implies

$$p(z) \geq (1-\delta n) \cdot p(y) + 0 \geq p(y) - \delta n$$

and

$$p(z) \leq 1 \cdot p(y) + \delta n \cdot 1 = p(y) + \delta n,$$

completing the proof. □

Formal definition of $p_{\mathcal{R}}$. Let w be a $(1/20)$-approximating polynomial for $\mathsf{AND}_{|\mathcal{R}|}$ of degree $O(|\mathcal{R}|^{1/2})$ whose existence is guaranteed by Lemma 2.6. We may assume that $w(x) \in [0,1]$ for all $x \in \{0,1\}^n$ (we will exploit this assumption in the proof of Lemma 3.7 below, as it allows us to apply Lemma 3.6 below to w). Let $q := w \circ V_{T,1/(20n)}$. Finally, let us define $p_{\mathcal{R}}$ to be the composition of q with the indicator functions $\{\mathbb{1}[x_i = r]\}_{i \in [N], r \in \mathcal{R}}$. For example, if $\mathcal{R} = \{1, \ldots, |\mathcal{R}|\}$, then

$$p_{\mathcal{R}} = q\left(\mathbb{1}[x_1 = 1], \mathbb{1}[x_2 = 1], \ldots, \mathbb{1}[x_N = 1], \mathbb{1}[x_1 = 2], \ldots, \mathbb{1}[x_N = |\mathcal{R}|]\right).$$

Observe that

$$\deg(p_{\mathcal{R}}) \leq \deg(w) \cdot \deg(V_{T,1/(20n)}) \cdot \deg(\mathbb{1}[x_i = r]) \leq O\left(|\mathcal{R}|^{1/2} \cdot T^{1/2} \log n \cdot \log R\right) \leq \tilde{O}\left(\sqrt{NT}\right).$$

Showing $p_{\mathcal{R}}$ approximates $\mathsf{SURJ}_{\mathcal{R}}$ on $P_{\mathcal{R}}$. Lemma 3.4 implies that:

$$|q(x) - \mathsf{AND}_{|\mathcal{R}|} \circ \mathsf{OR}(x)| \leq 1/10 \text{ for all } x \in \left(H^N_{\leq T}\right)^{|\mathcal{R}|}. \tag{3.11}$$

An immediate consequence is the following lemma.

Lemma 3.5. $|p_{\mathcal{R}}(x) - \mathsf{SURJ}_{\mathcal{R}}(x)| \leq 1/10$ *for all* $x \in \mathcal{P}_{\mathcal{R}}$.

Bounding $p_{\mathcal{R}}$ outside of $P_{\mathcal{R}}$. For an input $x \in [R]^N$ outside of $\mathcal{P}_{\mathcal{R}}$, let $b_{\mathcal{R}}(x)$ be the number of range items that appear more than T times, i. e.,

$$b_{\mathcal{R}}(x) = |\{r \in \mathcal{R} : \#_r(x) > T\}|. \tag{3.12}$$

We claim that $|p_{\mathcal{R}}(x)| \leq \exp\left(b_{\mathcal{R}}(x) \cdot \tilde{O}(\sqrt{T})\right)$. This bound relies on the following elementary lemma.

Lemma 3.6. *Let $p : \mathbb{R}^n \to \mathbb{R}$ be a multilinear polynomial with $p(x) \in [0,1]$ for all $x \in \{0,1\}^n$. Then for $x \in \mathbb{R}^n$, we have*

$$|p(x)| \leq \prod_{i=1}^{n} (|1 - x_i| + |x_i|).$$

Proof. We prove the lemma by induction on the number of variables n. If $n = 0$, then p is a constant in the interval $[0,1]$ and the claim is true.

Now suppose the claim is true for $n-1$ variables, and let $p : \mathbb{R}^n \to \mathbb{R}$ with $p(x) \in [0,1]$ for $x \in \{0,1\}^n$. We begin by decomposing

$$p(x) = (1 - x_n) \cdot q_0(x_1, \ldots, x_{n-1}) + x_n \cdot q_1(x_1, \ldots, x_{n-1})$$

where q_0 and q_1 are themselves multilinear polynomials. Since $p(x) \in [0,1]$ for all $x \in \{0,1\}^n$, this is in particular true when $x_n = 0$. Hence $q_0(x') \in [0,1]$ for all $x' \in \{0,1\}^{n-1}$. Similarly, setting $x_n = 1$ reveals that $q_1(x') \in [0,1]$ for all $x' \in \{0,1\}^{n-1}$. Now for any $x \in \mathbb{R}^n$ with $x' = (x_1, \dots, x_{n-1})$ we have

$$\begin{aligned}|p(x)| &= \left|(1-x_n)\cdot q_0(x') + x_n \cdot q_1(x')\right| \\ &\le |1-x_n| \cdot |q_0(x')| + |x_n| \cdot |q_1(x')| \\ &\le (|1-x_n| + |x_n|) \cdot \prod_{i=1}^{n-1} (|1-x_i| + |x_i|)\end{aligned}$$

where the final inequality uses the inductive hypothesis. This proves the claim. □

Lemma 3.7. *There exists a function $a(x) = \exp\left(b_{\mathcal{R}}(x) \cdot \tilde{O}\left(\sqrt{T}\right)\right)$ such that for any $\mathcal{R} \subseteq [R]$, the polynomial $p_{\mathcal{R}} : [R]^N \to \mathbb{R}$ has degree $\tilde{O}(\sqrt{NT})$ and satisfies:*

$$p_{\mathcal{R}}(x) \in \begin{cases} [0, 1/10] & \text{if } x \in \mathcal{P}_{\mathcal{R}} \text{ and } \mathsf{SURJ}_{\mathcal{R}}(x) = 0 \\ [9/10, 1] & \text{if } x \in \mathcal{P}_{\mathcal{R}} \text{ and } \mathsf{SURJ}_{\mathcal{R}}(x) = 1 \\ [-a(x), a(x)] & \text{if } x \notin \mathcal{P}_{\mathcal{R}}. \end{cases}$$

Proof. The first two cases are an immediate consequence of Lemma 3.5.

To upper bound the value of $|p_{\mathcal{R}}(x)|$ for $x \notin \mathcal{P}_{\mathcal{R}}$, we exploit the structure of $p_{\mathcal{R}}$ as a multilinear polynomial w of degree $O(\sqrt{N})$ over the variables $z_1, \dots, z_{|\mathcal{R}|}$, where each z_r is the output of the r^{th} polynomial from Lemma 3.3. That is,

$$z_r = V_{T,1/(20n)}\left(\mathbb{1}[x_1 = r], \dots, \mathbb{1}[x_N = r]\right).$$

If $b_{\mathcal{R}}(x)$ range items appear greater than T times, this means that up to $b_{\mathcal{R}}(x)$ of the variables z_r might take values outside $[0,1]$. However, by Lemma 3.3, each of these $b_{\mathcal{R}}(x)$ variables is still at most $\exp(\tilde{O}(\sqrt{T}))$. By Lemma 3.6,

$$|w(z)| \le \prod_{r \in \mathcal{R}} (|1-z_r| + |z_r|) \le \exp\left(b_{\mathcal{R}}(x) \cdot \tilde{O}\left(\sqrt{T}\right)\right),$$

since each z_r that is in $[0,1]$ contributes a factor of exactly 1 to the product, whereas each of the remaining $b_{\mathcal{R}}(x)$ variables contributes a factor of at most $\exp(\tilde{O}(\sqrt{T}))$ to the product. □

3.4.2 Controlling the easy inputs

Intuition. Unlike the example of the NOR function, where all the inputs outside the "hard" set $\mathcal{P}$ (cf. Equation (3.1)) were in $\mathsf{NOR}^{-1}(1)$, for surjectivity there are both 0- and 1-inputs outside of the hard set $\mathcal{P}$ (cf. Equation (3.6)) . So our remaining task is not simply a matter of constructing a polynomial that detects if the input is outside of $\mathcal{P}$, as it was in the case of the NOR function.

However, we will show that, for surjectivity, the inputs outside of $\mathcal{P}$ are easy to handle in a different sense—they are easy because we can construct a reduction from inputs outside of $\mathcal{P}$ to inputs in $\mathcal{P}$. To gain some intuition, consider the task of designing a randomized algorithm for surjectivity where we want to reduce the general case to the case where there is a set $\mathcal{R} \subseteq [R]$ such that all range items

$r \in \mathcal{R}$ have $\#_r(x) \le T$. To do this, we can simply sample a large number of elements x_i and remove from consideration all range items appearing at least once in the sample, because we know these range items all appear in the input at least once. After this step, we have a new set $\mathcal{R} \subseteq [R]$ consisting of all range items that have not been seen in the sampling stage. Every $r \in \mathcal{R}$ is likely to have $\#_r(x) \le T$ because elements that appeared too frequently would have (with high probability) been observed in the sampling stage. Thus it now suffices to solve $\mathsf{SURJ}_{\mathcal{R}}$ on the input under the assumption that all range items $r \in \mathcal{R}$ have $\#_r(x) \le T$.

Informally, the above discussion states that we want to construct the polynomial described in Polynomial 2.

Polynomial 2 Informal description of the polynomial approximation for SURJ

1: Sample $S = \tilde{\Theta}(N^{3/4})$ items and remove all range items seen from $[R]$. Let the remaining set be $\mathcal{R} \subseteq [R]$.
2: Solve $\mathsf{SURJ}_{\mathcal{R}}$ under the promise that all $r \in \mathcal{R}$ have $\#_r(x) \le T$, where $T = \tilde{\Theta}(\sqrt{N})$.

We now have to construct a polynomial that represents this algorithmic idea. We have already constructed an (unbounded approximating) polynomial for $\mathsf{SURJ}_{\mathcal{R}}$ under the promise $\mathcal{P}_{\mathcal{R}}$ in the previous section, so the second step of this construction is done.

For Step 1 of Polynomial 2, we need to construct a polynomial to represent the idea of sampling input elements and evaluating a polynomial that *depends on* the sampled elements. To build up to this, consider a deterministic algorithm that queries a subset $\mathcal{S} \subseteq [N]$ of input elements, checks if the sampled string equals another fixed string y and outputs 1 if true and 0 if false. If we denote the input $x \in [R]^N$ restricted to the subset $\mathcal{S} \subseteq [N]$ as $x_{\mathcal{S}}$, then this algorithm outputs 1 if and only if $x_{\mathcal{S}} = y$. Interpreting the input as an element of $\{0,1\}^{N \log R}$ rather than $[R]^N$, it is easy to see that a deterministic query algorithm querying $|\mathcal{S}| \log R$ bits of x can solve this problem. Consequently, there is a polynomial of degree $|\mathcal{S}| \log R$ that outputs 1 if $x_{\mathcal{S}} = y$ and outputs 0 otherwise. For any fixed $\mathcal{S} \subseteq [N]$ and $y \in [R]^{|\mathcal{S}|}$, we denote this polynomial by $\mathbb{1}_y(x_{\mathcal{S}})$.

Now we can construct a polynomial which samples S elements of the input and then outputs a bit depending on the elements seen. Let $\mathcal{S} \subseteq [N]$ be a subset of indices with $|\mathcal{S}| = S$. Let $f(\mathcal{S}, x_{\mathcal{S}})$ be an arbitrary Boolean function that tells us whether to output 0 or 1 on seeing the sample $(\mathcal{S}, x_{\mathcal{S}})$. Then the following polynomial samples a random set $\mathcal{S}$ of size S, reads the input x restricted to the set $\mathcal{S}$, and outputs the bit $f(\mathcal{S}, x_{\mathcal{S}})$:

$$\frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \sum_{y \in [R]^S} \mathbb{1}_y(x_{\mathcal{S}}) f(\mathcal{S}, y) \tag{3.13}$$

Note that for any function f, this is a polynomial of degree $S \log R$ in the variables $x_1, \ldots, x_N$, because $\mathbb{1}_y(x_{\mathcal{S}})$ is a multilinear polynomial over those variables, and $f(\mathcal{S}, y)$ is simply a hard-coded bit that does not depend on x. The value of this polynomial on an input x equals

$$\frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \sum_{y \in [R]^S} \mathbb{1}_y(x_{\mathcal{S}}) f(\mathcal{S}, y) = \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} f(\mathcal{S}, x_{\mathcal{S}}). \tag{3.14}$$

We can now generalize this construction to allow for the possibility that the function f is itself a polynomial. This is what we need to implement Polynomial 2, in which we sample a random set $\mathcal{S} \subseteq [N]$ of size S, query $x_{\mathcal{S}}$ (using $S \log R$ queries), and then run a polynomial that depends on the results. Specifically, if we sample the set $\mathcal{S} \subseteq [N]$, and learn that $x_{\mathcal{S}}$ equals the string $y \in [R]^{|\mathcal{S}|}$, then we want to run the polynomial $p_{\mathcal{R}}(y)$ from Lemma 3.7 for the set $\mathcal{R}(y)$ of all elements in $[R]$ that do not appear in y, i. e.,

$$\mathcal{R}(y) = [R] \setminus \{r : \exists i \; y_i = r\}. \tag{3.15}$$

Formal description of the approximation to surjectivity. Using the tools from Section 3.4.1 and the above discussion, we can now construct a polynomial that corresponds to the informal description in Polynomial 2:

$$r(x) = \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \sum_{y \in [R]^S} \mathbb{1}_y(x_{\mathcal{S}}) p_{\mathcal{R}(y)}(x). \tag{3.16}$$

This is a polynomial of degree $S \log R + \max_y \{\deg(p_{\mathcal{R}(y)})\} = \tilde{O}(S + \sqrt{NT}) = \tilde{O}(N^{3/4})$, using $S = \tilde{\Theta}(N^{3/4})$ and $T = \tilde{\Theta}(\sqrt{N})$ (where the factors hidden by the $\tilde{\Theta}$ notation will be chosen later). The value of the polynomial on input x is

$$r(x) = \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} p_{\mathcal{R}(x_{\mathcal{S}})}(x). \tag{3.17}$$

where recall that $\mathcal{R}(x_{\mathcal{S}})$ is as defined in Equation (3.15). The right hand side of Equation (3.17) is precisely the expected value of the polynomial $p_{\mathcal{R}(x_{\mathcal{S}})}(x)$ when $\mathcal{S}$ is a uniformly random set of size S. We will now show that $r(x)$ is an approximating polynomial for surjectivity, i. e., that for all $x \in [R]^N$, $|\mathsf{SURJ}(x) - r(x)| \leq 1/3$.

Recalling that $b_{\mathcal{R}(x_{\mathcal{S}})}$ is the number of range items $r \in \mathcal{R}$ that appear in x greater than T times (cf. Equation (3.12)), we compute the value of the polynomial $r(x)$ on an input x:

$$\begin{aligned} r(x) &= \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} p_{\mathcal{R}(x_{\mathcal{S}})}(x) && (3.18)\\ &= \frac{1}{\binom{N}{S}} \left(\sum_{\mathcal{S} \in \binom{[N]}{S}} \sum_{b=0}^{N/T} \mathbb{1}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = b] \cdot p_{\mathcal{R}(x_{\mathcal{S}})}(x) \right) && (3.19)\\ &= \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \mathbb{1}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = 0] \cdot p_{\mathcal{R}(x_{\mathcal{S}})}(x) && (3.20)\\ &\qquad + \sum_{b=1}^{N/T} \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \mathbb{1}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = b] \cdot p_{\mathcal{R}(x_{\mathcal{S}})}(x), && (3.21) \end{aligned}$$

where we split up the sum into the $b = 0$ (3.20) and $b \geq 1$ (3.21) cases. We first show that the $b \geq 1$ term (3.21) has essentially no contribution to the final value.

When $b_{\mathcal{R}(x_{\mathcal{S}})}(x) \geq 1$, we know by Lemma 3.7 that the magnitude of the polynomial $|p_{\mathcal{R}}(x)|$ is at most $\exp\left(b_{\mathcal{R}(x_{\mathcal{S}})}(x) \cdot \tilde{O}\left(\sqrt{T}\right)\right)$. Thus the magnitude of the term in Expression (3.21) is at most

$$\sum_{b=1}^{N/T} \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \mathbb{1}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = b] \cdot \exp(b \cdot \tilde{O}(\sqrt{T})) \tag{3.22}$$

$$= \sum_{b=1}^{N/T} \Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = b] \cdot \exp\left(b \cdot \tilde{O}\left(\sqrt{T}\right)\right). \tag{3.23}$$

We now need to compute the value of $\Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = b]$ for each b. Intuitively, this roughly corresponds to the probability that we sample $S = \tilde{\Theta}(N^{3/4})$ elements from the input and miss all bT elements that correspond to the $T = \tilde{\Theta}(\sqrt{N})$ copies of the b range items that appear at least T times. If we were to sample $N/(bT)$ elements, the probability of seeing none of the bT elements would be $\Theta(1)$. Since we are sampling $S = \tilde{\Theta}(N^{3/4}) \gg N/(bT)$ elements, the probability of not seeing one of the bT elements is $\exp(-bST/N) = \exp(-b \cdot \tilde{\Omega}(N^{1/4}))$. We make this heuristic calculation formal in the following lemma.

Lemma 3.8. *Let $S, T \leq N$. Let $\mathcal{S} \subset [N]$ be a random subset of size S. Then for every $x \in [R]^N$ and every $b \geq 1$,*

$$\Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) \geq b] \leq \exp(-b \cdot (ST/N - \log N)),$$

recalling that $b_{\mathcal{R}(x_{\mathcal{S}})}$ is the number of range items $r \in \mathcal{R}$ that appear in x more than T times, but do not appear in $x_{\mathcal{S}}$.

Proof. To calculate the probability of interest, we begin by analyzing a simplified experiment. Suppose there are b range items $r_1, \ldots, r_b$ each appearing greater than T times in x. We compute the probability that *none* of the items $r_1, \ldots, r_b$ appear in the sample $x_{\mathcal{S}}$. If $T \geq N - S$, then this probability is zero. Otherwise, by direct calculation, we have for each $i = 1, \ldots, b$,

$$\Pr_{\mathcal{S}}[r_i \notin x_{\mathcal{S}} | r_1 \notin x_{\mathcal{S}}, \ldots, r_{i-1} \notin x_{\mathcal{S}}] \leq \prod_{j=0}^{S-1} \left(1 - \frac{T}{N-j}\right) \leq \left(1 - \frac{T}{N}\right)^S \leq \exp\left(-\frac{ST}{N}\right).$$

Hence the probability that none of $r_1, \ldots, r_b$ appear is

$$\Pr_{\mathcal{S}}[r_1 \notin x_{\mathcal{S}} \wedge \cdots \wedge r_b \notin x_{\mathcal{S}}] = \prod_{i=1}^{b} \Pr_{\mathcal{S}}[r_i \notin x_{\mathcal{S}} | r_1 \notin x_{\mathcal{S}}, \ldots, r_{i-1} \notin x_{\mathcal{S}}] \leq \exp\left(-\frac{bST}{N}\right).$$

Now fix an arbitrary x, and let $r_1, \ldots, r_k$ be the range items appearing greater than T times in x. Observe that $k \leq N/T$. We estimate the probability of interest by taking a union bound over all subsets of $r_1, \ldots, r_k$ of size b:

$$\begin{aligned}
\Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) \geq b] &\leq \sum_{Y \subseteq [k] : |Y| = b} \Pr_{\mathcal{S}}[r_{i_j} \notin x_{\mathcal{S}} \quad \forall j \in Y] \\
&\leq \binom{k}{b} \exp\left(-\frac{bST}{N}\right) \\
&\leq \exp\left(-\frac{bST}{N} + b \log N\right).
\end{aligned}$$

□

From Lemma 3.8 we know that $\Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = b] \le \exp(-b \cdot \tilde{\Omega}(ST/N))$. Hence the $b \ge 1$ term (3.21) is at most

$$\sum_{b=1}^{N/T} \exp(-b \cdot \tilde{\Omega}(ST/N)) \cdot \exp(b \cdot \tilde{O}(\sqrt{T})) = o(1), \tag{3.24}$$

by choosing $T = \tilde{\Theta}(\sqrt{N})$ and $S = \tilde{\Theta}(N^{3/4})$ appropriately. This shows that the term in (3.21) does not significantly influence the value of the polynomial for any input x.

Thus we have for any input x,

$$r(x) = \frac{1}{\binom{N}{S}} \sum_{\mathcal{S} \in \binom{[N]}{S}} \mathbb{1}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = 0] \cdot p_{\mathcal{R}(x_{\mathcal{S}})}(x) + o(1). \tag{3.25}$$

By Lemma 3.7, we know that if $b_{\mathcal{R}(x_{\mathcal{S}})}(x) = 0$ (and hence $x \in \mathcal{P}_{\mathcal{R}(x_{\mathcal{S}})}$), then $p_{\mathcal{R}(x_{\mathcal{S}})}(x)$ is a $(1/10)$-approximation to $\mathsf{SURJ}(x)$. Applying Lemma 3.8 once again shows that

$$\Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) \ge 1] \le \exp(-\tilde{\Omega}(N^{1/4})),$$

so $\Pr_{\mathcal{S}}[b_{\mathcal{R}(x_{\mathcal{S}})}(x) = 0] = 1 - o(1)$. Hence, for all $x \in [R]^N$, $|r(x) - \mathsf{SURJ}(x)| \le 1/10 + o(1) \le 1/3$.

This completes the proof of Theorem 3.1.

4 Lower bound for surjectivity

The goal of this section is to show the following improved lower bound on the approximate degree of the surjectivity function.

Theorem 4.1. *For some $N = O(R)$, the $(1/3)$-approximate degree of $\mathsf{SURJ}_{N,R}$ is $\tilde{\Omega}(R^{3/4})$.*

To prove Theorem 4.1, we combine the following theorem with the reductions of Proposition 2.10 and Corollary 2.16.

Theorem 4.2. *Let $N = c \cdot R$ for a sufficiently large constant $c > 0$. Let $F^{\le N} \colon H_{\le N}^{N \cdot R} \to \{-1,1\}$ equal $\mathsf{AND}_R \circ \mathsf{OR}_N$ restricted to inputs in $H_{\le N}^{N \cdot R} = \{x \in \{-1,1\}^{N \cdot R} : |x| \le N\}$. Then $\widetilde{\mathrm{ubdeg}}(F^{\le N}) \ge \tilde{\Omega}(R^{3/4})$.*

The proof of Theorem 4.2 entails using dual witnesses for the high approximate degree of AND_R and OR_N to construct a dual witness for the higher approximate degree of $F^{\le N}$. As indicated in Section 1.3, the construction is essentially the same as in [33], except that we observe that a dual witness for OR constructed and used in prior work satisfies an exponentially stronger decay condition than has been previously realized.

The construction can be thought of as consisting of three steps:

Step 1. We begin by constructing a dual witness ψ for the fact that the unbounded approximate degree of the OR_N function is $\Omega\left(\sqrt{T}\right)$ even when promised that the input has Hamming weight at most $T = \Theta(\sqrt{R})$. The dual witness ψ is a small variant of the one in [33], but we give a more careful analysis of its tail decay. In particular, we make use of the fact that for all $t \geq 1$, the ℓ_1 weight that ψ places on the t-th layer of the Hamming cube is upper bounded by $\exp(-\Omega(t/\sqrt{T}))/t^2$.

Step 2. We combine ψ with a dual witness Φ for AND_R to obtain a preliminary dual witness $\Phi \star \psi$ for $F = \mathsf{AND}_R \circ \mathsf{OR}_N$. The dual witness $\Phi \star \psi$ shows that F has approximate degree $\Omega(\sqrt{R} \cdot \sqrt{T}) = \Omega(R^{3/4})$. However, $\Phi \star \psi$ places weight on inputs of Hamming weight larger than N, and hence does not give an *unbounded* approximate degree lower bound for the promise variant $F^{\leq N}$.

Step 3. Using Proposition 2.22 we zero out the mass that $\Phi \star \psi$ places on inputs of Hamming weight larger than N, while maintaining its pure high degree and correlation with $F^{\leq N}$. This yields the final desired *unbounded* approximate degree dual witness ζ for $F^{\leq N}$, as per Proposition 2.5.

4.1 Step 1: A dual witness for OR_N

We begin by constructing a univariate function which captures the properties we need of our inner dual witness for OR_N. The construction slightly modifies the dual polynomial for OR_N given by Špalek [82]. We provide a careful analysis of its decay as a function of the input's Hamming weight.

Proposition 4.3. *Let $T \in \mathbb{N}$ and $1/T \leq \delta \leq 1/2$. There exist constants $c_1, c_2 \in (0,1)$ and a function $\omega : [T]_0 \to \mathbb{R}$ such that*

$$\omega(0) - \sum_{t=1}^{T} \omega(t) \geq 1 - \delta, \tag{4.1}$$

$$\sum_{t=0}^{T} |\omega(t)| = 1, \tag{4.2}$$

$$\textit{For all univariate polynomials } q\colon \mathbb{R} \to \mathbb{R},\ \deg q < c_1\sqrt{\delta T} \implies \sum_{t=0}^{T} \omega(t) \cdot q(t) = 0, \tag{4.3}$$

$$|\omega(t)| \leq \frac{170 \exp(-c_2 t\sqrt{\delta}/\sqrt{T})}{\delta \cdot t^2} \qquad \forall t = 1, \dots, T. \tag{4.4}$$

Proof of Proposition 4.3. By renormalizing, it suffices to construct a function $\omega : [T]_0 \to \mathbb{R}$ such that

$$\omega(0) - \sum_{t=1}^{T} \omega(t) \geq (1 - \delta)\|\omega\|_1, \tag{4.5}$$

$$\text{For all univariate polynomials } q\colon \mathbb{R} \to \mathbb{R},\ \deg q < c_1\sqrt{\delta T} \implies \sum_{t=0}^{T} \omega(t) \cdot q(t) = 0, \tag{4.6}$$

$$|\omega(t)| \leq \frac{170 \exp(-c_2 t\sqrt{\delta}/\sqrt{T}) \|\omega\|_1}{\delta \cdot t^2} \qquad \forall t = 1, \dots, T. \tag{4.7}$$

Let $c = \lceil 8/\delta \rceil$ below. We will freely use the fact that since $\delta \le 1/2$, we have $c \le c^2/(c-1) \le 10/\delta$. Let $m = \lfloor \sqrt{T/2c} \rfloor$ and define the set

$$S = \{1, c\} \cup \{2ci^2 : 0 \le i \le m\}.$$

Note that $|S| \ge c_1\sqrt{\delta T}$ for some absolute constant $c_1 > 0$. Define the function

$$\omega(t) = \frac{(-1)^{t+(T-m)}}{T!}\binom{T}{t}\prod_{r\in[T]_0\setminus S}(t-r).$$

Property (4.6) follows from the following well-known combinatorial identity.

Fact 4.4 (e. g., [43] Equation (5.23) or [63]). *Let $T \in \mathbb{N}$, and let p be a polynomial of degree less than T. Then*

$$\sum_{t=0}^{T}(-1)^t\binom{T}{t}p(t) = 0.$$

Expanding the binomial coefficient in the definition of ω reveals that

$$|\omega(t)| = \begin{cases} \prod_{r\in S\setminus\{t\}} \frac{1}{|t-r|} & \text{for } t \in S, \\ 0 & \text{otherwise.} \end{cases}$$

We now use this characterization to establish the improved tail decay property (4.7). This clearly holds for $t = 1$ with $c_2 = 1/10$, since $|\omega(1)| \le \|\omega\|_1$ and

$$\frac{170}{\delta}\exp(-c_2\sqrt{\delta}/\sqrt{T}) \ge 340 \cdot \mathrm{e}^{-\sqrt{2}/10} > 1.$$

For $t = c$, we have

$$\begin{aligned} \frac{|\omega(c)|}{\omega(0)} &= \frac{c\prod_{i=1}^{m}(2ci^2)}{c(c-1)\prod_{i=1}^{m}(2ci^2-c)} \\ &= \frac{1}{c-1}\left(\prod_{i=1}^{m}\frac{i^2-1/2}{i^2}\right)^{-1} \\ &\le \frac{1}{c-1}\left(1-\sum_{i=1}^{m}\frac{1}{2i^2}\right)^{-1} \\ &\le \frac{1}{c-1}\left(1-\frac{\pi^2}{12}\right)^{-1} \le \frac{6}{c-1} \end{aligned} \tag{4.8}$$

where the first inequality follows from the fact that $\prod_{i=1}^{m}(1-a_i) \ge 1-\sum_{i=1}^{m}a_i$ for $a_i \in (0,1)$. Now note that

$$\begin{aligned}
\frac{|\omega(c)|}{\|\omega\|_1} &\le \frac{|\omega(c)|}{\omega(0)} \\
&\le \frac{6}{c-1} \\
&\le \frac{60}{\delta \cdot c^2} && \text{since } 10/\delta \ge c^2/(c-1) \\
&\le \frac{170}{\delta \cdot c^2} \cdot \mathrm{e}^{-c\sqrt{\delta}/(10\sqrt{T})},
\end{aligned}$$

since $\delta \ge 1/T$ and hence $\mathrm{e}^{-c\sqrt{\delta}/(10\sqrt{T})} \ge \mathrm{e}^{-1}$. Thus (4.7) holds for $t = c$, recalling that $c_2 = 1/10$.

For $t = 2cj^2$ with $j \ge 1$, we get

$$\begin{aligned}
\frac{|\omega(t)|}{\omega(0)} &= \frac{c\prod_{i=1}^m (2ci^2)}{(2cj^2-1)(2cj^2-c)\prod_{i\in[m]_0\setminus\{j\}} |2ci^2-2cj^2|} \\
&= \frac{c(m!)^2}{(4c^2j^4-(2c^2+2c)j^2+c)\prod_{i\in[m]_0\setminus\{j\}}(i+j)|i-j|} \\
&= \frac{c}{4c^2j^4-(2c^2+2c)j^2+c} \cdot \frac{(m!)^2}{(m+j)!(m-j)!}.
\end{aligned}$$

For $j \ge 1$, the first factor is bounded by

$$\frac{c}{4c^2j^4-(2c^2+2c)j^2+c} \le \frac{3c}{(2cj^2)^2},$$

using the fact that $c \ge 2$. We control the second factor by

$$\begin{aligned}
\frac{(m!)^2}{(m+j)!(m-j)!} &= \frac{m}{m+j} \cdot \frac{m-1}{m+j-1} \cdot \ldots \cdot \frac{m-j+1}{m+1} \\
&\le \left(\frac{m}{m+j}\right)^j \\
&\le \left(1 - \frac{j}{2m}\right)^j \\
&\le \mathrm{e}^{-j^2/(2m)},
\end{aligned}$$

where the last inequality uses the fact that $1-x \le \mathrm{e}^{-x}$ for all x. Since

$$\frac{|\omega(t)|}{\|\omega\|_1} \le \frac{|\omega(2cj^2)|}{\omega(0)} \le \frac{3c}{(2cj^2)^2} \cdot \mathrm{e}^{-2cj^2/(4cm)} \le \frac{170}{\delta \cdot t^2} \cdot \mathrm{e}^{-t\sqrt{\delta}/(10\sqrt{T})},$$

this establishes (4.7).

What remains is to perform the correlation calculation to establish (4.5). For $t = 1$, we observe

$$\frac{|\omega(1)|}{\omega(0)} = \frac{c\prod_{i=1}^m (2ci^2)}{(c-1)\prod_{i=1}^m (2ci^2-1)} \ge \prod_{i=1}^m \frac{i^2}{i^2 - 1/(2c)} \ge 1.$$

Next, we observe that the total contribution of $t > c$ to $\|\omega\|_1/\omega(0)$ is at most

$$\sum_{t>c} \frac{|\omega(t)|}{\omega(0)} \leq \sum_{j=1}^{m} \frac{3c}{(2cj^2)^2} < \sum_{j=1}^{\infty} \frac{3}{4cj^4} = \frac{\pi^4}{120c}. \tag{4.9}$$

Next, we calculate

$$\begin{aligned}
\omega(0) - \sum_{t=1}^{T} \omega(t) &\geq \omega(0) - \omega(1) - \left(\sum_{t=c}^{T} |\omega(t)| \right) \\
&\geq \omega(0) - \omega(1) - \left(\omega(c) + \omega(0) \cdot \frac{\pi^4}{120c} \right) && \text{by (4.9)} \\
&\geq -\omega(1) + \omega(0) \left(1 - \frac{6}{c-1} - \frac{\pi^4}{120c} \right) && \text{by (4.8)} \\
&\geq -\omega(1) + (1-\delta)\omega(0) && \text{by our choice of } c \geq 8/\delta. \qquad (4.10)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\|\omega\|_1 &\leq \omega(0) - \omega(1) + \omega(c) + \omega(0) \cdot \frac{\pi^4}{120c} && \text{by (4.9)} \\
&\leq -\omega(1) + \omega(0) \left(1 + \frac{6}{c-1} + \frac{\pi^4}{120c} \right) && \text{by (4.8)} \\
&\leq -\omega(1) + (1+\delta)\omega(0) && \text{since } c \geq 8/\delta. \qquad (4.11)
\end{aligned}$$

Combining (4.10) and (4.11), and using the fact that $-\omega(1) \geq \omega(0)$ shows that

$$\frac{\omega(0) - \sum_{t=1}^{k} \omega(t)}{\|\omega\|_1} \geq \frac{-\omega(1) + (1-\delta)\omega(0)}{-\omega(1) + (1+\delta)\omega(0)} \geq \frac{2-\delta}{2+\delta} \geq 1 - \delta.$$

This establishes (4.5), completing the proof. □

The following construction of a dual polynomial for OR_N, with $N \geq T$, is an immediate consequence of Minsky-Papert symmetrization (Lemma 2.7), combined with Proposition 4.3.

Proposition 4.5. *Let $T, N \in \mathbb{N}$ with $T \leq N$, and let $\delta > 1/T$. Define ω as in Proposition 4.3. Define the function $\psi : \{-1,1\}^N \to [-1,1]$ by $\psi(x) = \omega(|x|)/\binom{N}{|x|}$ for $x \in H^N_{\leq T}$ and $\psi(x) = 0$ otherwise. Then*

$$\langle \psi, \mathsf{OR}_N \rangle \geq 1 - \delta, \tag{4.12}$$

$$\|\psi\|_1 = 1, \tag{4.13}$$

$$\textit{For any polynomial } p \colon \{-1,1\}^N \to \mathbb{R},\ \deg p < c_1\sqrt{\delta T} \implies \langle \psi, p \rangle = 0. \tag{4.14}$$

4.2 Step 2: Constructing a preliminary dual witness for $\mathsf{AND}_R \circ \mathsf{OR}_N$

The following proposition, when combined with Proposition 2.20, shows that there is a function $\Phi : \{-1,1\}^R \to [-1,1]$ such that the dual block composition $\Phi \star \psi$ is a good dual polynomial for $\mathsf{AND}_R \circ \mathsf{OR}_N$. In the next section, we will modify $\Phi \star \psi$ to zero out the weight it places outside $H^{N\cdot R}_{\leq N}$.

Proposition 4.6. *Let* $\mathsf{OR}_N : \{-1,1\}^N \to \{-1,1\}$ *and* $\mathsf{AND}_R \colon \{-1,1\}^R \to \{-1,1\}$. *Let* $\psi \colon \{-1,1\}^N \to [-1,1]$ *be a function such that*

$$\|\psi\|_1 = 1 \qquad \text{and} \qquad \langle \psi, \mathsf{OR}_N \rangle \geq 19/20.$$

Then there exists a function $\Phi \colon \{-1,1\}^R \to [-1,1]$ *with pure high degree* $\Omega(\sqrt{R})$ *and* $\|\Phi\|_1 = 1$ *such that*

$$\langle \Phi \star \psi, \mathsf{AND}_R \circ \mathsf{OR}_N \rangle > 2/3.$$

The proof of Proposition 4.6 is implicit in the results of [29, 73].

4.3 Step 3: Constructing the final dual witness

Proposition 4.7. *Let R be sufficiently large. There exist* $N = O(R)$, $D = \tilde{\Omega}(N^{3/4})$, *and* $\zeta : (\{-1,1\}^N)^R \to \mathbb{R}$ *such that*

$$\zeta(x) = 0 \text{ for all } x \notin H^{N\cdot R}_{\leq N}, \tag{4.15}$$

$$\sum_{x \in H^{N\cdot R}_{\leq N}} \zeta(x) \cdot (\mathsf{AND}_R \circ \mathsf{OR}_N)(x) > 1/3, \tag{4.16}$$

$$\|\zeta\|_1 = 1, \text{ and} \tag{4.17}$$

$$\text{For every polynomial } p \colon (\{-1,1\}^N)^R \to \mathbb{R} \text{ of degree less than } D, \text{ we have } \langle p, \zeta \rangle = 0. \tag{4.18}$$

Proof. We start by fixing choices of several key parameters:

- $d = \Theta(\sqrt{R})$ is the pure high degree of the dual witness Φ for AND_R in Proposition 4.6,
- $T = \lfloor (R/d)^{1/2} \rfloor^2 = \Theta(\sqrt{R})$,
- $\hat{D} = c_1 \sqrt{T} \cdot d = \Theta(R^{3/4})$, where c_1 is the constant from Proposition 4.3,
- $\delta = 1/20$,
- $\alpha = 170/\delta = 3400$,
- $\beta = c_2 \cdot \sqrt{\delta}/\sqrt{T} = \Theta(1/R^{1/4})$, where c_2 is the constant from Proposition 4.3,
- $N = \lceil 20\sqrt{\alpha} \rceil R = 693R$.

Let $\psi : \{-1,1\}^N \to [-1,1]$ be the function constructed in Proposition 4.5 with $\delta := 1/20$. Let $\Phi : \{-1,1\}^R \to [-1,1]$ be the function constructed in Proposition 4.6, and define $\xi = \Phi \star \psi$. Then by Proposition 2.20, ξ satisfies the following conditions:

$$\langle \xi, \mathsf{AND}_R \circ \mathsf{OR}_N \rangle > 2/3, \tag{4.19}$$
$$\|\xi\|_1 = 1, \tag{4.20}$$
$$\text{For every polynomial } p \text{ of degree less than } D, \text{ we have } \langle \xi, p \rangle = 0. \tag{4.21}$$

Recall that ψ was obtained by symmetrizing the function ω constructed in Proposition 4.3. Proposition 2.22 guarantees that for some $\Delta \geq \beta\sqrt{\alpha}R/(4\ln^2 R) = \tilde{\Omega}(R^{3/4})$, the function ξ can be modified to produce a function $\zeta : (\{-1,1\}^N)^R \to \mathbb{R}$ such that

$$\begin{aligned}
&\zeta(x) = 0 \text{ for all } x \notin H^{N\cdot R}_{\leq N},\\
&\langle \zeta, \mathsf{AND}_R \circ \mathsf{OR}_N \rangle \geq \langle \xi, \mathsf{AND}_R \circ \mathsf{OR}_N \rangle - \|\zeta - \xi\|_1 \geq 2/3 - 2/9 > 1/3,\\
&\|\zeta\|_1 = 1,\\
&\text{For every polynomial } p \text{ of degree less than } \min\{\hat{D}, \Delta\}, \text{ we have } \langle \zeta, p \rangle = 0.
\end{aligned}$$

Observing that

$$D = \min\{\hat{D}, \Delta\} = \tilde{\Omega}(R^{3/4})$$

shows that the function ζ satisfies the conditions necessary to prove Proposition 4.7. □

Theorem 4.2 follows by combining Proposition 4.7 with the dual characterization of unbounded approximate degree given in Proposition 2.5. By Corollary 2.16, we conclude that $\widetilde{\deg}(\mathsf{dSURJ}_{N,R}) = \tilde{\Omega}(R^{3/4})$. Theorem 4.1 follows by Proposition 2.10.

5 Lower bound for k-distinctness

Our goal is to prove the following lower bound on the approximate degree of the k-distinctness function.

Theorem 5.1. *For $k \geq 2$ and some $N = O_k(R)$, the $(1/3)$-approximate degree of $\mathsf{DIST}^k_{N,R}$ is*

$$\tilde{\Omega}_k(R^{3/4-1/(2k)}).$$

The same lower bound holds for the quantum query complexity of $\mathsf{DIST}^k_{N,R}$.

Here, the notation O_k hides factors depending only on k, and $\tilde{\Omega}_k$ hides factors logarithmic in R and factors depending only on k.

Theorem 5.1 is a consequence of applying the reductions of Proposition 2.13 and Corollary 2.18 to the following, which is the ultimate goal of this section.

Theorem 5.2. *Let $G^{\leq N} : H^{N\cdot R}_{\leq N} \to \{-1,1\}$ equal $\mathsf{OR}_R \circ \mathsf{THR}^k_N$ restricted to inputs in $H^{N\cdot R}_{\leq N}$. Then for some $N = O_k(R)$, we have $\widetilde{\mathrm{ubdeg}}(G^{\leq N}) \geq \tilde{\Omega}_k(R^{3/4-1/(2k)})$.*

The proof of Theorem 5.2 will follow the same basic outline as the proof of Theorem 4.2. We will construct a dual polynomial for $G^{\leq N}$ via the following three steps:

Step 1. We first construct a dual witness ψ showing that the unbounded-approximate degree of the THR_N^k function is $\Omega_k\left(\sqrt{TN^{-1/k}}\right)$, even when promised that the input has Hamming weight at most $T = \Theta_k(\sqrt{R})$. Moreover, this dual witness satisfies additional conditions that are exploited in Step 2 below.

Step 2. We combine ψ with a dual witness Φ for OR_R to obtain a preliminary dual witness $\Phi \star \psi$ for $\mathsf{OR}_R \circ \mathsf{THR}_N^k$. The dual witness $\Phi \star \psi$ shows that $\mathsf{OR}_R \circ \mathsf{THR}_N^k$ has approximate degree $\Omega(\sqrt{R} \cdot \sqrt{T}) = \Omega_k(R^{3/4-1/(2k)})$. However, $\Phi \star \psi$ places weight on inputs of Hamming weight larger than N, and hence does not give an *unbounded* approximate degree lower bound for the promise variant $G^{\leq N}$.

Step 3. Using Proposition 2.22 we zero out the mass that $\Phi \star \psi$ places on inputs of Hamming weight larger than N, while maintaining its pure high degree and correlation with $G^{\leq N}$. This yields the final desired dual witness ζ for $G^{\leq N}$.

Additional Notation. For functions $f : \mathcal{X} \to \{-1,1\}$ and $\psi : \mathcal{X} \to \mathbb{R}$, define the error regions

$$E_+(\psi, f) = \{x \in \mathcal{X} : \psi(x) > 0, f(x) = -1\},$$
$$E_-(\psi, f) = \{x \in \mathcal{X} : \psi(x) < 0, f(x) = +1\}.$$

These are the regions where ψ disagrees in sign with f. We refer to E_+ as the set of "false positive" errors made by ψ, and E_- as the set of "false negative" errors.

5.1 Step 1: A dual witness for THR_N^k

We begin by constructing a univariate version of our dual witness for THR_N^k. Properties (5.1) and (5.2) below amount to more refined conditions on the correlation between ω and the (symmetrized) THR_N^k function. These properties will be needed in order to execute Step 2 of the construction in Section 5.2.

Proposition 5.3. *Let $k, T, N \in \mathbb{N}$ with $k \leq T$. There exist constants $c_1, c_2 \in (0,1]$ and a function ω : $\{0,1,\ldots,T\} \to \mathbb{R}$ such that*

$$\sum_{\omega(t)>0, t\geq k} |\omega(t)| \leq \frac{1}{48N}, \tag{5.1}$$

$$\sum_{\omega(t)<0, t<k} |\omega(t)| \leq \left(\frac{1}{2} - \frac{2}{4^k}\right), \tag{5.2}$$

$$\|\omega\|_1 := \sum_{t=0}^{T} |\omega(t)| = 1, \tag{5.3}$$

For all univariate polynomials $q \colon \mathbb{R} \to \mathbb{R}$,

$$\deg p < c_1\sqrt{k^{-1} \cdot T \cdot N^{-1/k}} \implies \sum_{t=0}^{T} \omega(t) \cdot q(t) = 0, \tag{5.4}$$

$$|\omega(t)| \leq \frac{(2k)^k \exp(-c_2 t/\sqrt{k \cdot T \cdot N^{1/k}})}{t^2} \qquad \forall t = 1,2,\ldots,T. \tag{5.5}$$

Proof. If $k = 1$, then the function defined by $\omega(0) = \frac{1}{2}$ and $\omega(1) = -\frac{1}{2}$ satisfies the conditions of the proposition for $c_1 = c_2 = 1$. In what follows, we treat the complementary case where $k \geq 2$.

Let $E_+ := \{t\colon \omega(t) > 0, t \geq k\}$, and $E_- := \{t\colon \omega(t) < 0, t < k\}$. By normalizing, it suffices to construct a function $\omega : [T]_0 \to \mathbb{R}$ such that

$$\sum_{t \in E_+} |\omega(t)| \leq \frac{1}{48N} \cdot \|\omega\|_1, \tag{5.6}$$

$$\sum_{t \in E_-} |\omega(t)| \leq \left(\frac{1}{2} - \frac{2}{4^k}\right) \cdot \|\omega\|_1, \tag{5.7}$$

For all univariate polynomials $q\colon \mathbb{R} \to \mathbb{R}$,

$$\deg p < c_1 \sqrt{k^{-1} \cdot T \cdot N^{-1/k}} \implies \sum_{t=0}^{k} \omega(t) \cdot q(t) = 0, \tag{5.8}$$

$$|\omega(t)| \leq \frac{(2k)^k \exp(-c_2 t / \sqrt{k \cdot T \cdot N^{1/k}}) \|\omega\|_1}{t^2} \qquad \forall t = 1, 2, \dots, T. \tag{5.9}$$

Let $c = 2k\lceil N^{1/k} \rceil$, and let $m = \lfloor \sqrt{T/c} \rfloor$. Define the set

$$S = \{1, 2, \dots, k\} \cup \{ci^2 : 0 \leq i \leq m\}.$$

Note that $|S| = \Omega(k^{-1/2} T^{1/2} N^{-1/(2k)})$. Define the polynomial

$$\omega(t) = \frac{(-1)^{t+(T-m)+1}}{T!} \binom{T}{t} \prod_{r \in [T]_0 \setminus S} (t - r).$$

The signs are chosen so that $\omega(k) < 0$. It is immediate from Fact 4.4 that ω satisfies (5.8) for $c_1 = 1/\sqrt{2}$. We now show that (5.9) holds. For $t = 1, \dots, k$, we have

$$\frac{(2k)^k \exp(-c_2 t / \sqrt{k \cdot T \cdot N^{1/k}})}{t^2} \geq \frac{(2k)^k \exp(-c_2 \sqrt{k})}{k^2} \geq 1$$

as long as $c_2 \leq 1/2$ and $k \geq 2$. Since $|\omega(t)| \leq \|\omega\|_1$, the bound holds for $t = 1, \dots, k$.

For $t = cj^2$ with $j \geq 1$, we expand out the binomial coefficient in the definition of ω to obtain

$$|\omega(t)| = \begin{cases} \prod_{r \in S \setminus \{t\}} \frac{1}{|t-r|} & \text{for } t \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For $t \in \{0, 1, \dots, k\}$, we observe that

$$\frac{|\omega(t)|}{|\omega(k)|} = \frac{k! \cdot \prod_{i=1}^{m} (ci^2 - k)}{t! \cdot (k-t)! \cdot \prod_{i=1}^{m} (ci^2 - t)} \leq \binom{k}{t}. \tag{5.10}$$

Meanwhile, for $t = cj^2$ with $j \geq 1$, we get

$$\begin{aligned}
\frac{|\omega(t)|}{|\omega(k)|} &= \frac{k! \cdot \prod_{i=1}^{m}(ci^2 - k)}{\prod_{i=1}^{k}(cj^2 - i) \cdot \prod_{i \in [m]_0 \setminus \{j\}} |ci^2 - cj^2|} \\
&\leq \frac{k! \cdot \prod_{i=1}^{m} ci^2}{(cj^2 - k)^k \cdot \prod_{i \in [m]_0 \setminus \{j\}} c(i+j)|i-j|} \\
&= \frac{k!}{j(cj^2 - k)^k} \cdot \frac{(m!)^2}{(m+j)!(m-j)!}.
\end{aligned}$$

The first factor is bounded above by

$$\frac{k!}{(c-k)^k j^{2k+1}}.$$

As long as $c \geq 2k$ and $k \geq 2$, this expression is at most

$$\frac{k^k}{(c/2)^k j^4} = \frac{(2k)^k}{c^k \cdot j^4}.$$

We control the second factor by

$$\begin{aligned}
\frac{(m!)^2}{(m+j)!(m-j)!} &= \frac{m}{m+j} \cdot \frac{m-1}{m+j-1} \cdot \ldots \cdot \frac{m-j+1}{m+1} \\
&\leq \left(\frac{m}{m+j}\right)^j \\
&\leq \left(1 - \frac{j}{2m}\right)^j \\
&\leq \mathrm{e}^{-j^2/(2m)},
\end{aligned}$$

where the last inequality uses the fact that $1 - x \leq \mathrm{e}^{-x}$ for all x. Hence,

$$\frac{|\omega(cj^2)|}{|\omega(k)|} \leq \frac{(2k)^k}{c^k \cdot j^4} \cdot \mathrm{e}^{-j^2/(2m)}. \tag{5.11}$$

This immediately yields

$$\frac{|\omega(cj^2)|}{\|\omega\|_1} \leq \frac{|\omega(cj^2)|}{|\omega(k)|} \leq \frac{(2k)^k}{(cj^2)^2} \cdot \mathrm{e}^{-cj^2/(2cm)},$$

which establishes (5.9) for all $t = cj^2 > k$.

Moreover, by (5.11)

$$\sum_{t>k} |\omega(t)| \leq |\omega(k)| \cdot \sum_{j=1}^{m} \frac{(2k)^k}{c^k \cdot j^4} \cdot \mathrm{e}^{-j^2/(2m)} \leq \frac{(2k)^k}{c^k} \cdot |\omega(k)| \cdot \sum_{j=1}^{m} \frac{1}{j^4} \leq \frac{|\omega(k)|}{48N}. \tag{5.12}$$

Hence, since $\omega(k) < 0$,

$$\sum_{t \in E_+} |\omega(t)| \leq \sum_{t>k} |\omega(t)| \leq \frac{|\omega(k)|}{48N} \leq \frac{\|\omega\|_1}{48N},$$

which gives (5.6).

Finally, to establish (5.7), we combine (5.10) and (5.12) to obtain

$$\frac{\|\omega\|_1}{|\omega(k)|} \le \sum_{t=0}^{k} \binom{k}{t} + \frac{1}{48N} < 2^k + 1 < \frac{1}{2} \cdot 4^k. \tag{5.13}$$

We calculate

$$\begin{aligned}
\frac{\|\omega\|_1}{2} - \sum_{t \in E_-} |\omega(t)| &= \sum_{t:\omega(t)<0} (-\omega(t)) - \sum_{t \in E_-} (-\omega(t)) \qquad\qquad \text{since } \langle \omega, \mathbf{1} \rangle = 0 \\
&= \sum_{t:\ \omega(t)<0, t \ge k} (-\omega(t)) \\
&\ge -\omega(k).
\end{aligned}$$

Rearranging and applying the bound (5.13),

$$\sum_{t \in E_-} |\omega(t)| \le \left(\frac{1}{2} + \frac{\omega(k)}{\|\omega\|_1} \right) \cdot \|\omega\|_1 \le \left(\frac{1}{2} - 2 \cdot 4^{-k} \right) \cdot \|\omega\|_1.$$

□

Applying Minsky-Papert symmetrization (Lemma 2.7) to ensure that the resulting function has the appropriate pure high degree, Proposition 5.3 yields a dual polynomial for THR_N^k.

Proposition 5.4. *Let $k, T, N \in \mathbb{N}$ with $k \le T \le N$. Define $\psi : \{-1,1\}^N \to \mathbb{R}$ by $\psi(x) = \omega(|x|)/\binom{N}{|x|}$ for $x \in H_{\le T}^N$ and $\psi(x) = 0$ otherwise, where ω is as constructed in Proposition 5.3. Then*

$$\sum_{x \in E_+(\psi, \mathsf{THR}_N^k)} |\psi(x)| \le \frac{1}{48N}, \tag{5.14}$$

$$\sum_{x \in E_-(\psi, \mathsf{THR}_N^k)} |\psi(x)| \le \frac{1}{2} - \frac{2}{4^k}, \tag{5.15}$$

$$\|\psi\|_1 = 1, \tag{5.16}$$

$$\textit{For any polynomial } p \colon \{-1,1\}^N \to \mathbb{R},\ \deg p < c_1 \sqrt{k^{-1} \cdot T \cdot N^{-1/k}} \implies \langle \psi, p \rangle = 0, \tag{5.17}$$

$$\sum_{|x|=t} |\psi(x)| \le (2k)^k \exp(-c_2 t / \sqrt{k \cdot T \cdot N^{1/k}})/t^2 \qquad \forall t = 1, 2, \ldots, N. \tag{5.18}$$

5.2 Step 2: A preliminary dual witness for $\mathsf{OR}_R \circ \mathsf{THR}_N^k$

5.2.1 Refined amplification lemmas

The dual witness Φ for OR_R that we construct will itself be obtained as a dual block composition $\rho \star \varphi$. Each constituent dual polynomial will play a distinct role in showing that $\Phi \star \psi$ is a good dual polynomial for $\mathsf{OR}_R \circ \mathsf{THR}_N^k$. The first function φ is an "error amplifier" in the sense that $\varphi \star \psi$ is much better correlated with $\mathsf{OR}_R \circ \mathsf{THR}_N^k$ than ψ is with THR_N^k. The second function ρ, on the other hand, is a "degree amplifier" in that it serves to increase the pure high degree of $\varphi \star \psi$.

While the amplification results we need are new, they are relatively straightforward extensions of similar results in [29, 30, 73]. Proofs appear below for completeness.

Amplifying error. The following proposition shows that if ψ is a dual witness for the high approximate degree of a Boolean function f, then there is a dual witness of the form $\varphi \star \psi$ for $\mathsf{OR}_M \circ f$ such that (a) $\varphi \star \psi$ may make slightly more false positive errors than ψ (by at most a factor of M), and (b) $\varphi \star \psi$ makes significantly fewer false-negative errors than ψ (exponentially smaller in M).

Proposition 5.5. *Let $f : \{-1,1\}^m \to \{-1,1\}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ be functions such that*

$$\sum_{x \in E_+(\psi, f)} |\psi(x)| \le \delta^+, \tag{5.19}$$

$$\sum_{x \in E_-(\psi, f)} |\psi(x)| \le \delta^-, \tag{5.20}$$

$$\|\psi\|_1 = 1. \tag{5.21}$$

For every $M \in \mathbb{N}$, there exists a function $\varphi : \{-1,1\}^M \to [-1,1]$ with $\|\varphi\|_1 = 1$ and pure high degree 1 such that

$$\sum_{x \in E_+(\varphi \star \psi, \mathsf{OR}_M \circ f)} |(\varphi \star \psi)(x)| \le M \cdot \delta^+, \tag{5.22}$$

$$\sum_{x \in E_-(\varphi \star \psi, \mathsf{OR}_M \circ f)} |(\varphi \star \psi)(x)| \le \frac{1}{2} \cdot (2\delta^-)^M. \tag{5.23}$$

Proof. Let $\varphi : \{-1,1\}^M \to \{-1,1\}$ be defined such that $\varphi(\mathbf{1}) = 1/2$, $\varphi(-\mathbf{1}) = -1/2$, and $\varphi(x) = 0$ for all other x. Notice that

$$\sum_{(x_1,\dots,x_M) \in \{-1,1\}^M} \varphi(x_1, \dots, x_M) = 0 \tag{5.24}$$

so Ψ has pure high degree 1, and that $\|\Psi\|_1 = 1$.

We now prove that Equation 5.22 holds. Let λ be the distribution on $\{-1,1\}^m$ given by $\lambda(x) = |\psi(x)|$, and let $\lambda^{\otimes M}$ be the product distribution on $(\{-1,1\}^m)^M$ given by $\lambda^{\otimes M}(x_1, \dots, x_M) = \prod_{i=1}^M |\psi(x_i)|$. Since ψ is orthogonal to the constant polynomial, it has expected value 0, and hence the string $(\dots, \mathrm{sgn}(\psi(x_i)), \dots)$ is distributed uniformly in $\{-1,1\}^M$ when one samples $(x_1, \dots, x_M)$ according to $\lambda^{\otimes M}$. This allows us to write

$$\begin{aligned}
&\sum_{(x_1,\dots,x_M) \in E^+(\varphi \star \psi, \mathsf{OR}_M \circ f)} |(\varphi \star \psi)(x_1, \dots, x_M)| \\
&= 2^M \mathbf{E}_{\lambda^{\otimes M}}[\varphi(\dots, \mathrm{sgn}(\psi(x_i)), \dots) \cdot \mathbb{I}(\varphi(\dots, \mathrm{sgn}(\psi(x_i)), \dots) > 0 \wedge \mathsf{OR}_M(\dots, f(x_i), \dots) = -1)] \\
&= \sum_{z : \varphi(z) > 0} \varphi(z) \cdot \Pr_{\lambda^{\otimes M}}[\mathsf{OR}_M(\dots, f(x_i), \dots) = -1 | (\dots, \mathrm{sgn}(\psi(x_i)), \dots) = z].
\end{aligned} \tag{5.25}$$

Observe that for any bit b,

$$\Pr_{x\sim\lambda}[f(x) \neq \mathrm{sgn}(\psi(x)) \,|\, \mathrm{sgn}(\psi(x)) = b] = 2\sum_{x\in A_b} |\psi(x)|,$$

where for brevity, we have written $A_{+1} = E_+(\psi, f)$ and $A_{-1} = E_-(\psi, f)$. Therefore, as noted in [75], for any given $z \in \{-1,1\}^M$, the following two random variables are identically distributed:

- The string $(\ldots, f(x_i), \ldots)$, when one chooses $(\ldots, x_i, \ldots)$ from $\lambda^{\otimes M}$ conditioned on $(\ldots, \mathrm{sgn}(\psi(x_i)), \ldots) = z$
- The string $(\ldots, y_i z_i, \ldots)$, where $y \in \{-1,1\}^M$ is a random string whose ith bit independently takes on value -1 with probability $2\sum_{x\in A_{z_i}} |\psi(x)|$.

Thus, Expression (5.25) equals

$$\sum_{z:\varphi(z)>0} \varphi(z) \cdot \Pr_y[\mathsf{OR}_M(\ldots, y_i z_i, \ldots) = -1], \tag{5.26}$$

where $y \in \{-1,1\}^M$ is a random string whose ith bit independently takes on value -1 with probability $2\sum_{x\in A_{z_i}} |\psi(x)|$. The only term in this sum corresponds to $z = \mathbf{1}$, which we now argue is at most $M\delta^+$. By (5.19), each $y_i = -1$ with probability $2\sum_{x\in A_1} |\psi(x)| = 2\sum_{x\in E^+(\psi,f)} |\psi(x)| \leq 2\delta^+$. Hence, for $z = \mathbf{1}$, we have

$$\Pr_y[\mathsf{OR}_M(\ldots, y_i, \ldots) = -1] \leq 2M\delta^+$$

by a union bound. Thus Expression (5.26) is at most $M\delta^+$, proving (5.22).

It now remains to prove the bound (5.23). By an identical argument as above, we have

$$\sum_{(x_1,\ldots,x_M)\in E^-(\varphi\star\psi, \mathsf{OR}_M\circ f)} |(\varphi\star\psi)(x_1,\ldots,x_M)| = \sum_{z:\varphi(z)<0} \varphi(z) \cdot \Pr_y[\mathsf{OR}_M(\ldots, y_i z_i, \ldots) = 1]. \tag{5.27}$$

The only term in the sum corresponds to $z = -\mathbf{1}$, which we argue takes value $\frac{1}{2}\cdot(2\delta^-)^M$. Here, each $y_i = -1$ independently with probability $\sum_{x\in A_{-1}} |\psi(x)| = 2\sum_{x\in E^-(\psi,f)} |\psi(x)| \leq 2\delta^-$, and $\mathsf{OR}_M(\ldots, -y_i, \ldots) = 1$ only if $y_i = -1$ for every i. Hence, we conclude that

$$\Pr_y[\mathsf{OR}_M(\ldots, -y_i, \ldots) = 1] \leq (2\delta^-)^M.$$

It follows that Expression (5.27) is at most $\frac{1}{2}\cdot(2\delta^-)^M$, establishing (5.22). This completes the proof. □

Amplifying degree. The following proposition states that if ψ is a dual polynomial for a Boolean function f, then there is a dual polynomial $\rho\star\psi$ for $\mathsf{OR}_M \circ f$ with significantly larger pure high degree that does not make too many more false positive and false negative errors than does ψ itself.

Proposition 5.6. *Let $f : \{-1,1\}^m \to \{-1,1\}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ be functions such that*

$$\sum_{x \in E_+(\psi,f)} |\psi(x)| \le \delta^+, \tag{5.28}$$

$$\sum_{x \in E_-(\psi,f)} |\psi(x)| \le \delta^-, \tag{5.29}$$

$$\|\psi\|_1 = 1 \tag{5.30}$$

For every $M \in \mathbb{N}$ there exists a function $\rho : \{-1,1\}^M \to \mathbb{R}$ with $\|\rho\|_1 = 1$ and pure high degree $\Omega(\sqrt{M})$ such that

$$\langle \rho \star \psi, \mathsf{OR}_M \circ f \rangle \ge \frac{9}{10} - 4M\delta^+ - 4\delta^-. \tag{5.31}$$

Proof. Lemma 2.6 shows that the function OR_M has $(9/10)$-approximate degree $\Omega(\sqrt{M})$. Hence, Proposition 2.3 guarantees the existence of a function $\rho : \{-1,1\}^M \to \mathbb{R}$ with $\|\rho\|_1 = 1$ and pure high degree $\Omega(\sqrt{M})$ such that

$$\langle \rho, \mathsf{OR}_M \rangle \ge \frac{9}{10}. \tag{5.32}$$

What remains is to establish the correlation bound (5.31). Letting λ denote the distribution $\lambda(x) = |\psi(x)|$ as in the proof of Proposition 5.5, we may write

$$\begin{aligned}
&\sum_{(x_1,\dots,x_M) \in (\{-1,1\}^m)^M} (\varphi \star \psi)(x_1,\dots,x_M) \cdot \mathsf{OR}_M(\dots, f(x_i), \dots) \\
&\quad = 2^M \mathbf{E}_{\lambda^{\otimes M}}[\varphi(\dots, \mathrm{sgn}(\psi(x_i)), \dots) \cdot \mathsf{OR}_M(\dots, f(x_i), \dots)] \\
&\quad = \sum_{z \in \{-1,1\}^M} \varphi(z) \cdot \mathbf{E}_{\lambda^{\otimes M}}[\mathsf{OR}_M(\dots, f(x_i), \dots) | (\dots, \mathrm{sgn}(\psi(x_i)), \dots) = z] \\
&\quad = \sum_{z \in \{-1,1\}^M} \varphi(z) \cdot \mathbf{E}_y[\mathsf{OR}_M(\dots, y_i z_i, \dots)],
\end{aligned} \tag{5.33}$$

where $y \in \{-1,1\}^M$ is a random string whose ith bit independently takes the value -1 with probability $2\sum_{x \in A_{z_i}} |\psi(x)|$. (Here, we are using the abbreviated notation $A_{+1} = E_+(\psi, f)$ and $A_{-1} = E_-(\psi, f)$.) We first consider the contribution of the term corresponding to $z = \mathbf{1}$ to the sum. Here, by a union bound,

$$\begin{aligned}
\mathbf{E}_y[\mathsf{OR}_M(\dots, y_i z_i, \dots)] &= 1 - 2\Pr_y[\mathsf{OR}_M(\dots, y_i, \dots) = -1] \\
&\ge 1 - 2M \cdot \left(2 \sum_{x \in A_{+1}} |\psi(x)| \right) \\
&\ge 1 - 4M\delta^+.
\end{aligned}$$

Hence, the term $z = \mathbf{1}$ contributes $\varphi(\mathbf{1}) \cdot (1 - 4M\delta^+)$ to the sum.

Now we consider the contribution of any term corresponding to $z \ne \mathbf{1}$. Given such a z, let i^* be an index such that $z_{i^*} = -1$. Then we have

$$\begin{aligned} -\mathbf{E}_y[\mathsf{OR}_M(\ldots, y_i z_i, \ldots)] &= 1 - 2\Pr_y[\mathsf{OR}_M(\ldots, y_i z_i, \ldots) = 1] \\ &\geq 1 - 2 \cdot \Pr_{y_{i^*}}[y_{i^*} = -1] \\ &= 1 - 2 \cdot \left(2 \sum_{x \in A_{-1}} |\psi(x)|\right) \\ &\geq 1 - 4\delta^-. \end{aligned}$$

We can now lower bound (5.33) by

$$\begin{aligned} \varphi(\mathbf{1}) \cdot (1 - 4M\delta^+) - \sum_{z \neq \mathbf{1}} \varphi(z)(1 - 4\delta^-) &\geq \sum_{z \in \{-1,1\}^M} \varphi(z)\mathsf{OR}_M(z) - 4M\delta^+|\varphi(\mathbf{1})| - 4\delta^- \sum_{z \neq 1} |\varphi(z)| \\ &\geq \frac{9}{10} - 4M\delta^+ - 4\delta^-. \end{aligned}$$ □

5.2.2 Constructing a dual witness for $\mathsf{OR}_R \circ \mathsf{THR}_N^k$

We now combine our amplification lemmas to construct a dual witness for $\mathsf{OR}_R \circ \mathsf{THR}_N^k$.

Proposition 5.7. *Let $k, T, N, R \in \mathbb{N}$ with $k \leq T \leq R \leq N$ and R divisible by 4^k. Let $\psi : \{-1,1\}^N \to \mathbb{R}$ be a function with $\|\psi\|_1 = 1$ and*

$$\sum_{x \in E_+(\psi, \mathsf{THR}_N^k)} |\psi(x)| \leq \frac{1}{48N},$$

$$\sum_{x \in E_-(\psi, \mathsf{THR}_N^k)} |\psi(x)| \leq \frac{1}{2} - \frac{2}{4^k}.$$

Then there exists a function $\Phi : \{-1,1\}^R \to \mathbb{R}$ with $\|\Phi\|_1 = 1$ and pure high degree $\Omega(2^{-k}\sqrt{R})$ such that

$$\langle \Phi \star \psi, \mathsf{OR}_R \circ \mathsf{THR}_N^k \rangle \geq 2/3.$$

Proof. Using the construction of Proposition 5.5 with $m = N$, $M = 4^k$, and $f = \mathsf{THR}_N^k$, we first obtain a function $\varphi : \{-1,1\}^{4^k} \to \mathbb{R}$ with $\|\varphi\|_1 = 1$ and pure high degree 1 such that

$$\sum_{x \in E_+(\varphi \star \psi, \mathsf{OR}_{4^k} \circ \mathsf{THR}_N^k)} |(\varphi \star \psi)(x)| \leq \frac{4^k}{48N},$$

$$\sum_{x \in E_-(\varphi \star \psi, \mathsf{OR}_{4^k} \circ \mathsf{THR}_N^k)} |(\varphi \star \psi)(x)| \leq \frac{1}{2} \cdot \left(1 - 4 \cdot 4^{-k}\right)^{4^k} \leq \frac{e^{-4}}{2}.$$

Now by the construction of Proposition 5.6 with $m = 4^k \cdot N$, $M = R/4^k$ and $f = \mathsf{OR}_{4^k} \circ \mathsf{THR}_N^k$, there exists a function $\rho : \{-1,1\}^{R/4^k} \to \mathbb{R}$ with $\|\rho\|_1 = 1$ and pure high degree $\Omega(2^{-k}\sqrt{R})$ such that

$$\langle \rho \star (\varphi \star \psi), \mathsf{OR}_{R/4^k} \circ (\mathsf{OR}_{4^k} \circ \mathsf{THR}_N^k) \rangle \geq \frac{9}{10} - \frac{4R}{48N} - 2e^{-4} \geq \frac{2}{3}.$$

Let $\Phi : \{-1,1\}^R \to \mathbb{R}$ be the dual block composition $\rho \star \varphi$. Since the dual block composition preserves ℓ_1-norms (Proposition 2.20, Condition (2.8)) and multiplies pure high degrees (Proposition 2.20, Condition (2.9)), the function Φ itself has ℓ_1-norm 1 and pure high degree $\Omega(2^{-k}\sqrt{R})$. The claim now follows from the associativity of dual block composition (Proposition 2.20, Condition (2.10)) and the fact that $\mathsf{OR}_R = \mathsf{OR}_{R/4^k} \circ \mathsf{OR}_{4^k}$. $\square$

5.3 Step 3: Completing the construction

We are now ready to apply Proposition 2.22 to zero out the mass that the construction of Proposition 5.7 places on inputs outside of $H^{N\cdot R}_{\leq N}$.

Proposition 5.8. *Let R be sufficiently large. There exist $N = O((2k)^{k/2} \cdot R)$, $D = \tilde{\Omega}(2^{-k/4}k^{(k-3)/4} \cdot R^{3/4-1/(2k)})$, and $\zeta : (\{-1,1\}^N)^R \to \mathbb{R}$ such that*

$$\zeta(x) = 0 \text{ for all } x \notin H^{N\cdot R}_{\leq N}, \tag{5.34}$$

$$\sum_{x \in H^{N\cdot R}_{\leq N}} \zeta(x) \cdot (\mathsf{OR}_R \circ \mathsf{THR}^k_N)(x) > 1/3, \tag{5.35}$$

$$\|\zeta\|_1 = 1, \text{ and} \tag{5.36}$$

$$\text{For every polynomial } p\colon (\{-1,1\}^N)^R \to \mathbb{R} \text{ of degree less than } D, \text{ we have } \langle p, \zeta \rangle = 0. \tag{5.37}$$

Proof. We start by fixing choices of several key parameters:

- $d = \Theta(2^{-k}\sqrt{R})$ is the pure high degree of the dual witness Φ for OR_R in Proposition 5.7,
- $T = \lfloor (8k)^{k/2}\sqrt{R} \rfloor$,
- $\alpha = (2k)^k$,
- $N = \lceil 20\sqrt{\alpha} \rceil R = \Theta((2k)^{k/2}R)$,
- $\hat{D} = c_1\sqrt{k^{-1} \cdot T \cdot N^{-1/k}} \cdot d = \Theta(2^{-k/4}k^{(k-3)/4} \cdot R^{3/4-1/(2k)})$, where c_1 is the constant from Proposition 5.3,
- $\beta = c_2/\sqrt{k \cdot T \cdot N^{1/k}} = \Theta(2^{-3k/4}k^{(-k-3)/4} \cdot R^{-1/4-1/(2k)})$, where c_2 is the constant from Proposition 5.3.

Let $\psi : \{-1,1\}^N \to [-1,1]$ be the function constructed in Proposition 5.4. Let $\Phi : \{-1,1\}^R \to [-1,1]$ be the function constructed in Proposition 5.7, and define $\xi = \Phi \star \psi$. Then by Proposition 2.20, ξ satisfies the following conditions:

$$\langle \xi, \mathsf{OR}_R \circ \mathsf{THR}^k_N \rangle > 2/3, \tag{5.38}$$

$$\|\xi\|_1 = 1, \tag{5.39}$$

$$\text{For every polynomial } p \text{ of degree less than } \hat{D}, \text{ we have } \langle \xi, p \rangle = 0. \tag{5.40}$$

Recall that ψ was obtained by symmetrizing the function ω constructed in Proposition 5.3. Proposition 2.22 guarantees that for some $\Delta \geq \beta\sqrt{\alpha}R/(4\ln^2 R) = \tilde{\Omega}(2^{-k/4}k^{(k-3)/4} \cdot R^{3/4-1/(2k)})$, the function ξ can be modified to produce a function $\zeta : (\{-1,1\}^N)^R \to \mathbb{R}$ such that

$$\begin{aligned}
&\zeta(x) = 0 \text{ for all } x \notin H^{N\cdot R}_{\leq N},\\
&\langle \zeta, \mathsf{OR}_R \circ \mathsf{THR}^k_N \rangle \geq \langle \xi, \mathsf{OR}_R \circ \mathsf{THR}^k_N \rangle - \|\zeta - \xi\|_1 \geq 2/3 - 2/9 > 1/3,\\
&\|\zeta\|_1 = 1,\\
&\text{For every polynomial } p \text{ of degree less than } \min\{\hat{D}, \Delta\}, \text{ we have } \langle \zeta, p \rangle = 0.
\end{aligned}$$

Observing that

$$D = \min\{\hat{D}, \Delta\} = \tilde{\Omega}(2^{-k/4}k^{(k-3)/4} \cdot R^{3/4-1/(2k)})$$

shows that the function ζ satisfies the conditions necessary to prove Proposition 5.8. □

Theorem 5.2 now follows by combining Proposition 5.8 with the dual characterization of unbounded approximate degree Proposition 2.5. The approximate degree lower bound in Theorem 5.1 is then a consequence of Proposition 2.13 and Corollary 2.18. The quantum query lower bound follows via the standard fact that the ε-error quantum query complexity of f is lower bounded by $1/2 \cdot \widetilde{\deg}_{2\varepsilon}(f)$ [13].

6 Lower bound for image size testing and its implications

6.1 Image size testing

The image size testing problem (IST for short) is defined as follows.

Definition 6.1. Given an input $s = (s_1, \ldots, s_N) \in [R]_0^N$, and $i \in [R]$, let $f_i = |\{j : s_j = i\}|$. The *image size* of s is the number of $i \in [R]$ such that $f_i > 0$. For $0 < \gamma < 1$, define:

$$\mathsf{IST}^{\gamma}_{N,R}(s_1, \ldots, s_N) = \begin{cases} -1 & \text{if the image size is } R,\\ 1 & \text{if the image size is at most } \gamma \cdot R,\\ \text{undefined} & \text{otherwise.} \end{cases}$$

Observe that the definition of IST ignores whether or not the range item 0 has positive frequency, just like the functions dSURJ and dDIST^k. We choose to define IST in this manner to streamline our analysis.

The goal of this section is to prove the following lower bound.

Theorem 6.2. *For some constant $c > 0$, and any constant $\gamma \in (0,1)$, $\widetilde{\deg}\left(\mathsf{IST}^{\gamma}_{N,R}\right) \geq \tilde{\Omega}(R^{1/2})$, where $N = c \cdot \gamma^{-1/2} \cdot R$. The same lower bound applies to the quantum query complexity of $\mathsf{IST}^{\gamma}_{N,R}$.*

Remark 6.3. It is possible to refine our analysis to show that even the *unbounded* approximate degree of $\mathsf{IST}^{\gamma}_{N,R}$ is $\tilde{\Omega}(R^{1/2})$, and that this holds even if the error parameter is $1 - 2^{-n^{\Omega(1)}}$. However, for brevity we do not explicitly establish this stronger result. We direct the interested reader to the subsequent work [34], which shows that the threshold degree of $\mathsf{SURJ}_{N,R}$ is $\tilde{\Omega}(R^{1/2})$ for $R \leq N/2$. The proof of that result can be extended with little difficulty to show that the threshold degree of $\mathsf{IST}^{\gamma}_{N,R}$ is $\tilde{\Omega}(R^{1/2})$.

6.1.1 Connecting symmetric promise properties and block compositions of partial functions

For any function f and symmetric function g, Section 2.5 described a connection between the symmetric property

$$F^{\text{prop}}(s_1,\ldots,s_N) = f(g(\mathbb{1}[s_1=1],\ldots,\mathbb{1}[s_N=1]),\ldots,g(\mathbb{1}[s_1=R],\ldots,\mathbb{1}[s_N=R]))$$

and the partial function

$$F^{\leq N}(x_1,\ldots,x_R) = \begin{cases} f(g(x_1),\ldots,g(x_R)) & \text{if } x_1,\ldots,x_R \in \{-1,1\}^N, |x_1|+\cdots+|x_R| \leq N, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

For simplicity and clarity, that discussion was restricted to total functions f and g (in particular, this avoided having to address the possibility that $f(g(x_1),\ldots,g(x_R))$ is undefined in the definitions of F^{prop} and $F^{\leq N}$). Because IST is a partial function, we need to explain that the same connection still holds even when f is a partial function. To do this, we need to introduce the notion of the *double-promise* approximate degree of $F^{\leq N}$.

Definition 6.4. Let $Y \subset \{-1,1\}^R$ and $f\colon Y \to \{-1,1\}$, and let $g\colon \{-1,1\}^N \to \{-1,1\}$ be a symmetric (total) function. Let $G = \{x_1,\ldots,x_R\colon (g(x_1),\ldots,g(x_R)) \in Y\}$. Let $F^{\leq N}$ be defined as above. Observe that $F^{\leq N}$ is defined at all inputs in $H^{N\cdot R}_{\leq N} \cap G$. The double-promise ε-approximate degree of $F^{\leq N}$, denoted $\widetilde{\text{dpdeg}}(F^{\leq N})$ is the least degree of a real polynomial p such that:

$$|p(x) - F^{\leq N}(x)| \leq \varepsilon \text{ for all } x \in H^{N\cdot R}_{\leq N} \cap G, \tag{6.1}$$

$$|p(x)| \leq 1+\varepsilon \text{ for all } x \in H^{N\cdot R}_{\leq N} \setminus G. \tag{6.2}$$

Observe that in the definition above, no restriction is placed on $p(x)$ for any inputs that are not in $H^{N\cdot R}_{\leq N}$.

Bun and Thaler's analysis from [33] (cf. Theorem 2.15) applies to partial functions f in the following way.

Theorem 6.5 (Bun and Thaler [33]). *Let $Y \subset \{-1,1\}^R$ and let $f : Y \to \{-1,1\}$ be any partial function. Let $g : \{-1,1\}^N \to \{-1,1\}$ be any symmetric function. Then for F^{prop} and $F^{\leq N}$ defined above, and for any $\varepsilon > 0$, we have*

$$\widetilde{\deg}_\varepsilon(F^{\text{prop}}) \geq \widetilde{\text{dpdeg}}_\varepsilon(F^{\leq N}).$$

We will require the following dual formulation of $\widetilde{\text{dpdeg}}_\varepsilon(F^{\leq N})$.

Proposition 6.6. *Let $F^{\leq N}$ and G be defined as above. Then $\widetilde{\text{dpdeg}}_\varepsilon(F^{\leq N}) \geq d$ if and only if there exists a function $\psi\colon \{-1,1\}^n \to \mathbb{R}$ satisfying the following conditions.*

$$\psi(x) = 0 \textit{ for all } x \notin H^{N\cdot R}_{\leq N}, \tag{6.3}$$

$$\sum_{x \in H^{N\cdot R}_{\leq N} \cap G} \psi(x)\cdot F^{\leq N}(x) - \sum_{x \in H^{N\cdot R}_{\leq N} \setminus G} |\psi(x)| > \varepsilon, \tag{6.4}$$

$$\sum_{x \in \{-1,1\}^n} |\psi(x)| = 1, \textit{ and} \tag{6.5}$$

$$\textit{For every polynomial } p\colon \{-1,1\}^n \to \mathbb{R} \textit{ of degree less than } d, \sum_{x\in\{-1,1\}^n} p(x)\cdot\psi(x) = 0. \tag{6.6}$$

We will need to define the following partial function.

Definition 6.7. Define $\mathsf{GapAND}_R^{\gamma}\colon H^R_{\leq(\gamma\cdot R)} \cup \{-\mathbf{1}\} \to \{-1,1\}$ via:

$$\mathsf{GapAND}_R^{\gamma}(x) = \begin{cases} -1 & \text{if } x_i = -1 \text{ for all } i, \\ 1 & \text{if } x \in H^R_{\leq(\gamma\cdot R)}, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

In the case where $f = \mathsf{GapAND}_R^{\gamma}$ and $g = \mathsf{OR}_N$, the function $F^{\text{prop}}(s_1,\dots,s_N)$ is precisely the $\mathsf{IST}^{\gamma}_{N,R}$ function. Hence:

Corollary 6.8. *Let $N,R \in \mathbb{N}$. Let Y be the domain of $\mathsf{GapAND}_R^{\gamma}$, and let $G = \{(x_1,\dots,x_R) \in \{-1,1\}^{N\cdot R} \colon (\mathsf{OR}_N(x_1),\dots,\mathsf{OR}_N(x_R)) \in Y\}$. Then for any $\varepsilon > 0$,*

$$\widetilde{\deg}_{\varepsilon}(\mathsf{IST}^{\gamma}_{N,R}) \geq \widetilde{\mathrm{dpdeg}}_{\varepsilon}(F^{\leq N})$$

where $F^{\leq N} : G \cap H^{N\cdot R}_{\leq N} \to \{-1,1\}$ is the partial function obtained by restricting $\mathsf{GapAND}_R^{\gamma} \circ \mathsf{OR}_N$ to $H^{N\cdot R}_{\leq N}$.

With Corollary 6.8 in hand, we now turn to proving a lower bound on $\widetilde{\mathrm{dpdeg}}_{\varepsilon}(F^{\leq N})$.

6.1.2 Completing the proof of Theorem 6.2

Proof. We construct a dual polynomial to witness the lower bound in Theorem 6.2.

Define the parameters

- $\delta = \gamma/4$,
- $\alpha = 170/\delta$,
- $T = N = \lceil 20\sqrt{\alpha}\rceil R \leq 310 \cdot \gamma^{-1/2} \cdot R$,
- $\beta = c_2 \cdot \sqrt{\delta}/\sqrt{T}$, where c_2 is the constant from Proposition 4.3.

Let ψ be the dual witness for OR_N from Proposition 4.5 with $T = N$. Define $\Phi\colon \{-1,1\}^R \to \mathbb{R}$ as follows:

$$\Phi(x) = \begin{cases} -1/2 & \text{if } x = (-1,-1,\dots,-1), \\ 1/2 & \text{if } x = (1,1,\dots,1), \\ 0 & \text{otherwise.} \end{cases}$$

The dual block composition $\Phi \star \psi$ is the same dual witness for $\mathsf{AND}_R \circ \mathsf{OR}_N$ which Bun and Thaler [30] used to show that $\widetilde{\deg}_{\varepsilon}(\mathsf{AND}_R \circ \mathsf{OR}_N) = \Omega(N^{1/2})$ for $\varepsilon = 1 - 2^{-R}$. This dual witness was also used in subsequent work [21, 79]. Curiously, we are interested in this dual witness for a completely different reason than these prior works. These prior works were interested in $\Phi \star \psi$ because its correlation with the target function $\mathsf{AND}_R \circ \mathsf{OR}_N$ is exponentially closer to 1 than is the correlation of ψ with OR_N. For our purposes, it will not be essential to exploit such a strong correlation guarantee—rather, we are interested in $\Phi\star\psi$ because most of its "ℓ_1-mass" lies on inputs either with full image or tiny image (i. e., most of its mass lies in the domain of $\mathsf{GapAND}_R^\gamma\circ\mathsf{OR}_N$).

As in Corollary 6.8, let G denote the set of all inputs on which $\mathsf{GapAND}_R^\gamma\circ\mathsf{OR}_N$ is defined, i. e.,

$$G=\{x_1,\dots,x_R\in\{-1,1\}^{N\cdot R}\colon (\mathsf{OR}_N(x_1),\dots,\mathsf{OR}_N(x_R))\in H^R_{\le\gamma\cdot R}\cup\{-\mathbf{1}\}\}.$$

The analysis in these prior works [21, 30] implies that $\Phi\star\psi$ satisfies the following three conditions.

$$\|\Phi\star\psi\|_1=1, \tag{6.7}$$
$$\sum_{x\in G}(\Phi\star\psi)(x)\cdot\left(\mathsf{GapAND}_R^\gamma\circ\mathsf{OR}_N\right)(x)-\sum_{x\in\{-1,1\}^{N\cdot R}\setminus G}|(\Phi\star\psi)(x)|\ge 9/10, \tag{6.8}$$
$$\text{For any polynomial } p\colon\{-1,1\}^{N\cdot R}\to\mathbb{R},\ \deg p<c_1\sqrt{\delta T}\implies\langle\Phi\star\psi,p\rangle=0, \tag{6.9}$$

where c_1 is the constant from Proposition 4.5. Indeed, Properties (6.7) and (6.9) are immediate from Proposition 2.20 on the properties of dual block composition. For completeness, we prove that Property (6.8) holds in Section 6.1.3 below, making use of the fact that $\delta=\gamma/4$ and taking R to be sufficiently large.

Proposition 2.22 (with α and β set as above) guarantees that for some $\Delta\ge\beta\sqrt{\alpha}R/(4\ln^2 R)=\tilde{\Omega}(R^{1/2})$, the function $\Phi\star\psi$ can be modified to produce a function $\zeta:(\{-1,1\}^N)^R\to\mathbb{R}$ such that

$$\begin{aligned}&\zeta(x)=0\text{ for all }x\notin H^{N\cdot R}_{\le N},\\&\langle\zeta,\mathsf{GapAND}_R^\gamma\circ\mathsf{OR}_N\rangle\ge\langle\Phi\star\psi,\mathsf{GapAND}_R^\gamma\circ\mathsf{OR}_N\rangle-\|\zeta-\Phi\star\psi\|_1\ge 9/10-2/9>1/3,\\&\|\zeta\|_1=1,\\&\text{For every polynomial } p \text{ of degree less than } D:=\min\{\hat{D},\Delta\},\text{ we have } \langle\zeta,p\rangle=0.\end{aligned}$$

Observing that

$$D=\min\{c_1\sqrt{\delta T},\Delta\}=\tilde{\Omega}(R^{1/2})$$

shows that the function ζ satisfies the conditions necessary to prove Theorem 6.2 via Proposition 6.6. □

6.1.3 Proof of Condition (6.8)

Lemma 6.9. *Let $\delta>0$, $\gamma>2\delta$, and let*

$$G=\{x_1,\dots,x_R\in\{-1,1\}^{N\cdot R}\colon (\mathsf{OR}_N(x_1),\dots,\mathsf{OR}_N(x_R))\in H^R_{\le\gamma\cdot R}\cup\{-\mathbf{1}\}\}.$$

Define $\Phi:\{-1,1\}^R\to[-1,1]$ by $\Phi(-\mathbf{1})=-1/2$, $\Phi(\mathbf{1})=1/2$ and $\Phi(z)=0$ otherwise. Let $\psi:\{-1,1\}^N\to[-1,1]$ be any dual witness for OR_N such that $\|\psi\|_1=1$, $\langle\psi,\mathbf{1}\rangle=0$, and $\langle\psi,\mathsf{OR}_N\rangle\ge 1-\delta$. Then

$$\sum_{x\in G}(\Phi\star\psi)(x)\cdot\left(\mathsf{GapAND}_R^\gamma\circ\mathsf{OR}_N\right)(x)-\sum_{x\in\{-1,1\}^{N\cdot R}\setminus G}|(\Phi\star\psi)(x)|\ge 1-\delta^R-\exp(-(\gamma-\delta)R/3).$$

The proof of Lemma 6.9 crucially relies on a special property, called *one-sided error*, that is satisfied by any dual polynomial for OR.

Definition 6.10. Let $f : \{-1,1\}^N \to \{-1,1\}$ and let $\psi : \{-1,1\}^N \to \mathbb{R}$. We say that ψ has one-sided error with respect to f if for all $x \in \{-1,1\}^N$,

$$f(x) = 1 \implies \psi(x) > 0. \tag{6.10}$$

The following lemma shows that *any* dual witness for the OR_N function has one-sided error.

Lemma 6.11 (Gavinsky and Sherstov [41])**.** *Let $\psi : \{-1,1\}^N \to \mathbb{R}$ be a function with pure high degree at least 1 such that $\langle \psi, \mathsf{OR}_N \rangle > 0$. Then ψ has one-sided error with respect to OR_N.*

In particular, if ψ is such a dual witness for OR_N, then we have

$$\sum_{x \in A_{+1}} |\psi(x)| \leq \frac{1}{2}(1 - \langle \psi, \mathsf{OR}_N \rangle), \qquad \sum_{x \in A_{-1}} |\psi(x)| = 0, \tag{6.11}$$

where the sets A_{+1} and A_{-1} are, respectively, the sets of false positive and false negative errors given by

$$\begin{aligned} A_{+1} &= E_+(\psi, \mathsf{OR}_N) = \{x \in \{-1,1\}^N : \psi(x) > 0, \mathsf{OR}_N(x) = -1\}, \\ A_{-1} &= E_-(\psi, \mathsf{OR}_N) = \{x \in \{-1,1\}^N : \psi(x) < 0, \mathsf{OR}_N(x) = +1\}. \end{aligned}$$

Proof of Lemma 6.9. We begin by observing that the quantity of interest can be written as

$$\begin{aligned} &\sum_{x \in \{-1,1\}^{N \cdot R}} (\Phi \star \psi)(x) \cdot (\mathsf{AND}_R \circ \mathsf{OR}_N)(x) - \\ &\qquad \left(\sum_{x \in \{-1,1\}^{N \cdot R} \setminus G} (\Phi \star \psi)(x) \cdot (\mathsf{AND}_R \circ \mathsf{OR}_N)(x) + \sum_{x \in \{-1,1\}^{N \cdot R} \setminus G} |(\Phi \star \psi)(x)| \right) \\ &\geq \sum_{x \in \{-1,1\}^{N \cdot R}} (\Phi \star \psi)(x) \cdot (\mathsf{AND}_R \circ \mathsf{OR}_N)(x) - 2 \sum_{x \in \{-1,1\}^{N \cdot R} \setminus G} |(\Phi \star \psi)(x)|. \end{aligned} \tag{6.12}$$

We estimate each term of Expression (6.12) separately, beginning with the first term. Just as in the proofs of Proposition 5.5 and Proposition 5.6, we have

$$\sum_{x \in \{-1,1\}^{N \cdot R}} (\Phi \star \psi)(x) \cdot (\mathsf{AND}_R \circ \mathsf{OR}_N)(x) = \sum_{z \in \{-1,1\}^R} \Phi(z) \cdot \mathbb{E}_y[\mathsf{AND}_R(\ldots, y_i z_i, \ldots)]$$

where $y \in \{-1,1\}^R$ is a random string whose ith bit independently takes the value -1 with probability $2\sum_{x \in A_{z_i}} |\psi(x)|$. For $z = -\mathbf{1}$, we have by (6.11) that $2\sum_{x \in A_{-1}} |\psi(x)| = 0$, so the contribution of the corresponding term to the sum is $1/2$. For $z = \mathbf{1}$, we use the fact that $2\sum_{x \in A_{+1}} |\psi(x)| \leq \delta$ to compute

$$\begin{aligned} \frac{1}{2} \cdot \mathbb{E}_y[\mathsf{AND}_R(\ldots, y_i, \ldots)] &= \frac{1}{2} \cdot \left(1 - 2 \Pr_y[\mathsf{AND}_R(\ldots, y_i, \ldots) = -1]\right) \\ &\geq \frac{1}{2} \cdot \left(1 - 2\delta^R\right). \end{aligned}$$

Hence, the first summand of (6.12) is at least $1 - \delta^R$.

We now estimate the second summand, $2\sum_{x\notin G}|(\Phi\star\psi)(x)|$. As in the proofs of Proposition 5.5 and Proposition 5.6, we let λ denote the distribution with probability mass function $\lambda(x)=|\psi(x)|$. Then

$$\begin{aligned}2\sum_{x\notin G}|(\Phi\star\psi)(x)| &= 2^{R+1}\mathbb{E}_{\lambda^{\otimes R}}[|\Phi(\dots,\operatorname{sgn}\psi(x_i),\dots)|\cdot\mathbb{I}(x\notin G)]\\ &= 2\sum_{z\in\{-1,1\}^R}|\Phi(z)|\cdot\Pr_{\lambda^{\otimes R}}[x\notin G|(\dots,\operatorname{sgn}(\psi(x_i)),\dots)=z]\\ &= \Pr_{\lambda^{\otimes R}}[x\notin G|(\dots,\operatorname{sgn}(\psi(x_i)),\dots)=-\mathbf{1}]+\Pr_{\lambda^{\otimes R}}[x\notin G|(\dots,\operatorname{sgn}(\psi(x_i)),\dots)=\mathbf{1}].\end{aligned}$$

We analyze each term of this sum separately. For the first term, observe that by one-sided error of ψ, it follows that if $x=(x_1,\dots,x_R)$ is any input $\operatorname{sgn}(\psi(x_i))=-1$ for all i then $\mathsf{OR}_N(x_i)=-1$ for all i. Thus we are guaranteed that $x\in G$, so the contribution of the first summand is zero. To analyze the second summand, let us denote by $r_i\in\{0,1\}$ the indicator random variable for the event $\mathsf{OR}_N(x_i)=-1$ when x_i is drawn from the conditional distribution $(\lambda\,|\operatorname{sgn}(\psi(x_i))=1)$. Then

$$\begin{aligned}\Pr[r_i=1] &= \Pr_{x_i\sim\lambda}[\mathsf{OR}_N(x_i)=-1\,|\operatorname{sgn}(\psi(x_i))=1]\\ &= 2\sum_{x\in A_{+1}}|\psi(x_i)|\\ &\le \delta\end{aligned}$$

by (6.11). Hence,

$$\begin{aligned}\Pr_{\lambda^{\otimes R}}[x\notin G|(\dots,\operatorname{sgn}(\psi(x_i)),\dots)=\mathbf{1}] &\le \Pr\left[\sum_{i=1}^R r_i>\gamma R\right]\\ &\le \exp\left(-\frac{(\gamma-\delta)R}{3}\right)\end{aligned}$$

by the multiplicative Chernoff bound.[6] Thus, we have

$$2\sum_{x\notin G}|(\Phi\star\psi)(x)|\le\exp\left(-\frac{(\gamma-\delta)R}{3}\right).$$

Putting everything together, we see that Expression (6.12) is at least $1-\delta^R-\exp(-(\gamma-\delta)R/3)$ as we wanted to show. □

6.2 Lower bound for junta testing

It follows from a reduction in Ambainis et al. [9, Section 6] that for any $N=O(R)$ and sufficiently small constant $\gamma>0$, a $\tilde{\Omega}(R^{1/2})$ lower bound for the approximate degree or quantum query complexity of $\mathsf{IST}^{\gamma}_{N,R}$ implies an $\tilde{\Omega}(k^{1/2})$ approximate degree or quantum query lower bound for k-junta testing for proximity parameter $\varepsilon=1/3$. Hence, Theorem 6.2 has the following corollary.

[6] The formulation we use here is as follows. Let $r_1,\dots,r_R$ be independent $\{0,1\}$-valued random variables, $S=\sum_{i=1}^R r_i$, and $\mu=\mathbb{E}[S]$. Then for $\eta>1$, we have $\Pr[S>(1+\eta)\mu]\le\exp(-\eta\mu/3)$. In this application, we are taking $\mu\le\delta R$ and $\eta=\gamma R/\mu-1>1$.

Corollary 6.12. *Any quantum tester that distinguishes k-juntas from functions that are $(1/3)$-far from any k-junta with error probability at most $1/3$ makes $\tilde{\Omega}(k^{1/2})$ queries to the function.*

6.3 Lower bound for SDU

The goal of this section is to derive a lower bound for approximating the statistical distance of an input distribution from uniform up to some additive constant error. We formalize this problem as follows.

Given an input $(s_1,\dots,s_N)\in[R]^N$, and $i\in[R]$, let $f_i=|\{j\colon s_j=i\}|$, and let p be the probability distribution over $[R]$ such that $p_i=f_i/N$. For $N\geq R$ and $0<\gamma_2<\gamma_1<1$, define the partial function $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$ as follows.

Definition 6.13. Define

$$\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}(s_1,\dots,s_N)=\begin{cases}-1 & \text{if } \frac{1}{2}\sum_{i=1}^{R}|p_i-1/R|\leq\gamma_1,\\ 1 & \text{if } \frac{1}{2}\sum_{i=1}^{R}|p_i-1/R|\geq\gamma_2,\\ \text{undefined} & \text{otherwise.}\end{cases}$$

Above, $\frac{1}{2}\sum_{i=1}^{R}|p_i-1/R|$ is the statistical distance between p and the uniform distribution.

The SDU problem reduces to IST in the sense that any approximating polynomial for SDU implies the existence of an approximation to IST of the same degree. Hence, the approximate degree of SDU is at least as large as that of IST. For intuition as to why this is true, let us relate $\mathsf{IST}^{1/2}_{N,R}$ to $\mathsf{SDU}^{1/2,0}_{N,R}$ in the special case where $N=R$ and no dummy (i. e., 0) items appear in the input to IST. If $(s_1,\dots,s_N)$ is a true input to $\mathsf{IST}^{1/2}_{N,R}$, i. e., $\mathsf{IST}^{1/2}_{N,R}(s_1,\dots,s_N)=-1$, then every index $i\in[R]$ must appear in the input list exactly once. Hence, the distribution represented by $(s_1,\dots,s_N)$ is exactly uniform, so $\mathsf{SDU}^{1/2,0}_{N,R}(s_1,\dots,s_N)=-1$. On the other hand, if $\mathsf{IST}^{1/2}_{N,R}(s_1,\dots,s_N)=1$, then at most $R/2$ indices $i\in[R]$ appear in the input list, so the list represents a distribution with statistical distance at least $1/2$ from uniform. Thus, an approximating polynomial for $\mathsf{SDU}^{1/2,0}_{N,R}$ is also an approximating polynomial for $\mathsf{IST}^{1/2}_{N,R}$.

We will actually need a more general relationship between the approximate degrees of SDU and IST to handle the fact that we cannot take N to be exactly equal to R in the IST lower bound, as well as to handle the occurrences of dummy items in the definition of IST.

Theorem 6.14. *For some $N=O(R)$, and some constants $0<\gamma_2<\gamma_1<1$, $\widetilde{\deg}\left(\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}\right)\geq\tilde{\Omega}(R^{1/2})$. The same lower bound applies to the quantum query complexity of $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$.*

Proof. Fix $R>0$, let c be the constant from Theorem 6.2, $\gamma<(2/3c)^2$ be a sufficiently small constant, and $N=c\cdot\gamma^{-1/2}\cdot R$. As inputs in $\mathsf{IST}^{\gamma}_{N,R}$ are interpreted as elements of $[R]_0^N$, we can equivalently interpret them as elements of $[R+1]^N$, i. e., as inputs to $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R+1}$, for any desired $0<\gamma_1<\gamma_2<1$.

Set $\gamma_1=1-3\gamma/2$ and $\gamma_2=1-\gamma^{1/2}/c$. Observe that since $\gamma<(2/3c)^2$, γ_1 is strictly greater than γ_2. We claim that

$$\left(\mathsf{IST}^{\gamma}_{N,R}\right)^{-1}(-1)\subseteq\left(\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R+1}\right)^{-1}(-1),\text{ and}$$
$$\left(\mathsf{IST}^{\gamma}_{N,R}\right)^{-1}(+1)\subseteq\left(\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R+1}\right)^{-1}(+1).$$

Indeed, since inputs in $\left(\mathsf{IST}^{\gamma}_{N,R}\right)^{-1}(-1)$ define a probability distribution over $[R+1]$ with support size at least R, with all probabilities being integer multiples of $1/N = \gamma^{1/2}/(cR)$, the statistical distance between any such distribution and the uniform distribution is at most $1-R/N = 1-\gamma^{1/2}/c$. This follows from the following calculation. Amongst probability distributions $(p_1,\dots,p_{R+1})$ over $[R+1]$ with support size at least R and all probabilities p_i being integer multiples of $1/N$, it is not hard to see that one maximizes the statistical distance from the uniform distribution over $[R+1]$ by setting $p_1 = 1-\frac{R-1}{N}$, $p_2 = p_3 = \cdots = p_R = 1/N$, and $p_{R+1} = 0$. The statistical distance from uniform is

$$\frac{1}{2}\left(\left(1-\frac{R-1}{N}-\frac{1}{R+1}\right)+(R-1)\left(\frac{1}{R+1}-\frac{1}{N}\right)+\frac{1}{R+1}\right)$$
$$=1-\frac{R}{N}+\frac{1}{N}-\frac{1}{R+1}\le 1-\frac{R}{N},$$

where we have assumed that $N \ge R+1$, which is true for sufficiently small choice of γ.

Similarly, since inputs in $\left(\mathsf{IST}^{\gamma}_{N,R}\right)^{-1}(1)$ define a probability distribution over $[R+1]$ with support size at most $\gamma\cdot R+1$, the statistical distance between any such distribution and the uniform distribution is at least $1-3\gamma/2$. To see this, let $p=(p_1,\dots,p_{R+1})$ be any distribution of support size at most $\gamma\cdot R$, and let $S=\{i\colon p_i=0\}$ and $\bar{S}$ be the complement of S. Then the statistical distance of p from uniform is at least

$$\frac{1}{2}\left(\sum_{i\in S}\left(p_i-\frac{1}{R+1}\right)\right)+\frac{1}{2}\left(\sum_{i\in\bar{S}}\left(p_i-\frac{1}{R+1}\right)\right)=\frac{1}{2}\left(\frac{|S|}{R+1}+\left(\sum_{i\in\bar{S}}p_i\right)-\frac{|\bar{S}|}{R+1}\right)$$
$$=\frac{1}{2}\left(\frac{|S|}{R+1}+1-\frac{|\bar{S}|}{R+1}\right)\ge\frac{1}{2}\left(\frac{(1-\gamma)R}{R+1}+1-\frac{\gamma R+1}{R+1}\right)=\frac{1}{2}\left(1+\frac{R-2\gamma R-1}{R+1}\right)$$
$$=1-\frac{\gamma R+1}{R+1}=1-\gamma-\frac{\gamma+1}{R+1}\ge 1-3\gamma/2.$$

It is an easy consequence of the above that any ε-approximating polynomial of degree d for $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R+1}$ implies an approximation to $\mathsf{IST}^{\gamma}_{N,R}$ of the same degree (i. e., that $\widetilde{\deg}_\varepsilon(\mathsf{IST}^{\gamma}_{N,R})\le d$). Theorem 6.14 then follows from Theorem 6.2. $\square$

6.4 Lower bound for entropy comparison and approximation

Given a distribution p over $[R]$, the Shannon entropy of p, denoted $H(p)$, is defined to be $H(p) := \sum_{i\in[R]} p_i \log_2(1/p_i)$. Following Goldreich and Vadhan [42], we define a partial function $\mathsf{GapCmprEnt}^{\alpha,\beta}_{N,R}$ capturing the problem of comparing the entropies of two distributions.

The function $\mathsf{GapCmprEnt}^{\alpha,\beta}_{N,R}$ takes as input two vectors in $[R]^N$ and interprets each vector $i\in\{1,2\}$ as a probability distribution p_i over $[R]$, with $p_i(j) = f_{i,j}/N$ where $f_{i,j}$ is the frequency of j in the ith vector. The function $\mathsf{GapCmprEnt}^{\alpha,\beta}_{N,R}$ evaluates to

$$\begin{cases}-1 & \text{if } H(p_1)-H(p_2)\le\beta,\\ 1 & \text{if } H(p_1)-H(p_2)\ge\alpha,\\ \text{undefined} & \text{otherwise.}\end{cases}$$

Theorem 6.15. *There exist constants* $0 < \beta < \alpha < 1$ *such that* $\widetilde{\deg}(\mathsf{GapCmprEnt}^{\alpha,\beta}_{N,R}) = \tilde{\Omega}(R^{1/2})$. *The same lower bound applies to the quantum query complexity of* $\mathsf{GapCmprEnt}^{\alpha,\beta}_{N,R}$.

Proof. Vadhan [88, Claim 4.4.2 and Remark 4.4.3] showed that as long as $H((1+\gamma_1)/2) < 1-\gamma_2-\lambda$, then $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$ is reducible to $\mathsf{GapCmprEnt}^{\alpha,\beta}_{4N,2R}$ for some constants α,β such that $\alpha-\beta=\lambda$. This reduction (described next for completeness) implies that $\widetilde{\deg}_\varepsilon(\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}) \le \widetilde{\deg}_\varepsilon(\mathsf{GapCmprEnt}^{\alpha,\beta}_{4N,2R})$.

For completeness, we sketch this transformation, closely following the presentation of Goldreich and Vadhan [42]. At a high level, the reduction transforms an input in $[R]^N$ to $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$ (interpreted as a distribution p over $[R]$) into two distributions p_1, p_2 over $[R]\times\{0,1\}$ as follows. Both p_1 and p_2 start by sampling an $s\in\{0,1\}$ at random. If $s=0$, then a random sample r is chosen from p, and if $s=1$, then r is set to a uniform random sample from $[R]$. Distribution p_2 outputs (r,s), while p_1 outputs (r,b) for a random $b\in\{0,1\}$.

The entropy of p_1 is always $v+1$, where $v=\frac{1}{2}H(p)+\frac{1}{2}\log_2(R)$. As for the entropy of p_2, if p is far from the uniform distribution, then the selection bit s will be essentially determined by the sample r. Hence, the entropy of p_2 will be approximately v, which is noticeably smaller than the entropy of p_1. On the other hand, if the two input distributions are close then (even conditioned on the sample selected) the selection bit s will be almost random and so $H(p_2)\approx v+1$, which is approximately the same as $H(p_1)$. Quantitatively, Vadhan [88] shows that if the statistical distance between p and the uniform distribution is δ, then $1-\delta \le H(p_1)-H(p_2) \le H((1+\delta)/2)$.

Since we are considering distributions specified as vectors in $[R]^N$, this transformation can be equivalently described as follows. Assume for simplicity that R divides N. If p is specified by a vector u in $[R]^N$, then p_2 is specified by a vector w in $([R]\times\{0,1\})^{4N}$ defined as follows. For all $i\in[N]$ and $j\in\{0,1\}$, $w_{i,j}=(u_i,0)$, and for $j\in\{2,3\}$, $w_{i,j}=(\lceil Ri/N\rceil,1)$. Similarly, p_1 is specified by a vector v in $([R]\times\{0,1\})^{4N}$. For $i\in[R]$ and $j\in\{0,1\}$, $v_{i,j}=(u_i,j)$, and for $j\in\{2,3\}$, $v_{i,j}=(\lceil Ri/N\rceil, j-2)$. Observe that when representing u, v, and w as vectors in $\{-1,1\}^{N\log_2(R)}$ or $\{-1,1\}^{4N\cdot\log_2(R)}$, each bit of v and w depends on at most one bit of u.

Recall that in the statement of Theorem 6.14, $\gamma_1 = 1-3\gamma/2$ and $\gamma_2 = 1-\gamma^{1/2}/c$, where γ is an arbitrary constant less than $(2/3c)^2$, where $c>1$ is the constant from Theorem 6.2. Clearly, $H((1+\gamma_1)/2)) = H(1-3\gamma/4) = H(3\gamma/4)$. Using the fact that for any $p\in[0,1/2]$,

$$\begin{aligned} H(p) &= -p\log_2(p)-(1-p)\log_2(1-p) \\ &\le -2p\log_2(p) \\ &\le 6p^{3/4}, \end{aligned}$$

it follows that for some constant $\gamma \le 1/(648c^4)$, we have $H(3\gamma/4) < \gamma^{1/2}/c$. Hence, Vadhan's reduction from $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$ to $\mathsf{GapCmprEnt}^{\alpha,\beta}_{4N,2R}$ applies to this setting of γ_1 and γ_2, and this shows that any degree d ε-approximating polynomial for $\mathsf{GapCmprEnt}^{\alpha,\beta}_{4N,2R}$ implies a degree d polynomial ε-approximating polynomial for $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$.

Combined with Theorem 6.14, this implies that $\widetilde{\deg}(\mathsf{GapCmprEnt}^{\alpha,\beta}_{4N,2R}) = \tilde{\Omega}(R^{1/2})$. □

Clearly, a quantum query algorithm that approximates entropy up to additive error $(\alpha-\beta)/4$ can be used to solve $\mathsf{GapCmprEnt}^{\alpha,\beta}_{4N,2R}$, by (approximately) computing the entropies of each of the two input distributions, and determining whether the difference is at most $(\beta+\alpha)/2$. Hence, Theorem 6.15 implies the following lower bound for approximating entropy to additive error $\alpha-\beta$.

Corollary 6.16. *Let $N = c \cdot R$ for a sufficiently large constant c. Interpret an input in $[R]^N$ as a distribution p in the natural way (i. e., for each $j \in [R]$, $p_j = f_j/N$, where f_j is the number of times j appears in the input). There is a constant $\varepsilon > 0$ such that any quantum algorithm that approximates the entropy of p up to additive error ε with probability at least $2/3$ requires $\tilde{\Omega}(R^{1/2})$ queries.*

7 Conclusion and open questions

We conclude by briefly describing some additional consequences of our results, as well as a number of open questions and directions for future work.

7.1 Additional consequences: approximate degree lower bounds for DNFs and AC^0

For any constant $k > 0$, k-distinctness is computed by a DNF of polynomial size. Our $\tilde{\Omega}\left(n^{3/4-1/(2k)}\right)$ is the best known lower bound on the approximate degree of polynomial size DNF formulae. The previous best was $\tilde{\Omega}(n^{2/3})$ for element distinctness (a.k.a., 2-distinctness) [3], although Bun and Thaler did establish, for any $\delta > 0$, an $\Omega(n^{1-\delta})$ lower bound on the approximate degree of *quasi*polynomial size DNFs.

Similarly, for any constant $k \geq 1$, Bun and Thaler exhibited an AC^0 circuit of depth $2k-1$ with approximate degree $\tilde{\Omega}\left(n^{1-2^{k-2}/3^{k-1}}\right)$. Our techniques can be used to give a polynomial improvement for any fixed $k \geq 2$, to $\tilde{\Omega}\left(n^{1-2^{-k}}\right)$ (Theorem 1.1 is the special case of $k = 2$, as SURJ is computed by an AC^0 circuit of depth three). We omit further details of this result for brevity.

7.2 Open problems

The most obvious direction for future work is to extend our techniques to determine the approximate degree and quantum query complexity of additional problems of interest in the study of quantum algorithms. These include triangle finding problem [53, 59], graph collision [59], and verifying matrix products [26, 51]. It would also be interesting to close the gap between our $\Omega(n^{3/4-1/(2k)})$ lower bound for k-distinctness and Belovs' $O\left(n^{3/4-1/(2^{k+2}-4)}\right)$ upper bound, especially for small values of k (e. g., $k = 3$). Note that Belovs' upper bound is known to be tight for k equal to 1 and 2 [3].

Although we prove a lower bound of $\tilde{\Omega}(R^{1/2})$ for $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$ for some constants $0 < \gamma_2 < \gamma_1$, we leave open whether or not $\mathsf{SDU}^{2/3,1/3}_{N,R} = \tilde{\Omega}(R^{1/2})$. It may be tempting to suspect that Theorem 6.14 implies an $\tilde{\Omega}(R^{1/2})$ lower bound on $\mathsf{SDU}^{2/3,1/3}_{N,R}$, by invoking the well-known Polarization Lemma of Sahai and Vadhan [67]. The Polarization Lemma reduces $\mathsf{SDU}^{\gamma_1,\gamma_2}_{N,R}$ for any pair of constant γ_1, γ_2 with $\gamma_2 < \gamma_1^2$ to $\mathsf{SDU}^{2/3,1/3}_{N',R'}$ for an appropriate choice of N' and R'. Unfortunately, N' and R' may be polynomially larger than N and R, so this reduction does not give an $\tilde{\Omega}(R^{1/2})$ lower bound for $\mathsf{SDU}^{2/3,1/3}_{N,R}$ itself.

Another important direction is to determine the approximate degree of specific classes of functions, especially polynomial size DNF formulae, and AC^0 circuits. As mentioned in the previous subsection, our k-distinctness lower bound (Theorem 1.2) gives the best known lower bound on polynomial size DNFs. A compelling candidate for improving this lower bound is the k-sum function, which may have approximate degree as large as $\Theta(n^{k/(k+1)})$ (it is known that the quantum query complexity of k-sum is $\tilde{\Theta}(n^{k/(k+1)})$ [8, 18]). On the upper bounds side, it may be possible to extend the techniques underlying our $\tilde{O}(n^{3/4})$ upper bound on the approximate degree of SURJ to yield a sublinear upper bound for *every* DNF formula of polynomial size.

Open Problem 7.1. For every constant $c > 0$ and every DNF formula $f : \{-1,1\}^n \to \{-1,1\}$ of size at most n^c, is there a $\delta > 0$ (depending only on c) such that $\widetilde{\deg}(f) = O(n^{1-\delta})$?

A positive answer to Open Problem 7.1 would have major algorithmic consequences, including a subexponential time algorithm for agnostically learning DNF formulae [46] (and PAC learning depth three circuits [49]) of any fixed polynomial size.

For general AC^0 circuits, an $\Omega(n^{1-\delta})$ approximate degree lower bound is already known [33]. It would be very interesting to improve this lower bound to an optimal $\Omega(n)$. Until recently, SURJ was a prime candidate for exhibiting such a lower bound. However, owing to Sherstov's upper bound [78] and Theorem 1.1, SURJ is no longer a candidate function. However, we are optimistic about the following closely related candidate. An approximate majority function is any total Boolean function that evaluates to -1 (or $+1$) whenever at least $2/3$ of its inputs are -1 ($+1$, respectively). It is well-known (via the probabilistic method) that there are approximate majorities computable by depth 3 circuits of quadratic size and logarithmic bottom fan-in [4]. It is possible that *every* approximate majority has approximate degree $\Omega(n)$; proving this would resolve a question of Srinivasan [40].

Acknowledgements

We are grateful to Sasha Sherstov for an inspiring conversation at the BIRS 2017 workshop on Communication Complexity and Applications, II, which helped to spark this work. We also thank the anonymous STOC and Theory of Computing reviewers for comments improving the presentation of this manuscript.

This work was done while M. B. was a postdoctoral researcher at Princeton University. Some of this work was performed when R. K. was a postdoctoral associate at MIT and was partly supported by NSF grant CCF-1629809.

References

[1] SCOTT AARONSON: Impossibility of succinct quantum proofs for collision-freeness. *Quantum Inf. Comput.*, 12(1-2):21–28, 2012. [arXiv:1101.0403] 4

[2] SCOTT AARONSON, SHALEV BEN-DAVID, AND ROBIN KOTHARI: Separations in query complexity using cheat sheets. In *Proc. 48th STOC*, pp. 863–876. ACM Press, 2016. [doi:10.1145/2897518.2897644] 5

[3] SCOTT AARONSON AND YAOYUN SHI: Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004. Preliminary version in STOC'02. [doi:10.1145/1008731.1008735] 4, 5, 6, 7, 8, 10, 62

[4] MIKLÓS AJTAI: Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. [doi:10.1016/0168-0072(83)90038-6] 63

[5] ANDRIS AMBAINIS: Quantum lower bounds by quantum arguments. *J. Comput. System Sci.*, 64(4):750–767, 2002. Preliminary version in STOC'00. [doi:10.1006/jcss.2002.1826] 4

[6] ANDRIS AMBAINIS: Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(3):37–46, 2005. [doi:10.4086/toc.2005.v001a003] 18

[7] ANDRIS AMBAINIS: Polynomial degree vs. quantum query complexity. *J. Comput. System Sci.*, 72(2):220–238, 2006. Preliminary version in FOCS'03. [doi:10.1016/j.jcss.2005.06.006] 5

[8] ANDRIS AMBAINIS: Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007. Preliminary version in FOCS'04. [doi:10.1137/S0097539705447311] 7, 63

[9] ANDRIS AMBAINIS, ALEKSANDRS BELOVS, ODED REGEV, AND RONALD DE WOLF: Efficient quantum algorithms for (gapped) group testing and junta testing. In *Proc. 27th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'16)*, pp. 903–922. ACM Press, 2016. [doi:10.1137/1.9781611974331.ch65] 5, 8, 13, 58

[10] ANDRIS AMBAINIS, ANDREW M. CHILDS, BEN REICHARDT, ROBERT ŠPALEK, AND SHENGYU ZHANG: Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer. *SIAM J. Comput.*, 39(6):2513–2530, 2010. Preliminary version in FOCS'07. [doi:10.1137/080712167] 4

[11] ALP ATICI AND ROCCO A. SERVEDIO: Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007. [doi:10.1007/s11128-007-0061-6] 8

[12] HOWARD BARNUM, MICHAEL SAKS, AND MARIO SZEGEDY: Quantum query complexity and semi-definite programming. In *Proc. 18th IEEE Conf. on Computational Complexity (CCC'03)*, pp. 179–193. IEEE Comp. Soc., 2003. [doi:10.1109/CCC.2003.1214419] 4

[13] ROBERT BEALS, HARRY BUHRMAN, RICHARD CLEVE, MICHELE MOSCA, AND RONALD DE WOLF: Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001. Preliminary version in FOCS'98. [doi:10.1145/502090.502097] 4, 5, 53

[14] PAUL BEAME AND WIDAD MACHMOUCHI: The quantum query complexity of AC^0. *Quantum Inf. Comput.*, 12(7–8):670–676, 2012. Link at ACM DL. [arXiv:1008.2422] 5, 7

[15] RICHARD BEIGEL: Perceptrons, PP, and the polynomial hierarchy. *Comput. Complexity*, 4:339–349, 1994. Preliminary version in SCT'92. [doi:10.1007/BF01263422] 4

[16] ALEKSANDRS BELOVS: Learning-graph-based quantum algorithm for k-distinctness. In *Proc. 53rd FOCS*, pp. 207–216. IEEE Comp. Soc., 2012. [doi:10.1109/FOCS.2012.18] 5, 7, 12

[17] ALEKSANDRS BELOVS: Span programs for functions with constant-sized 1-certificates. In *Proc. 44th STOC*, pp. 77–84. ACM Press, 2012. [doi:10.1145/2213977.2213985] 7

[18] ALEKSANDRS BELOVS AND ROBERT ŠPALEK: Adversary lower bound for the k-sum problem. In *Proc. 4th Symp. Innovations in Theoretical Computer Science (ITCS'13)*, pp. 323–328. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2013. [doi:10.1145/2422436.2422474] 63

[19] ERIC BLAIS: Testing juntas nearly optimally. In *Proc. 41st STOC*, pp. 151–158. ACM Press, 2009. [doi:10.1145/1536414.1536437] 8

[20] ANDREJ BOGDANOV, YUVAL ISHAI, EMANUELE VIOLA, AND CHRISTOPHER WILLIAMSON: Bounded indistinguishability and the complexity of recovering secrets. In *Proc. 36th Annual Intern. Cryptology Conf. (CRYPTO'16)*, pp. 593–618, 2016. [doi:10.1007/978-3-662-53015-3_21] 4

[21] ADAM BOULAND, LIJIE CHEN, DHIRAJ HOLDEN, JUSTIN THALER, AND PRASHANT NALINI VASUDEVAN: On the power of statistical zero knowledge. *SIAM J. Comput.*, 49(4):1–58, 2019. Preliminary version in FOCS'17. [doi:10.1137/17M1161749] 4, 9, 10, 55, 56

[22] GILLES BRASSARD, PETER HØYER, AND ALAIN TAPP: Quantum counting. In *Proc. 25th Internat. Colloq. on Automata, Languages and Programming (ICALP'98)*, pp. 820–831. Springer, 1998. [doi:10.1007/BFb0055105] 27

[23] SERGEY BRAVYI, ARAM WETTROTH HARROW, AND AVINATAN HASSIDIM: Quantum algorithms for testing properties of distributions. *IEEE Trans. Inform. Theory*, 57(6):3971–3981, 2011. Preliminary version in STACS'10. [doi:10.1109/TIT.2011.2134250] 5, 8

[24] HARRY BUHRMAN, RICHARD CLEVE, RONALD DE WOLF, AND CHRISTOF ZALKA: Bounds for small-error and zero-error quantum algorithms. In *Proc. 40th FOCS*, pp. 358–368. IEEE Comp. Soc., 1999. [doi:10.1109/SFFCS.1999.814607] 5

[25] HARRY BUHRMAN, ILAN NEWMAN, HEIN RÖHRIG, AND RONALD DE WOLF: Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007. Preliminary version in STACS'05. [doi:10.1007/s00224-006-1313-z] 30, 31

[26] HARRY BUHRMAN AND ROBERT ŠPALEK: Quantum verification of matrix products. In *Proc. 17th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'06)*, pp. 880–889. ACM Press, 2006. Link at ACM DL. 62

[27] HARRY BUHRMAN, NIKOLAI K. VERESHCHAGIN, AND RONALD DE WOLF: On computation and communication with small bias. In *Proc. 22nd IEEE Conf. on Computational Complexity (CCC'07)*, pp. 24–32. IEEE Comp. Soc., 2007. [doi:10.1109/CCC.2007.18] 4

[28] MARK BUN, ROBIN KOTHARI, AND JUSTIN THALER: The polynomial method strikes back: Tight quantum query bounds via dual polynomials. In *Proc. 50th STOC*, pp. 297–310. ACM Press, 2018. [doi:10.1145/3188745.3188784] 1

[29] MARK BUN AND JUSTIN THALER: Dual lower bounds for approximate degree and Markov-Bernstein inequalities. *Inform. Comput.*, 243:2–25, 2015. Preliminary version in ICALP'13. [doi:10.1016/j.ic.2014.12.003] 9, 10, 11, 42, 48

[30] MARK BUN AND JUSTIN THALER: Hardness amplification and the approximate degree of constant-depth circuits. In *Proc. 42nd Internat. Colloq. on Automata, Languages and Programming (ICALP'15)*, pp. 268–280. Springer, 2015. [doi:10.1007/978-3-662-47672-7_22] 9, 13, 48, 55, 56

[31] MARK BUN AND JUSTIN THALER: Dual polynomials for collision and element distinctness. *Theory of Computing*, 12(16):1–34, 2016. [doi:10.4086/toc.2016.v012a016] 14

[32] MARK BUN AND JUSTIN THALER: Improved bounds on the sign-rank of AC^0. In *Proc. 43rd Internat. Colloq. on Automata, Languages and Programming (ICALP'16)*, pp. 37:1–37:14. Schloss Dagstuhl–Leibniz-Zentr. fuer Informatik, 2016. [doi:10.4230/LIPIcs.ICALP.2016.37] 10

[33] MARK BUN AND JUSTIN THALER: A nearly optimal lower bound on the approximate degree of AC^0. In *Proc. 58th FOCS*, pp. 1–12. IEEE Comp. Soc., 2017. [doi:10.1109/FOCS.2017.10] 7, 10, 11, 12, 13, 18, 19, 20, 21, 22, 37, 38, 54, 63

[34] MARK BUN AND JUSTIN THALER: The large-error approximate degree of AC^0. In *Proc. 23rd Internat. Workshop on Randomization and Computation (RANDOM'19)*, pp. 55:1–55:16. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.APPROX-RANDOM.2019.55] 53

[35] KARTHEKEYAN CHANDRASEKARAN, JUSTIN THALER, JONATHAN ULLMAN, AND ANDREW WAN: Faster private release of marginals on small databases. In *Proc. 5th Symp. Innovations in Theoretical Computer Science (ITCS'14)*, pp. 387–402. ACM Press, 2014. [doi:10.1145/2554797.2554833] 4

[36] ARKADEV CHATTOPADHYAY AND ANIL ADA: Multiparty communication complexity of disjointness. *Electron. Colloq. Comput. Complexity*, TR08-002:1–15, 2008. [ECCC] 4, 9

[37] E.W. CHENEY: *Introduction to Approximation Theory*. AMS Chelsea Pub., 1982. 30

[38] MATEI DAVID AND TONIANN PITASSI: Separating NOF communication complexity classes RP and NP. *Electron. Colloq. Comput. Complexity*, TR08-014:1–11, 2008. [ECCC] 4

[39] MATEI DAVID, TONIANN PITASSI, AND EMANUELE VIOLA: Improved separations between nondeterministic and randomized multiparty communication. *ACM Trans. Comput. Theory*, 1(2):5:1–5:20, 2009. Preliminary version in RANDOM'08. [doi:10.1145/1595391.1595392] 4

[40] YUVAL FILMUS, HAMED HATAMI, STEVEN HEILMAN, ELCHANAN MOSSEL, RYAN O'DONNELL, SUSHANT SACHDEVA, ANDREW WAN, AND KARL WIMMER: Real analysis in computer science: A collection of open problems, 2014. Preprint at the Simons Institute. 63

[41] DMITRY GAVINSKY AND ALEXANDER A. SHERSTOV: A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(10):227–245, 2010. [doi:10.4086/toc.2010.v006a010] 4, 57

[42] ODED GOLDREICH AND SALIL P. VADHAN: On the complexity of computational problems regarding distributions (a survey). In ODED GOLDREICH, editor, *Studies in Complexity and Cryptography*, volume 6650 of *LNCS*, pp. 390–405. Springer, 2011. [doi:10.1007/978-3-642-22670-0_27, ECCC:TR11-004] 60, 61

[43] RONALD L. GRAHAM, DONALD E. KNUTH, AND OREN PATASHNIK: *Concrete Mathematics: A Foundation for Computer Science*. Addison–Wesley, 2nd edition, 1994. 39

[44] PETER HØYER, TROY LEE, AND ROBERT ŠPALEK: Negative weights make adversaries stronger. In *Proc. 39th STOC*, pp. 526–535. ACM Press, 2007. [doi:10.1145/1250790.1250867] 4, 5

[45] JEFF KAHN, NATHAN LINIAL, AND ALEX SAMORODNITSKY: Inclusion-exclusion: Exact and approximate. *Combinatorica*, 16(4):465–477, 1996. [doi:10.1007/BF01271266] 4

[46] ADAM TAUMAN KALAI, ADAM R. KLIVANS, YISHAY MANSOUR, AND ROCCO A. SERVEDIO: Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008. Preliminary version in FOCS'05. [doi:10.1137/060649057] 4, 63

[47] VARUN KANADE AND JUSTIN THALER: Distribution-independent reliable learning. In *Proc. 27th Ann. Conf. on Learning Theory (COLT'14)*, pp. 3–24, 2014. MLR Press. [arXiv:1402.5164] 4

[48] HARTMUT KLAUCK, ROBERT ŠPALEK, AND RONALD DE WOLF: Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007. Preliminary version in FOCS'04. [doi:10.1137/05063235X] 5

[49] ADAM R. KLIVANS AND ROCCO A. SERVEDIO: Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. System Sci.*, 68(2):303–318, 2004. Prelim. version in STOC'01. [doi:10.1016/j.jcss.2003.07.007] 4, 63

[50] ADAM R. KLIVANS AND ROCCO A. SERVEDIO: Toward attribute efficient learning of decision lists and parities. *J. Mach. Learn. Res.*, 7:587–602, 2006. JMLR, prelim. version in COLT'04. 4

[51] ROBIN KOTHARI AND ASHWIN NAYAK: Quantum algorithms for matrix multiplication and product verification. In *Encyclopedia of Algorithms*, pp. 1673–1677. Springer, 2016. [doi:10.1007/978-1-4939-2864-4_303] 62

[52] SOPHIE LAPLANTE AND FRÉDÉRIC MAGNIEZ: Lower bounds for randomized and quantum query complexity using Kolmogorov arguments. *SIAM J. Comput.*, 38(1):46–62, 2008. Preliminary version in CCC'04. [doi:10.1137/050639090] 4

[53] FRANÇOIS LE GALL: Improved quantum algorithm for triangle finding via combinatorial arguments. In *Proc. 55th FOCS*, pp. 216–225. IEEE Comp. Soc., 2014. [doi:10.1109/FOCS.2014.31] 62

[54] TROY LEE: A note on the sign degree of formulas, 2009. [arXiv:0909.4607] 11, 19

[55] TROY LEE, RAJAT MITTAL, BEN W. REICHARDT, ROBERT ŠPALEK, AND MARIO SZEGEDY: Quantum query complexity of state conversion. In *Proc. 52nd FOCS*, pp. 344–353. IEEE Comp. Soc., 2011. [doi:10.1109/FOCS.2011.75] 4, 5

[56] TROY LEE AND ADI SHRAIBMAN: An approximation algorithm for approximation rank. In *Proc. 24th IEEE Conf. on Computational Complexity (CCC'09)*, pp. 351–357. IEEE Comp. Soc., 2009. [doi:10.1109/CCC.2009.25] 4

[57] TROY LEE AND ADI SHRAIBMAN: Disjointness is hard in the multiparty number-on-the-forehead model. *Comput. Complexity*, 18(2):309–336, 2009. Preliminary version in CCC'08. [doi:10.1007/s00037-009-0276-2] 9

[58] TONGYANG LI AND XIAODI WU: Quantum query complexity of entropy estimation. *IEEE Trans. Inform. Theory*, 65(5):2899–2921, 2018. [doi:10.1109/TIT.2018.2883306] 5, 7, 9

[59] FRÉDÉRIC MAGNIEZ, MIKLOS SANTHA, AND MARIO SZEGEDY: Quantum algorithms for the triangle problem. *SIAM J. Comput.*, 37(2):413–424, 2007. Preliminary version in SODA'05. [doi:10.1137/050643684] 62

[60] MARVIN MINSKY AND SEYMOUR PAPERT: *Perceptrons – An Introduction to Computational Geometry*. MIT Press, 1969. [doi:10.7551/mitpress/11301.001.0001] 4, 9, 16

[61] ASHLEY MONTANARO: The quantum complexity of approximating the frequency moments. *Quantum Inf. Comput.*, 16(13–14):1169–1190, 2016. Link at ACM DL. [arXiv:1505.00113] 7

[62] NOAM NISAN AND MARIO SZEGEDY: On the degree of Boolean functions as real polynomials. *Comput. Complexity*, 4(4):301–313, 1994. Preliminary version in STOC'92. [doi:10.1007/BF01263419] 15, 26

[63] RYAN O'DONNELL AND ROCCO A. SERVEDIO: New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010. Preliminary version in STOC'03. [doi:10.1007/s00493-010-2173-3] 4, 39

[64] ANUP RAO AND AMIR YEHUDAYOFF: Simplified lower bounds on the multiparty communication complexity of disjointness. In *Proc. 30th IEEE Conf. on Computational Complexity (CCC'15)*, pp. 88–101. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015. [doi:10.4230/LIPIcs.CCC.2015.88] 4

[65] ALEXANDER A. RAZBOROV AND ALEXANDER A. SHERSTOV: The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. Prelim. version FOCS'08. [doi:10.1137/080744037] 11, 22

[66] BEN W. REICHARDT: Reflections for quantum query algorithms. In *Proc. 22nd Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'11)*, pp. 560–569. ACM Press, 2011. [doi:10.1137/1.9781611973082.44] 4, 5

[67] AMIT SAHAI AND SALIL VADHAN: A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003. Preliminary version in FOCS'97. [doi:10.1145/636865.636868] 62

[68] ROCCO A. SERVEDIO, LI-YANG TAN, AND JUSTIN THALER: Attribute-efficient learning and weight–degree tradeoffs for polynomial threshold functions. In *Proc. 25th Ann. Conf. on Learning Theory (COLT'12)*, pp. 14:1–14:19, 2012. MLR Press. 4

[69] ALEXANDER A. SHERSTOV: Communication lower bounds using dual polynomials. *Bull. EATCS*, 95:59–93, 2008. [ECCC:TR08-057] 4

[70] ALEXANDER A. SHERSTOV: Approximate inclusion-exclusion for arbitrary symmetric functions. *Comput. Complexity*, 18(2):219–247, 2009. Preliminary version in CCC'08. [doi:10.1007/s00037-009-0274-4] 4

[71] ALEXANDER A. SHERSTOV: Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in STOC'07. [doi:10.1137/08071421X] 4, 9

[72] ALEXANDER A. SHERSTOV: The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011. Preliminary version in STOC'08. [doi:10.1137/080733644] 4, 9, 15

[73] ALEXANDER A. SHERSTOV: Approximating the AND-OR tree. *Theory of Computing*, 9(20):653–663, 2013. [doi:10.4086/toc.2013.v009a020] 9, 11, 42, 48

[74] ALEXANDER A. SHERSTOV: Communication lower bounds using directional derivatives. In *Proc. 45th STOC*, pp. 921–930. ACM Press, 2013. [doi:10.1145/2488608.2488725] 4

[75] ALEXANDER A. SHERSTOV: The intersection of two halfspaces has high threshold degree. *SIAM J. Comput.*, 42(6):2329–2374, 2013. [doi:10.1137/100785260] 9, 11, 19, 49

[76] ALEXANDER A. SHERSTOV: Making polynomials robust to noise. *Theory of Computing*, 9(18):593–615, 2013. Preliminary version in STOC'12. [doi:10.4086/toc.2013.v009a018] 10

[77] ALEXANDER A. SHERSTOV: The multiparty communication complexity of set disjointness. *SIAM J. Comput.*, 45(4):1450–1489, 2016. Preliminary version in STOC'12. [doi:10.1137/120891587] 4

[78] ALEXANDER A. SHERSTOV: Algorithmic polynomials. In *Proc. 50th STOC*, pp. 311–324. ACM Press, 2018. Journal v. to appear in *SIAM J. Comp.* [doi:10.1145/3188745.3188958] 5, 6, 7, 63

[79] ALEXANDER A. SHERSTOV: Breaking the Minsky–Papert barrier for constant-depth circuits. *SIAM J. Comput.*, 47(5):1809–1857, 2018. Prelim. v. STOC'14. [doi:10.1137/15M1015704] 9, 10, 55

[80] ALEXANDER A. SHERSTOV: The power of asymmetry in constant-depth circuits. *SIAM J. Comput.*, 47(6):2362–2434, 2018. Preliminary version in FOCS'15. [doi:10.1137/16M1064477] 7, 9, 10

[81] YAOYUN SHI AND YUFAN ZHU: Quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5):444–460, 2009. [arXiv:0710.0095] 9, 11, 19

[82] ROBERT ŠPALEK: A dual polynomial for OR, 2008. [arXiv:0803.4516] 9, 38

[83] ROBERT ŠPALEK AND MARIO SZEGEDY: All quantum adversary methods are equivalent. *Theory of Computing*, 2(1):1–18, 2006. Prelim. version ICALP'05. [doi:10.4086/toc.2006.v002a001] 5

[84] AVISHAY TAL: Shrinkage of De Morgan formulae by spectral techniques. In *Proc. 55th FOCS*, pp. 551–560. IEEE Comp. Soc., 2014. [doi:10.1109/FOCS.2014.65] 4

[85] AVISHAY TAL: Formula lower bounds via the quantum method. In *Proc. 49th STOC*, pp. 1256–1268. ACM Press, 2017. [doi:10.1145/3055399.3055472] 4

[86] JUSTIN THALER: Lower bounds for the approximate degree of block-composed functions. In *Proc. 43rd Internat. Colloq. on Automata, Languages and Programming (ICALP'16)*, pp. 17:1–17:15. Schloss Dagstuhl–Leibniz-Zentr. fuer Informatik, 2016. [doi:10.4230/LIPIcs.ICALP.2016.17] 9, 10

[87] JUSTIN THALER, JONATHAN ULLMAN, AND SALIL P. VADHAN: Faster algorithms for privately releasing marginals. In *Proc. 39th Internat. Colloq. on Automata, Languages and Programming (ICALP'12)*, pp. 810–821. Springer, 2012. [doi:10.1007/978-3-642-31594-7_68] 4

[88] SALIL PRAVIN VADHAN: *A Study of Statistical Zero-knowledge Proofs*. Ph. D. thesis, Massachusetts Institute of Technology, 1999. Link at Springer. 13, 61

[89] GREGORY VALIANT AND PAUL VALIANT: Estimating the unseen: An $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. In *Proc. 43rd STOC*, pp. 685–694. ACM Press, 2011. [doi:10.1145/1993636.1993727] 8, 9

[90] EMANUELE VIOLA: Special Topics in Complexity Theory. Lecture notes, Northeastern Univ., 2017. ECCC. 22

[91] MARK ZHANDRY: A note on the quantum collision and set equality problems. *Quantum Inf. Comput.*, 15(7&8):557–567, 2015. Link at ACM DL. [arXiv:1312.1027] 4

[92] SHENGYU ZHANG: On the power of Ambainis lower bounds. *Theoret. Comput. Sci.*, 339(2):241–256, 2005. Preliminary version in ICALP'04. [doi:10.1016/j.tcs.2005.01.019] 4, 5

AUTHORS

Mark Bun
Assistant professor
Department of Computer Science
Boston University
Boston, MA, USA
mbun@bu.edu
http://cs-people.bu.edu/mbun/

Robin Kothari
Senior researcher
Microsoft Quantum and Microsoft Research
Redmond, WA, USA
robin.kothari@microsoft.com
http://www.robinkothari.com

Justin Thaler
Assistant professor
Department of Computer Science
Georgetown University
Washington, DC, USA
justin.thaler@georgetown.edu
http://people.cs.georgetown.edu/jthaler/

ABOUT THE AUTHORS

MARK BUN is an Assistant Professor in the Department of Computer Science at Boston University. He completed his Ph. D. in the Theory of Computation group at Harvard University, where he was advised by Salil Vadhan. After that, he was a postdoctoral researcher at Princeton University and a Google Research Fellow at the Simons Institute. His research interests include computational complexity, differential privacy, cryptozoology, and learning theory. As a native Seattleite, he enjoys coffee, rain, composting, 4-way stop signs, hiking, cheese rolling, and competitive duck herding.

ROBIN KOTHARI received his Ph. D. in Computer Science from the University of Waterloo in 2014, under the supervision of Andrew Childs and John Watrous. He was then a postdoctoral associate at the Center for Theoretical Physics at the Massachusetts Institute of Technology for three years. He is currently a Senior Researcher at Microsoft Quantum and Microsoft Research. His research interests include quantum algorithms, quantum complexity theory, query complexity, and communication complexity.

JUSTIN THALER is an Assistant Professor in the Department of Computer Science at Georgetown University. He completed his Ph. D. in the Theory of Computation group at Harvard University, where he was advised by Michael Mitzenmacher. After that, he was a postdoctoral researcher at the Simons Institute, and a Research Scientist at Yahoo Labs in New York City. His research interests include computational complexity, verifiable computing, and algorithms for massive datasets. In his spare time, he enjoys running and playing low-quality tennis.