





Low-Degree Testing over Grids

Prashanth Amireddy   

School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Srikanth Srinivasan   

Department of Computer Science, Aarhus University, Denmark

Madhu Sudan   

School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

Abstract

We study the question of local testability of low (constant) degree functions from a product domain $\mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ to a field \mathbb{F} , where $\mathcal{S}_i \subseteq \mathbb{F}$ can be arbitrary constant sized sets. We show that this family is locally testable when the grid is “symmetric”. That is, if $\mathcal{S}_i = \mathcal{S}$ for all i , there is a probabilistic algorithm using constantly many queries that distinguishes whether f has a polynomial representation of degree at most d or is $\Omega(1)$ -far from having this property. In contrast, we show that there exist asymmetric grids with $|\mathcal{S}_1| = \cdots = |\mathcal{S}_n| = 3$ for which testing requires $\omega_n(1)$ queries, thereby establishing that even in the context of polynomials, local testing depends on the structure of the domain and not just the distance of the underlying code.

The low-degree testing problem has been studied extensively over the years and a wide variety of tools have been applied to propose and analyze tests. Our work introduces yet another new connection in this rich field, by building low-degree tests out of tests for “junta-degrees”. A function $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \rightarrow \mathcal{G}$, for an abelian group \mathcal{G} is said to be a junta-degree- d function if it is a sum of d -juntas. We derive our low-degree test by giving a new local test for junta-degree- d functions. For the analysis of our tests, we deduce a small-set expansion theorem for spherical/hamming noise over large grids, which may be of independent interest.

2012 ACM Subject Classification Theory of computation \rightarrow Probabilistic computation

Keywords and phrases Property testing, Low-degree testing, Small-set expansion, Local testing

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.41

Category RANDOM

Related Version *Full Version:* <https://arxiv.org/pdf/2305.04983.pdf>

Funding *Prashanth Amireddy:* Supported in part by a Simons Investigator Award and NSF Award CCF 2152413 to Madhu Sudan.

Srikanth Srinivasan: Supported by a start-up package from Aarhus University. Work done while visiting the Meta-Complexity program at the Simons Institute for the Theory of Computing, UC Berkeley.

Madhu Sudan: Supported in part by a Simons Investigator Award and NSF Award CCF 2152413.

1 Introduction

The main problem considered in this paper is “low-degree testing over grids”. Specifically given a degree parameter $d \in \mathbb{Z}^{\geq 0}$ and proximity parameter $\delta > 0$ we would like to design a tester (a randomized oracle algorithm) that is given oracle access to a function $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \rightarrow \mathbb{F}$ where \mathbb{F} is a field and $\mathcal{S}_1, \dots, \mathcal{S}_n \subseteq \mathbb{F}$ are arbitrary finite sets, and accepts if f is a polynomial of degree at most d while rejecting with constant probability (say $1/2$) if f is δ -far (in relative Hamming distance) from every degree d polynomial. The main goal here is to identify settings where the test makes $O(1)$ queries when $d, 1/\delta$ and $\max_{i \in [n]} \{|\mathcal{S}_i|\}$ are all considered constants. (In particular the goal is to get a query complexity independent of n .)



© Prashanth Amireddy, Srikanth Srinivasan, and Madhu Sudan;
licensed under Creative Commons License CC-BY 4.0

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023).

Editors: Nicole Megow and Adam D. Smith; Article No. 41; pp. 41:1–41:22



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Low-degree testing

The low-degree testing problem over grids is a generalization of the classical low-degree testing problem which corresponds to the special case where \mathbb{F} is a finite field and $\mathcal{S}_1 = \cdots = \mathcal{S}_n = \mathbb{F}$. Versions of the classical problem were studied in the early 90s [5, 6, 11] in the context of program checking and (multi-prover) interactive proofs. The problem was formally defined and systematically studied by Rubinfeld and Sudan [24] and played a central role in the PCP theorem [2, 3] and subsequent improvements. While the initial exploration of low-degree testing focused on the case where $d \ll |\mathbb{F}|$ (and tried to get bounds that depended polynomially, or even linearly, on d), a later series of works starting with that of Alon, Kaufman, Krivelevich, Litsyn and Ron [1] initiated the study of low degree testing in the setting where $d > |\mathbb{F}|$. [1] studied the setting of $\mathbb{F} = \mathbb{F}_2$ and this was extended to the setting of other constant sized fields in [17, 19]. An even more recent sequence of works [10, 14, 15, 18] explores so-called “optimal tests” for this setting and these results have led to new applications to the study of the Gowers uniformity norm, proofs of XOR lemmas for polynomials [10], and novel constructions of small set expanders [8].

Part of the reason for the wide applicability of low-degree testing is the fact that evaluations of polynomials form error-correcting codes, a fact that dates back at least to the work of Ore [22]. Ore’s theorem (a.k.a. the Schwartz-Zippel lemma) however applies widely to the evaluations of polynomial on entire “grids”, i.e., sets of the form $\mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ and bounds the distance between low-degree functions in terms of the degree d and minimum set size $\min_i \{|\mathcal{S}_i|\}$. This motivated Bafna, Srinivasan and Sudan [7] to introduce the low-degree testing problem over grids. They proposed and analyzed a low-degree test for the special case of the Boolean grid, i.e., where $|\mathcal{S}_1| = \cdots = |\mathcal{S}_n| = 2$. This setting already captures the setting considered in [1] while also including some novel settings such as testing the Fourier degree of Boolean functions (here the domain is $\{-1, +1\}^n$ while the range is \mathbb{R}). The main theorem in [7] shows that there is a tester with constant query complexity, thus qualitatively reproducing the theorem of [1] (though with a worse query complexity than [1] which was itself worse than the optimal result in [10]), while extending the result to many new settings.

In this work we attempt to go beyond the restriction of a Boolean grid. We discuss our results in more detail shortly, but the main outcome of our exploration is that the problem takes on very different flavors depending on whether the grid is symmetric ($\mathcal{S}_1 = \cdots = \mathcal{S}_n$) or not. In the former case, we get constant complexity testers for constant $|\mathcal{S}_i|$ whereas in the latter setting we show that even when $|\mathcal{S}_i| = 3$ low-degree (even $d = 1$) testing requires superconstant query complexity. (See Theorem 3 for details.) In contrast to previous testers, our tester goes via “junta-degree-tests”, a concept that has been explored in the literature but not as extensively as low-degree tests, and not been connected to low-degree tests in the past. We describe this problem and our results for this problem next.

Junta-degree testing

A function $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \rightarrow \mathcal{G}$ for an arbitrary set \mathcal{G} is said to be a d -junta if it depends only on d of the n variables. When \mathcal{G} is an abelian group, a function $f : \mathcal{S}_1 \times \cdots \times \mathcal{S}_n \rightarrow \mathcal{G}$ is said to be of junta-degree d if it is the sum of d -juntas (where the sum is over \mathcal{G}).¹ In the special case where $|\mathcal{S}_i| = 2$ for all i and \mathcal{G} is a field, junta degree coincides with the

¹ While in principle the problem could also be considered over non-abelian groups, in such a case it not clear if there is a fixed bound on the number of juntas that need to be summed to get to a function of bounded junta-degree.

usual notion of degree. More generally every degree d polynomial has junta degree d , while a function of junta-degree d is a polynomial of degree at most $d \cdot \max_i \{|\mathcal{S}_i|\}$. Thus junta-degree is softly related to algebraic degree and our work provides a step towards low-degree testing via the problem of junta-degree testing.

Junta-degree testing considers the task of testing if a given function has junta-degree at most d or if it is far from all functions of junta-degree at most d . While this problem has not been considered in full generality before, two works do consider this problem for the special case of $d = 1$. Dinur and Golubev [13] considered this problem in the setting where $\mathcal{G} = \mathbb{F}_2$, while Bogdanov and Prakeriya [12] consider this for general abelian groups. This special case corresponds to the problem of testing if a function is a direct sum, thus relating to other interesting classes of properties studied in testing. Both works give $O(1)$ query testers in their settings, but even the case of $d = 2$ remained open.

In our work we give testers for this problem for general constant d in the general asymmetric domain setting with the range being an arbitrary finite group \mathcal{G} , though with the restriction that the maximum set size $|\mathcal{S}_i|$ is bounded. We then use this tester to design our low-degree test over symmetric grids. We turn to our results below. Even though our primary motivation in studying low-junta-degree testing is to ultimately use it for low-degree testing, we note that junta-degree testing even for the case of \mathcal{G} being the additive group of \mathbb{R} (or \mathbb{C}) and $\mathcal{S}_i = \Omega$ (which is some finite set) for all i , is by itself already interesting as in this case, junta-degree corresponds to the “degree” of the Fourier representation of the function (in any basis). Low-Fourier-degree functions and such approximations form a central object in complexity theory and computational learning theory, at least when the domain size is $|\Omega| = 2$. The problem of *learning* low-Fourier-degree functions in particular has received much attention over the years [20, 21], and hence *testing* the same family, over general domains Ω , is an interesting corollary of our results, especially since our techniques are more algebraic than analytic (modulo the usage of a hypercontractivity theorem).

1.1 Our results

We start by stating our theorem for junta-degree testing. (For a formal definition of a tester, see Definition 7).

► **Theorem 1.** *The family of junta-degree- d functions from $\mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ to \mathcal{G} is locally testable with a non-adaptive one-sided tester that makes $O_{s,d}(1)$ queries to the function being tested, where $s = \max_i |\mathcal{S}_i|$.*

In the special case where $|\mathcal{S}_i| = s$ for all i , the tester makes $s^{O(s^2d)}$ queries.

In particular, if we treat all the parameters above except n as constant, this gives a test that succeeds with high probability by making only a constant number of queries. Taking $(\mathcal{G}, +) = (\mathbb{R}, +)$ or $(\mathbb{C}, +)$, the above theorem results in a local tester for Fourier-degree:

► **Corollary 2.** *The family of functions $f : \Omega^n \rightarrow \mathbb{R}$ of Fourier-degree at most d is locally testable in $s^{O(s^2d)} = O_{s,d}(1)$ queries, where $s = |\Omega|$.²*

We now turn to the question of testing whether a given function $f : \mathcal{S}^n \rightarrow \mathbb{F}$ is *degree- d* , i.e., whether there is a polynomial of degree at most d agreeing with f , or δ -far from it. Here \mathcal{S} can be any arbitrary finite subset of the field. Note that being junta-degree- d is a necessary condition for f being degree- d . Combining the above JUNTA-DEG with an additional test (called WEAK-DEG), we can test low-degree functions over a field, or rather over *any subset* of a field.

² The same result also holds if the co-domain is \mathbb{C} instead of \mathbb{R} .

► **Theorem 3.** *For any subset $S \subseteq \mathbb{F}$ of size s , the family of degree- d functions from S^n to \mathbb{F} is locally testable with a non-adaptive, one-sided tester that makes $(sd)^{O(s^3d)} = O_{s,d}(1)$ queries to the function being tested.*

The special case of $S = \mathbb{F} = \mathbb{F}_q$ (finite field of size q) is especially interesting. Although this was already established for general finite fields first by Kaufman and Ron [19] and an optimal query complexity (in terms of d , for constant prime q) was achieved by Haramaty, Shpilka and Sudan [15], we nevertheless present it as a corollary of Theorem 3.

► **Corollary 4** (Kaufman and Ron [19]). *The family of degree- d functions $f : \mathbb{F}_q^n \rightarrow \mathbb{F}_q$ is locally testable in $(qd)^{O(q^3d)} = O_{q,d}(1)$ queries.*

Turning our attention to more general product domains, we show that while junta-degree testing is still locally testable over there more general grids, testing degree in constantly many queries, even for $d = 1$, is intractable for all sufficiently large fields \mathbb{F} .

► **Theorem 5.** *For a growing parameter n , there exists a field \mathbb{F} and its subsets $\mathcal{S}_1, \dots, \mathcal{S}_n$ of constant size (i.e., 3) such that testing the family of degree-1 functions $f : \mathcal{S}_1 \times \dots \times \mathcal{S}_n \rightarrow \mathbb{F}$ requires $\Omega(\log n)$ queries.*

► **Remark 6.** A recent work of Arora, Bhattacharyya, Fleming, Kelman and Yoshida [4] considers low-degree testing over the reals and tests whether a given $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is degree- d or ε -far with respect to a distribution \mathcal{D} . They give a test with query complexity independent of n for their problem ([4, Theorem 1.1]). This seems to contradict our result which seems to include the special case of their setting for $\mathcal{D} = \text{Unif}(\mathcal{S}_1 \times \dots \times \mathcal{S}_n)$ and $\mathbb{F} = \mathbb{R}$, where Theorem 5 shows that a dependence on n is necessary. The seeming contradiction is resolved by noting that the models in our paper and that of [4] are quite different. In particular, while in our setting the function f can only be queried on the support of the distribution \mathcal{D} (namely $\mathcal{S}_1 \times \dots \times \mathcal{S}_n$), in [4] the function can be queried at any point in \mathbb{R}^n and the distribution \mathcal{D} only shows up when defining the distance between two functions. (So in their model a function f that happens to agree with a degree d polynomial on the support of \mathcal{D} but disagrees outside the support may be rejected with positive probability, while in our model such a function must be accepted with probability one.)

1.2 Technical contributions

All low-degree tests roughly follow the following pattern: Given a function f on n variables x_1, \dots, x_n they select some $k = O_d(1)$ new variables $y = (y_1, \dots, y_k)$ and substitute $x_i = \sigma_i(y)$, where σ_i 's are simple random functions, to get an $O(1)$ -variate function $g(y) = f(\sigma(y))$; and then verify g is a low-degree polynomial in k variables by brute force. When the domain is \mathbb{F}_q^n for some field \mathbb{F}_q , σ_i 's can be chosen to be an affine form in y – this preserves the domain and ensures degree of g is at most the degree of f , thus at least ensuring completeness. While soundness of the test was complex to analyze, a key ingredient in the analysis is that for any pair of points $a \neq b \in \mathbb{F}_q^k$, $\sigma(a)$ and $\sigma(b)$ are uniform independent elements of \mathbb{F}_q^n (over the randomness of σ). At least in the case where f is roughly $1/q^k$ distance from the degree d family, this ensures that with constant probability g will differ from a degree d polynomial in exactly one point making the test reject. Dealing with cases where f is much further away is the more complex part that we won't get into here.

When the domain is not \mathbb{F}_q^n affine substitutions no longer preserve the domain and so we can't use them in our tests. In the cases of the domain being $\{-1, +1\}^n$, [7] used much simpler affine substitutions of the form $x_i = c_i y_{j(i)}$ where $c_i \in \{-1, +1\}$ uniformly and independently over i and $j(i) \in \{1, \dots, k\}$ uniformly though not independently over i . Then [7] iteratively

reduce the number of variables as follows: When only r variables x_1, \dots, x_r remain, they pick two uniformly random indices $i \neq j \in \{1, \dots, r-1\}$ and identify x_j with x_i , and then rename the $r-1$ remaining variables as x_1, \dots, x_{r-1} . At the end when $r = k$, they pick a random bijection between x_1, \dots, x_k and y_1, \dots, y_k . This iterative identification eventually maps every variable x_i to some variable $y_{j(i)}$. The nice feature of this identification scheme is it leads to a sequence of functions f_n, f_{n-1}, \dots, f_k with f_r being a function of r variables on the same domain, and of degree at most d if $f = f_n$ has degree at most d . If however we start with f_n being *very* far from degree d polynomials, there must exist r such that f_r is very far from high-degree functions while f_{r-1} is only *moderately* far. The probability of a bad event can be bounded (via some algebraic arguments) by $O(d^2/r^2)$. This step is the key to this argument and depends on the fact that f_{r-1} involves very small changes to f_r . Summing over r then gives the constant probability that the final function f_k (or equivalently g) is far from degree d polynomials. This still leaves [7] with the problem of dealing with functions f that are close to codewords: Here they use the fact that this substitution ensures that $\sigma(a)$ is distributed uniformly in $\{-1, +1\}^n$ for every $a \in \{-1, +1\}^k$. It is however no longer true that $\sigma(b)$ is uniform conditioned on $\sigma(a)$ for $b \neq a$, but it is still the case that if b is moderately far in Hamming distance from a then $\sigma(b)$ has sufficient entropy conditioned on $\sigma(a)$. (Specifically $\sigma(b)$ is distributed uniformly on a sphere of distance $\Omega(n)$ from $\sigma(a)$.) This entropy, combined with appropriate small-set expansion bounds on the Boolean hypercube, and in particular a spherical hypercontractivity result due to Polyanskiy [23]), ensures that if f is somewhat close to a low-degree polynomial then g is far from every degree d polynomial on an appropriately chosen subset of $\{-1, +1\}^k$ and so the test rejects.

To extend this algorithm and analysis to the setting on non-Boolean domains we are faced with two challenges: (1) We cannot afford to negate variables (using the random variables $c(i)$ above) when the domain is not $\{-1, +1\}$ – we can only work with identification of variables (or something similar). (2) The increase in the domain size forces us to seek a general spherical hypercontractivity result on non-Boolean alphabets and this is not readily available. Overcoming either one of the restrictions on its own seems plausible, but doing it together (while also ensuring that the sequence of restrictions/identifications do not make the distance to the family being tested to abruptly drop in distance as we go from f_n, f_{n-1}, \dots to f_k) turns out to be challenging and this is where we find it critical to go via junta-degree testing.

As a first step in our proof we extend the approach of [7] to junta-degree testing over the domain \mathcal{S}^n for arbitrary finite \mathcal{S} . (It is relatively simple to extend this further to the case of $\mathcal{S}_1 \times \dots \times \mathcal{S}_n$ – we don't discuss that here.) This is achieved by using substitutions of the form $x_i = \pi_i(y_{j(i)})$ where $\pi_i : \mathcal{S} \rightarrow \mathcal{S}$ is a random bijection. While this might increase the degree of the function, this preserves the junta-degree (or reduces it) and makes it suitable for analysis of the junta-degree test, which we now describe: Following the template of a low-degree test stated at the beginning of this subsection, the junta-tester would simply check whether $g(y) = f(\sigma(y))$ is of junta-degree at most d where σ is the random function induced by the identifications $j(\cdot)$ and permutations π_i of variables. The permutations π_i here serve the same purpose as the coefficients c_i 's do in the substitutions $x_i = c_i y_{j(i)}$ of [7] which is to ensure that for any $a \in \mathcal{S}^k$, $\sigma(a)$ is uniformly distributed in \mathcal{S}^n . With this idea in place extending the analysis of [7] to our setting ends up with a feasible path, except we had to address a few more differences; one such challenge is that in the analysis the rejection probability of junta-degree test on functions that are close to being junta-degree- d , we will need to analyze the effect of a spherical noise operator on grids (i.e., a subset of coordinates of fixed size is chosen uniformly at random and each coordinate in that subset is changed to

a different value uniformly at random). While [23] shows that such a noise operator has the desired hypercontractivity behavior, and the corresponding small-set expansion theorem was used in the test of [7], this was only for a Boolean alphabet. In this paper, when the alphabet size $s = |\mathcal{S}|$ is more than 2, by doing Fourier analysis over \mathbb{Z}_s^n , we are able to relate it to the more standard Bernoulli i.i.d. noise operator for which we do have a small-set expansion theorem available – we believe this can be of independent interest.³

The other differences of our junta-degree test analysis compared to that of [7] are mainly to account for the fact that we are aiming for junta-degree testing over any (abelian) group whereas the low-degree testing ideas of [7] and other prior work utilize the properties of polynomials over fields. We also give a cleaner proof as compared to [7] for the fact that the sequence of functions of fewer and fewer variables obtained by the random identifications (along with permutations) does not abruptly decrease in distance to the junta-degree- d family like we pointed out earlier (see “large-distance lemma” Lemma 16).

We then return to the task of low-degree testing: For this we design a new test: We first test the given oracle for junta-degree d , then if it passes, we pick a fresh random identification scheme setting $x_i = y_{j(i)}$ for uniform independent $j(i) \in \{1, \dots, k\}$ and verify (by brute-force) that the resulting k variate function has degree at most d . The advantage with this two stage tester is that in the second stage the given function is already known to be close to a polynomial of degree at most sd where $s = \max_i \{|\mathcal{S}_i|\}$. This makes the testing problem closer to a polynomial identity testing problem, though the problem takes some care to define, and many careful details to be worked out in the analysis. A particular challenge arises from the fact that the first phase only proves that our function is only *close* to a low-degree polynomial and may not be low-degree exactly – so in the second stage we have to be careful to sample the function on essentially uniform inputs. This prevents us from using all of \mathcal{S}^k when looking at the restricted function $g(y)$, but only allows us to use balanced inputs in \mathcal{S}^k (where a balanced input has an equal number of coordinates with each value $v \in \mathcal{S}$). In turn understanding what the lowest degree of function can be given its values on a balanced set leads to new algebraic questions. Section 4 gives a full proof of the low-degree test and analysis spelling out the many technical questions and our solutions to those.

The final testing-related result we prove is an impossibility result, showing that while low-degree functions are locally testable over \mathcal{S}^n , this cannot be extended to general grids $\mathcal{S}_1 \times \dots \times \mathcal{S}_n$ for large enough fields (Theorem 5). From a coding theory perspective, this reveals that local testability of even polynomial evaluations codes requires more structure than simply having a large distance. To sketch the idea, let $d = 1$ and \mathbb{F} be any field of size at least $n + 2$ with distinct elements $\{0, 1, a_1, \dots, a_n\}$. For $i \in [n]$, let $\mathcal{S}_i = \{0, 1, a_i\}$. We will refer to 0, 1 as Boolean elements and the remaining as non-Boolean ones. Let $\zeta(b) = b$ if b is Boolean and \star otherwise. As degree-1 functions over $\mathcal{S}_1 \times \dots \times \mathcal{S}_n$ form a linear subspace over \mathbb{F} , by a result due to Ben-Sasson, Harsha and Raskhodnikova [9] any test can be converted to a one-sided, non-adaptive one without changing the number of queries or the error by more than a factor of 2. Thus, we may assume that the test (call it TEST) is of the following form: TEST samples a matrix $M \in \mathbb{F}^{\ell \times n}$ according to a distribution \mathcal{D} with rows $x^{(1)}, \dots, x^{(\ell)} \in \mathcal{S}_1 \times \dots \times \mathcal{S}_n$ and accepts f if and only if $P(f(x^{(1)}), \dots, f(x^{(\ell)}))$ is true where P is some fixed predicate. We will show that if TEST accepts degree-1 functions with probability 1 and rejects $\Omega(1)$ -far functions with probability $\Omega(1)$, then $\ell = \Omega(\log n)$. By the one-sidedness, we must have for all $M \in \text{sup}(\mathcal{D})$ that if $f_M := (f(x^{(1)}), \dots, f(x^{(\ell)})) \in \text{colspace}(M)$, then TEST accepts, where sup denotes the support and colspace denotes the column space. Picking $i \in [n]$ uniformly at random, note that the function $g(x) := x_i(x_i - 1)$ is $\Omega(1)$ -far from degree-1.

³ We note that the hypercontractivity setting we are considering and analyzing in this part is not sufficient to get a direct analysis of low-degree testing. Such an analysis would require hypercontractivity for more delicate noise models than the simpler “ q -ary symmetric” models we analyze here.

Then we argue that the “evaluation vector” $g_M := (g(x^{(1)}), \dots, g(x^{(\ell)}))$ lies in the column space of M with high probability (over i) if $\ell = o(\log n)$, so TEST faultily accepts g . The idea is that if $\ell = o(\log n)$ then there are at least two columns (say $i \neq j \in [n]$) of M that are identical under the ζ mapping, so subtracting one from the other gives a vector in \mathbb{F}^ℓ that is (up to a constant factor) equal to g_M . This is because for $k \in [\ell]$, the k -th coordinate of the difference vector is $a_i - a_j$ (some constant) if $\zeta(x_i^{(k)}) = \star$ and 0 otherwise. Similarly the k -th coordinate of g_M is $a_i(a_i - 1)$ (some constant) if $\zeta(x_i^{(k)}) = \star$ and 0 otherwise.

2 Preliminaries

We denote $[n] = \{1, \dots, n\} \subseteq \mathbb{Z}$, $[m..n] = [n] \setminus [m - 1]$ and $\mathbb{Z}_s = \mathbb{Z}/s\mathbb{Z} = \{0, 1, \dots, s - 1\}$ for $s \geq 2$. Throughout the paper, let $(\mathcal{G}, +)$ be an arbitrary abelian group and $(\mathbb{F}, +, \cdot)$ an arbitrary field. \mathbb{F}_q^n is a vector space over the finite field of q elements, to which we associate an inner product (bilinear form) as: $\langle x, y \rangle = \sum_{i=1}^n x_i \cdot y_i$.

For any finite set \mathcal{S} and $a \in \mathcal{S}^n$ we denote the Hamming weight of a by $\#a = \{i \in [n] : a_i \neq 0\}$, assuming \mathcal{S} contains an element called 0. If $I \subseteq [n]$, we use a^I to denote the tuple a restricted to the coordinates of I , i.e., $a^I = (a_i)_{i \in I}$. Similarly $\mathcal{S}^I = \{a^I : a \in \mathcal{S}^n\}$. For disjoint subsets $I, J \subseteq [n]$, and $a \in \mathcal{S}^I$ and $b \in \mathcal{S}^J$, we denote their concatenation by $a \circ b \in \mathcal{S}^{I \cup J}$. Denoting a product domain/grid by $\bar{\mathcal{S}} = \mathcal{S}_1 \times \dots \times \mathcal{S}_n$, we let $\bar{\mathcal{S}}^I = \times_{i \in I} \mathcal{S}_i$ denote the Cartesian product of sets restricted to the coordinates of I .

We use $\binom{[n]}{\leq d}$ to denote the set of subsets of $[n]$ of size at most d . For m a multiple of s , let “balanced set” $\mathcal{B}(\mathcal{S}, m) \subseteq \mathcal{S}^m$ be the set of points that contain exactly m/s many repetitions of each element of \mathcal{S} . Abusing notation, sometimes we may think of $\mathcal{B}(\mathcal{S}, m)^{m'}$ as a subset of $\mathcal{S}^{mm'}$ by flattening the tuple of m -tuples. The *group-integer multiplication* operation $\cdot : \mathcal{G} \times \mathbb{Z} \rightarrow \mathcal{G}$ is defined by $g \cdot m = g + \dots + g$ ($|m|$ times) if $m \geq 0$ and $-g - \dots - g$ ($|m|$ times) otherwise.

2.1 Local testability

The *distance* between $f : \bar{\mathcal{S}} \rightarrow \mathcal{G}$ and a family of functions \mathcal{F} with the same domain $\bar{\mathcal{S}}$ is $\delta_{\mathcal{F}}(f) = \min_{g \in \mathcal{F}} \delta(f, g)$, where $\delta(f, g) = \Pr_{x \sim \bar{\mathcal{S}}} [f(x) \neq g(x)]$. We say that f is δ -far from \mathcal{F} if $\delta_{\mathcal{F}}(f) \geq \delta$. When \mathcal{F} is the family of junta-degree- d functions, we denote $\delta_{\mathcal{F}}(\cdot)$ by simply $\delta_d(\cdot)$. Similarly f is δ -close to \mathcal{F} if $\delta_{\mathcal{F}}(f) \leq \delta$.

► **Definition 7** (Local testability). *A randomized algorithm \mathcal{A} with an oracle access to a function $f : \bar{\mathcal{S}} \rightarrow \mathcal{G}$ as its input, is said to be q -local if it performs at most q queries for any given f . For a family of functions \mathcal{F} with domain $\bar{\mathcal{S}}$ and co-domain \mathcal{G} , we say that \mathcal{F} is q -locally testable for $q = q(\mathcal{F})$ if there exists a q -local test \mathcal{A} that accepts f with probability 1 if $f \in \mathcal{F}$, and rejects f with probability at least $\delta_{\mathcal{F}}(f)/2$ if $f \notin \mathcal{F}$. Further if $q(\mathcal{F}) = O(1)$ (i.e., independent of the number of variables n), we simply refer to \mathcal{F} as being locally testable.*

One can define more general two-sided error and adaptive tests, but in the context of this paper, the above definition for local testability is without loss of generality as we know from the work of Ben-Sasson, Harsha and Raskhodnikova [9] that for *linear properties*⁴, any “test” can be transformed to be one-sided and non-adaptive without altering the query complexity (locality) and success probability by more than constant factors.

⁴ i.e., for families \mathcal{F} for which $f \in \mathcal{F}$ and $g \in \mathcal{F}$ implies $c_1 f + c_2 g \in \mathcal{F}$ for all $c_1, c_2 \in \mathbb{F}$.

The family of functions \mathcal{F} of our study (namely “junta-degree- d ” and “degree- d ” to be formally defined shortly) are parameterized by $s = \max_i |\mathcal{S}_i|$ and an integer d which we treat as constants. All tests we are going to present are $O_{s,d}(1)$ -local, one-sided and non-adaptive. However, the probability of rejection in case of $f \notin \mathcal{F}$ is only $\Omega_{s,d}(\delta_{\mathcal{F}}(f))$; nevertheless by repeating the test an appropriate $O_{s,d}(1)$ number of times, we get a $O_{s,d}(1)$ -local test for \mathcal{F} that succeeds with probability $\delta_{\mathcal{F}}(f)/2$ when $f \notin \mathcal{F}$ and with probability 1 when $f \in \mathcal{F}$.

2.2 Junta-polynomials and polynomials

For this section, we let $\bar{\mathcal{S}} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ denote an arbitrary finite product domain (or grid) and $s = \max_i \{s_i\}$, where $s_i = |\mathcal{S}_i|$.

► **Definition 8** (Junta-degree). *A function $f : \bar{\mathcal{S}} \rightarrow \mathcal{G}$ ⁵ is said to be junta-degree- d if*

$$f(x) = f_1(x^{D_1}) + \cdots + f_t(x^{D_t})$$

for some $t \in \mathbb{Z}$, $D_j \in \binom{[n]}{\leq d}$ and functions $f_j : \bar{\mathcal{S}}^{D_j} \rightarrow \mathcal{G}$ for $j \leq t$. If $t = 1$, we call f a d -junta.

The junta-degree of f is the minimum $d \geq 0$ such that f is junta-degree- d .

For junta-degree testing over arbitrary grids $\bar{\mathcal{S}} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$, we may assume that $\mathcal{S}_i = \mathbb{Z}_{s_i}$ without loss of generality, where $s_i = |\mathcal{S}_i|$. The following claims about junta-polynomials are analogous to standard facts about multi-variate polynomials over a field.

► **Claim 9.** Any junta-degree- d function $f : \mathbb{Z}_s^n \rightarrow \mathcal{G}$ can be uniquely⁶ expressed as

$$f(x_1, \dots, x_n) = \sum_{\substack{a \in \mathbb{Z}_s^n \\ \#a \leq d}} g_a \cdot \prod_{i \in [n]: a_i \neq 0} \delta_{a_i}(x_i), \quad (1)$$

where $g_a \in \mathcal{G}$ and $\delta_b : \mathbb{Z}_s \rightarrow \mathbb{Z}$ is defined as $\delta_b(y) = 1$ if $b = y$ and 0 otherwise.

► **Definition 10** (Junta-polynomial). *We will call such a representation as a junta-polynomial, and the degree of a junta-polynomial is defined as $\max_{a \in \mathbb{Z}_s^n: g_a \neq 0} \#a$. It can be seen that the degree of a junta-polynomial is exactly equal to the junta-degree of the function it computes, assuming that the degree of the identically 0 junta-polynomial is 0.*

We will refer to the summands in (1) as terms, the constants g_a as coefficients, the integer products $\prod_{i \in [n]: a_i \neq 0} \delta_{a_i}(x_i)$ as monomials. We say that a is a root of a junta-polynomial P if $P(a) = 0$ and a is a non-root otherwise.

► **Claim 11.** Any non-zero junta-polynomial $P : \mathbb{Z}_s^n \rightarrow \mathcal{G}$ of degree at most d has at least s^{n-d} non-roots.

We will now discuss standard facts about formal polynomials. Let \mathbb{F} be a field and $\mathcal{S} \subseteq \mathbb{F}$ be of size $s \geq 2$. For a polynomial $P(x_1, \dots, x_n) \in \mathbb{F}[x_1, \dots, x_n]$ the individual degree of x_i is the largest degree x_i takes in any (non-zero) monomial of P . The individual degree of P is the largest individual degree of any variable x_i . We say that P is degree- d if its degree is at most d . We say that $f : \mathcal{S}^n \rightarrow \mathbb{F}$ is degree- d iff there is a degree- d polynomial $P \in \mathbb{F}[x_1, \dots, x_n]$ computing f . For the analysis of our degree-tester, we also need a notion of degree- d for non-product domains: for any $\mathcal{T} \subseteq \mathcal{S}^n$, we say that $f : \mathcal{T} \rightarrow \mathbb{F}$ is degree- d if there is a degree- d polynomial $P \in \mathbb{F}[x_1, \dots, x_n]$ computing f .

⁵ Here we treat a tuple of sets as the domain of the function

⁶ up to the commutativity of the Σ (group addition) and Π (integer multiplication) operations

▷ **Claim 12.** Any degree- d function $f : \mathcal{S}^n \rightarrow \mathbb{F}$ has a unique polynomial representation with degree at most d and individual degree at most $s - 1$.

By setting $d = n(s - 1)$ (or ∞) in the above claim, we see that the set of all functions from \mathcal{S}^n to \mathbb{F} is a vector space over \mathbb{F} of dimension s^n – the monomials with individual degree at most $s - 1$ form a basis. More generally, for any $\mathcal{T} \subseteq \mathcal{S}^n$ the set of functions from \mathcal{T} to \mathbb{F} forms a vector space of dimension $|\mathcal{T}|$ with an inner product defined for $f, g : \mathcal{T} \rightarrow \mathbb{F}$ as $\langle f, g \rangle = \sum_{x \in \mathcal{T}} f(x) \cdot g(x)$. For any d , the set of degree- d functions is a subspace of this vector space.

It is easy to see that if $f : \mathcal{S}^n \rightarrow \mathbb{F}$ is degree- d , then it is also junta-degree- d (w.r.t. to the additive group of \mathbb{F}). Conversely, if $f : \mathcal{S}^n \rightarrow \mathbb{F}$ is junta-degree- d , then it is degree- $(s - 1)d$: this follows by applying Claim 12 to the d -junta components of f . If $s = 2$, the degree is exactly equal to the junta-degree.

Let $\delta'_d(f)$ denote the distance of f to the degree- d family.

2.3 Fourier analysis

► **Definition 13** (Fourier representation). Any function $f : \mathbb{Z}_s^n \rightarrow \mathbb{C}$ can be uniquely expressed as

$$f(x) = \sum_{\alpha \in \mathbb{Z}_s^n} \widehat{f}(\alpha) \chi_\alpha(x) \tag{2}$$

where the characters are defined as $\chi_\alpha(x) = \prod_{i \in [n]} \chi_{\alpha_i}(x_i)$ where $\chi_\beta(y) = \omega^{\beta y \bmod s}$ for $\beta, y \in \mathbb{Z}_s$ and $\omega \in \mathbb{C}$ is a (fixed) primitive s -th root of unity.

► **Definition 14** (Noise operators). For $\nu \in [0, 1]$ and $x \in \mathbb{Z}_s^n$, we define $N_\nu(x)$ ⁷ to be the distribution over \mathbb{Z}_s^n where each coordinate of x is unchanged with probability $1 - \nu$, and changes to a different value uniformly at random with probability ν . Similarly, the spherical noise corresponds to $S_\nu(x)$ where a subset $J \subseteq [n]$ of fixed size νn is chosen uniformly at random and the coordinates outside J are unchanged and those within J are changed to a uniformly different value. Let \mathcal{D}_ν denote the probability distribution over \mathbb{Z}_s with mass $1 - \nu$ at 0 and $\nu/(s - 1)$ at all the other points. Let \mathcal{E}_ν denote the uniform distribution over $\{y \in \mathbb{Z}_s : \#y = \nu n\}$. For $\mu_1 \sim \mathcal{D}_\nu^{\otimes n}$ and $\mu_2 \sim \mathcal{E}_\nu$, note that $N_\nu(x)$ and $x + \mu_1$ are identically distributed; so are $S_\nu(x)$ and $x + \mu_2$.

3 Low-junta-degree testing

We note that junta-degree- d functions with domain $\mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ such that $|\mathcal{S}_i| = s$ for all i are “equivalent” to those with domain \mathbb{Z}_s^n as one can fix an arbitrary ordering of elements in each \mathcal{S}_i and treat the function as being over \mathbb{Z}_s^n : this does not change the junta-degree. Hence, we will fix $\mathcal{S}_i = \mathbb{Z}_s$. The more general case of unequal domain sizes is handled in the full version of the paper.

We claim that the following test works to check if a given function $f : \mathbb{Z}_s^n \rightarrow \mathcal{G}$ is junta-degree- d .

⁷ This is different from the standard usage N_ρ where ρ denotes the probability of “retention” and not of noise.

The junta-degree test (JUNTA-DEG)

For a parameter $k = O_{s,d}(1)$ that is yet to be fixed, the junta-degree test (which we shall refer to as JUNTA-DEG) for $f : \mathbb{Z}_s^n \rightarrow \mathcal{G}$ is the following algorithm with $I = [n]$, $r = n$ and $f_r = f$:

- Test $T_{I,r}(f_r)$:** gets query access to $f_r : \mathbb{Z}_s^I \rightarrow \mathcal{G}$ where $I \subseteq [n]$ is of size r .
1. If $r \leq k$, accept iff f_r is junta-degree- d (check this by querying f_r at all points in its domain). Otherwise,
 2. Choose $i \neq j \in I$ and a permutation $\pi_j : \mathbb{Z}_s \rightarrow \mathbb{Z}_s$ independently and uniformly at random. Let $I' = I \setminus \{j\}$.
 3. Apply the test $T_{I',r-1}(f_{r-1})$ where $f_{r-1} : \mathbb{Z}_s^{I'} \rightarrow \mathcal{G}$ is the function obtained by setting $x_j = \pi_j(x_i)$ in f_r : that is, $f_{r-1}(a^{I'}) := f_r(a^{I'} \circ (\pi_j(a_i))^{\{j\}})$ for $a \in \mathcal{S}^{I'}$.

The query complexity of the JUNTA-DEG test is $s^k = O_{s,d}(1)$ regardless of the randomness within the test. Furthermore, if the function f happens to be a junta-degree- d function, then the test JUNTA-DEG always accepts it, since permuting variables and substituting some variables with other variables does not change the junta-degree, so Step 1 succeeds. In this section, we will show that if $\delta := \delta_d(f) > 0$, then $\Pr[\text{JUNTA-DEG rejects } f] \geq \varepsilon\delta$ for appropriate $\varepsilon = \Omega_{s,d}(1)$.

We follow the same approach as [7] (which itself follows [10]) and argue that if $\delta_d(f)$ is “small”, then we will be able to prove $\Pr[\text{JUNTA-DEG rejects } f] \geq \varepsilon \cdot \delta_d(f)$ and if not, at least we will be able to find some $r \in [k+1..n]$ such that $\delta_d(f_r)$ is small enough (but importantly, not too small). Then, we apply the small-distance analysis to that f_r .

We state here the two main lemmas to prove that the correctness of the junta-degree tester. Here, the parameters $\varepsilon_0 \leq \varepsilon_1$ and ε will be chosen to be at least $s^{-O(k)}$. In the context of the test $T_{I,r}(f_r)$ described above, we will set $k = \psi s^2 d$ for a sufficiently large but constant ψ to be fixed in the proofs of the below lemmas⁸.

► **Lemma 15** (Small-distance lemma). *For any $I \subseteq [n]$ of size $r > k$, if $\delta = \delta_d(f_r) \leq \varepsilon_1$, then*

$$\Pr[T_{I,r} \text{ rejects } f_r] \geq \varepsilon\delta.$$

► **Lemma 16** (Large-distance lemma). *For any $I \subseteq [n]$ of size $r > k$, if $\delta_d(f_r) > \varepsilon_1$, then*

$$\Pr_{i,j,\pi_j} [\delta_d(f_{r-1}) \leq \varepsilon_0] \leq k^2/2r(r-1).$$

Assuming the above two lemmas to be true, the proof of Theorem 1, at least for symmetric domains $\mathcal{S}_1 \times \cdots \times \mathcal{S}_n = \mathbb{Z}_s^n$, follows the same approach as in [7] and we omit it here. We also defer the proof of the large-distance lemma to the full version (and Appendix B). The case of junta-degree testing over *general* grids can be reduced to that of symmetric grids and we refer the reader to the full version for details.

3.1 Small-distance lemma

Proof of Lemma 15. We will “unroll” the recursion of the JUNTA-DEG test and state it more directly as follows: Fix an arbitrary $r > k$. As r is fixed, we denote f_r by f (not to be confused with the initial function on n variables). For the proof, we will need the following alternate description of $T_{I,r}$ (subsequently, we shall drop the subscript I). Here, $\sigma : [r] \rightarrow [k]$ is a map chosen according to the following random process:

⁸ For $d = 0$, we can take $k = \psi s^2$ so that it is non-zero.

- For $i = 1$ to k , set $\sigma(i) = i$.
- For $i = k + 1$ to r , set $\sigma(i) = j$ with probability $|\{i' < i : \sigma(i') = j\}| / (i - 1)$, for each $j \in [k]$.

The only property we need about the above distribution of σ is that it is “well-spread”, which was already shown in [7] as the following lemma.

► **Lemma 17** (Corollary 6.9 in [7]). *With probability at least $1/2^{O(k)}$, we have $|\sigma^{-1}(j)| \geq r/4k$ for all $j \in [k]$ – we call such a σ good.*

Test $T_r(f)$: gets query access to $f : \mathcal{S}^{[r]} \rightarrow \mathcal{G}$ with variables x_1, \dots, x_r .

1. Choose a tuple of permutations of \mathbb{Z}_s , $\pi = (\pi_1, \dots, \pi_r)$ u.a.r.
2. Choose a bijection $\mu : [r] \rightarrow [r]$ u.a.r.
3. Choose a map $\sigma : [r] \rightarrow [k]$ according to the distribution described above Lemma 17.
4. For $y = (y_1, \dots, y_k) \in \mathbb{Z}_s^k$, define $x_{\pi\sigma\mu}(y) = (\pi_1(y_{\sigma(\mu^{-1}(1))}), \dots, \pi_r(y_{\sigma(\mu^{-1}(r))}))$.
5. Accept iff $f'(y) := f(x_{\pi\sigma\mu}(y))$ is junta-degree- d .

Let $\delta = \delta_d(f_r) = \delta(f, P) \leq \varepsilon_1$ where $P : \mathbb{Z}_s^{[r]} \rightarrow \mathcal{G}$ is junta-degree- d and $E \subseteq \mathcal{S}^r$ be the points where f and P differ. Our objective is to show that

$$\Pr_{\pi, \sigma, \mu} [T_r \text{ rejects } f] = \Pr_{\pi, \sigma, \mu} [f' \text{ is not junta-degree-}d] \geq \varepsilon\delta. \quad (3)$$

Let the functions $f' : \mathcal{S}^k \rightarrow \mathcal{G}$ and $P' : \mathcal{S}^k \rightarrow \mathcal{G}$ be defined by $f'(y) = f(x_{\pi\sigma\mu}(y))$ and $P'(y) = P(x_{\pi\sigma\mu}(y))$ respectively (these functions depend on π, σ, μ) and $E' \subseteq \mathcal{S}^k$ be the points where these two restricted functions differ.

To proceed further, we will need a subset U of \mathbb{Z}_s^k with the following properties (we defer the proof to Appendix A and the full version):

▷ **Claim 18.** Let $w = \lceil \log(8\psi s^2)d \rceil < k$. There exists a set $U \subseteq \mathbb{Z}_s^k$ of size 2^w such that

1. (Code) For all $y \neq y' \in U$,

$$k/4 \leq \Delta(y, y') \leq 3k/4$$

where $\Delta(y, y')$ denotes the number of coordinates where y and y' differ.

2. (Hitting) No two junta-degree- d functions $P : \mathcal{S}^k \rightarrow \mathcal{G}$ and $Q : \mathcal{S}^k \rightarrow \mathcal{G}$ can differ at exactly one point in U .

Let $V = \{x_{\pi\sigma\mu}(y) : y \in U\} \subseteq \mathbb{Z}_s^r$. Because the mapping $y \mapsto x_{\pi\sigma\mu}(y)$ is one-one conditioned on σ being good, it holds that $|V \cap E| = |U \cap E'|$ under this conditioning. Now suppose the randomness is such that $|U \cap E'| = 1$. Then, since no two junta-degree- d functions can disagree at exactly one point in U (Property 2 of Claim 18), it must be the case that f' be of junta-degree greater than d (as P' , being a restriction of a junta-degree- d function is already junta-degree- d). Therefore, for (3) we can set $\varepsilon := \Pr_{\sigma}[\sigma \text{ good}] \geq 1/2^{O(k)}$ and show

$$\Pr_{\pi, \sigma, \mu} [|U \cap E'| = 1 \mid \sigma \text{ good}] = \Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [|U \cap E'| = 1] \geq \delta.$$

By a simple inclusion-exclusion, the above probability is

$$\Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [|V \cap E| = 1] \geq \sum_{y \in U} \Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [x_{\pi\sigma\mu}(y) \in E] - \sum_{y \neq y' \in U} \Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [x_{\pi\sigma\mu}(y) \in E \text{ and } x_{\pi\sigma\mu}(y') \in E] \quad (4)$$

41:12 Low-Degree Testing over Grids

For any $y \in U$, $x_{\pi\sigma\mu}(y) = (\pi_1(y_{\sigma(\mu^{-1}(1))}), \dots, \pi_r(y_{\sigma(\mu^{-1}(r))}))$ is uniformly distributed over \mathbb{Z}_s^r since π_1, \dots, π_r are random permutations of \mathbb{Z}_s . Hence the first part of (4) is

$$\sum_{y \in U} \Pr[x_{\pi\sigma\mu}(y) \in E] = |U| \cdot \frac{|E|}{s^r} = |U| \cdot \delta. \quad (5)$$

For any fixed $y \neq y' \in U$ and good σ , using Property 1 of Claim 18 for points in U we claim that the random variables $x := x_{\pi\sigma\mu}(y)$ and $x' := x_{\pi\sigma\mu}(y')$ are related as follows:

▷ **Claim 19.** $x' \sim S_\nu(x)$, for some $\nu \in [1/32, 31/32]$.

For the second term of (4),

$$\begin{aligned} \Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [x_{\pi\sigma\mu}(y) \in E \text{ and } x_{\pi\sigma\mu}(y') \in E] &= \Pr_{\substack{x \sim \mathbb{Z}_s^r \\ x' \sim S_\nu(x)}} [x \in E \text{ and } x' \in E] \\ &\text{(for some } \nu \in [1/32, 31/32] \text{ depending on } \sigma, \text{ using Claim 19)} \\ &\leq C \cdot \delta^{1+\lambda} \quad \text{for some constant } C \text{ and } \lambda = 1/2^{14} \log s. \end{aligned} \quad (6)$$

(Using spherical noise small-set expansion (Theorem 23))

Plugging the bounds (5) and (6) back in (4), we get

$$\Pr_{\substack{\pi, \mu, \\ \sigma \text{ good}}} [|V \cap E| = 1] \geq |U| \delta - |U|^2 C \delta^{1+\lambda} \geq |U| \delta / 2 \geq \delta.$$

The above inequalities follow from $|U| = 2^w$ and $\delta \leq \varepsilon_1$; this is where we set $\varepsilon_1 := (1/2C|U|)^{1/\lambda} = (1/2C2^w)^{2^{14} \log s} \geq 1/s^{O(\log(8\psi s^2)d)} \geq 1/s^{O(k)}$. Hence we conclude that

$$\Pr_{\pi, \sigma, \mu} [T_r \text{ rejects } f] \geq \Pr_{\sigma} [\sigma \text{ good}] \cdot \Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [|U \cap E'| = 1] \geq \varepsilon \Pr_{\substack{\pi, \mu \\ \sigma \text{ good}}} [|V \cap E| = 1] \geq \varepsilon \delta. \quad \blacktriangleleft$$

4 Low-degree testing

We will describe our low-degree test now.

The degree test (DEG)

Given query access to $f : \mathcal{S}^n \rightarrow \mathbb{F}$, the following test (called DEG) works to test whether f is degree- d . We may assume that $s = |\mathcal{S}| \geq 2$ as f is a constant function otherwise.

Test DEG(f): gets query access to $f : \mathcal{S}^n \rightarrow \mathbb{F}$.

- Run JUNTA-DEG(f) to check if f is junta-degree- d .
- Run WEAK-DEG(f).
- Accept iff both the above tests accept.

In the above description, the sub-routine WEAK-DEG corresponds to the following test.

Test WEAK-DEG(f): gets query access to $f : \mathcal{S}^n \rightarrow \mathbb{F}$.

- Choose a map $\mu : [n] \rightarrow [K]$ u.a.r. where $K = t(d+1)$ and $t = s^3$.
- For $y = (y_1, \dots, y_K) \in \mathcal{S}^K$, define $x_\mu(y) = (y_{\mu(1)}, \dots, y_{\mu(n)})$.
- Define the function $f' : \mathcal{B}(\mathcal{S}, t)^{K/t} \rightarrow \mathbb{F}$ as $f'(y) = f(x_\mu(y))$, where $\mathcal{B}(\mathcal{S}, t)$, defined in Section 2, is the “balanced” subset of \mathcal{S}^t .
- Accept iff f' is degree- d .

We now move on to the analysis of this test, proving Theorem 3.

Proof of Theorem 3. Let $g : \mathcal{S}^n \rightarrow \mathbb{F}$ be a closest junta-degree- d function to f i.e., $\delta_d(f) = \delta(f, g)$. We shall assume that $\delta_d(f) \leq \varepsilon_2$ for a small enough $\varepsilon_2 = K^{-O(K)}$, and neither f nor g are degree- d . Otherwise, we will be able to appeal to the correctness of JUNTA-DEG.

Let $E \subseteq \mathcal{S}^n$ be the points where f and g differ; thus $|E|/s^n = \delta_d(f) \leq \varepsilon_2$. Let

$$V_\mu = \{x_\mu(y) : y \in \mathcal{B}(\mathcal{S}, t)^{K/t}\}.$$

Suppose $\mu : [n] \rightarrow [K]$ is such that $V_\mu \cap E = \emptyset$ and WEAK-DEG rejects g . Then WEAK-DEG does not distinguish between f and g and hence rejects f as well. We will show that both these events occur with good probability. For the first probability, we will upper bound

$$\begin{aligned} \Pr_\mu [V_\mu \cap E \neq \emptyset] &= \Pr_\mu [\exists y \in \mathcal{B}(\mathcal{S}, t)^{K/t} : x_\mu(y) \in E] \\ &\leq \left| \mathcal{B}(\mathcal{S}, t)^{K/t} \right| \cdot \Pr_\mu [\text{For fixed arbitrary } y \in \mathcal{B}(\mathcal{S}, t)^{K/t}, x_\mu(y) \in E]. \end{aligned}$$

Note that since all points in $\mathcal{B}(\mathcal{S}, t)^{K/t}$ contain an equal number of occurrences of all the elements of \mathcal{S} , $x_\mu(y)$ is uniformly distributed in \mathcal{S}^n for a uniformly random μ . Hence, the above probability is

$$\Pr_\mu [V_\mu \cap E \neq \emptyset] \leq |\mathcal{S}|^K \cdot \frac{|E|}{s^n} \leq s^K \varepsilon_2 < \frac{1}{2K^d}. \quad (\text{by setting } \varepsilon_2 := 1/4s^K K^d \geq K^{-O(K)})$$

We show that WEAK-DEG indeed rejects g with good probability.

▷ **Claim 20.** $\Pr_\mu[\text{WEAK-DEG rejects } g] \geq 1/K^d$.

Assuming this claim,

$$\begin{aligned} \Pr[\text{DEG rejects } f] &\geq \Pr[\text{WEAK-DEG rejects } f] \\ &\geq \Pr[\text{WEAK-DEG rejects } g] - \Pr[V_\mu \cap E \neq \emptyset] \geq \frac{1}{2K^d} \geq \frac{\delta'_d(f)}{2K^d}. \end{aligned}$$

This finishes the analysis of the low-degree test assuming Claim 20. ◀

4.1 Soundness of WEAK-DEG

Proof of Claim 20. We will need the following lemma about the vector space formed by functions over $\mathcal{B}(\mathcal{S}, t)^{K/t} \subseteq \mathcal{S}^K$.

► **Lemma 21.** For $\mathcal{T} \subseteq \mathcal{S}^K$, the vector space of functions from \mathcal{T} to \mathbb{F} has a basis $\{m_1, \dots, m_\ell\}$ such that for any $f : \mathcal{T} \rightarrow \mathbb{F}$ of the form $f = c_1 m_1 + \dots + c_\ell m_\ell$ for some $c_i \in \mathbb{F}$, we have

$$f \text{ is degree-}d \iff \forall i, c_i = 0 \text{ or } m_i \text{ is degree-}d. \quad (7)$$

41:14 Low-Degree Testing over Grids

Let $g' : \mathcal{B}(\mathcal{S}, t)^{K/t} \rightarrow \mathbb{F}$ be defined as $g'(y) = g(x_\mu(y))$ where $x_\mu(y) = (y_{\mu(1)}, \dots, y_{\mu(n)})$. Then, recall that WEAK-DEG rejects g iff g' is not degree- d . We use Lemma 21 above to fix a suitable basis m_1, \dots, m_ℓ for functions from $\mathcal{T} = \mathcal{B}(\mathcal{S}, t)^{K/t}$ to \mathbb{F} . Then we can write each $g'(y)$ obtained above uniquely as

$$g'(y) = \sum_{i=1}^{\ell} c_i \cdot m_i$$

where the coefficients c_i are some functions of μ . We will treat $\mu : [n] \rightarrow [K]$ as a random element of $[K]^n$.

We argue that each c_i is junta-degree- d (as a function of μ). To see this, we recall that g is junta-degree- d . Consider the case when g is a function of only x_{i_1}, \dots, x_{i_s} for some $s \leq d$. In this case, clearly the polynomial g' depends only on $\mu_{i_1}, \dots, \mu_{i_s}$. In particular, each c_i is just an s -junta. Extending the argument by linearity, we see that for any g that is junta-degree- d , the underlying coefficients c_i of $g'(y)$ are junta-degree- d polynomials in the co-ordinates of μ .

Now assume that there exists a $\mu^* : [n] \rightarrow [K]$ such that $g'(y)$ is not degree- d (we will show the existence of such a “good” μ in the next subsection). Thus, by Lemma 21 there exists $i^* \in [\ell]$ such that m_{i^*} is not degree- d and $c_{i^*}(\mu^*) \neq 0$. In particular, the function c_{i^*} is non-zero.

We have argued that there is an m_{i^*} in the basis such that the associated coefficient c_{i^*} is a non-zero junta-degree- d polynomial. In particular, Claim 11 implies that the probability that $c_{i^*}(\mu) \neq 0$ for a random μ is at least $1/K^d$. Therefore, using Lemma 21,

$$\Pr_{\mu} [\text{WEAK-DEG rejects } g] = \Pr_{\mu} [g' \text{ is not degree-}d] \geq \Pr_{\mu} [c_{i^*}(\mu) \neq 0] \geq 1/K^d. \quad \blacktriangleleft$$

4.2 Existence of a good map μ

We will show for any function $g : \mathcal{S}^n \rightarrow \mathbb{F}$ that is not degree- d , there exists a map $\mu : [n] \rightarrow [K]$ such that the function $g'(y) = g(x_\mu(y))$ defined for $y \in \mathcal{B}(\mathcal{S}, t)^{K/t}$ is also not degree- d . This is easy to prove if the domain of g' were to be \mathcal{S}^K , but is particularly tricky in our setting.

Let $D = d + 1$. We will give a map $\mu : [n] \rightarrow [t] \times [D] \equiv [tD] = [K]$ instead. Let P be the polynomial with individual degree at most $s - 1$ representing g ; suppose the degree of P is $d' > d$ and let $m(x) = c \cdot x_{i_1}^{a_1} \cdots x_{i_\ell}^{a_\ell}$ be a monomial of $P(x)$ of degree d' for some non-zero $c \in \mathbb{F}$, where $a_j \geq 1$ for all j and i_1, \dots, i_ℓ are some distinct elements of $[n]$ and $\ell \leq d$ as g is junta-degree- d . Then, we define μ as follows for $i \in [n]$:

$$\mu(i) = \begin{cases} (1, j), & \text{if } i = i_j \text{ for some } j \in [\ell] \\ (1, D), & \text{otherwise.} \end{cases}$$

It is easy to inspect that $P(x_\mu(y))$ (call it $Q(y)$) is a polynomial in variables $y_{(1,1)}, \dots, y_{(1,D)}$, and is of degree $d' > d$ – this is because the monomial $m(x)$, upon this substitution turns to

$$m(x_\mu(y)) = c \cdot y_{(1,1)}^{a_1} \cdots y_{(1,\ell)}^{a_\ell},$$

which cannot be cancelled by $m'(x_\mu(y))$ for any other monomial $m'(x)$ of $P(x)$, as if m' contains the variable x_i for some $i \notin \{i_1, \dots, i_\ell\}$ then $m'(x_\mu(y))$ contains the variable $y_{(1,D)}$ and on the other hand if m' only contains variables x_i for some $i \in \{i_1, \dots, i_\ell\}$, then the individual degree of $y_{(1,j)}$ in the two substitutions differs for some j . Hence, the degree of $Q(y)$ is $a_1 + \dots + a_\ell = d'$. As we can express the function $y_{(1,D)}^a$ for $a > s - 1$ as a polynomial

in $y_{1,D}$ of individual degree at most $s - 1$, we can further transform $Q(y)$ so that it has individual degree at most $s - 1$, while maintaining the properties that it still only contains the variables $y_{(1,1)}, \dots, y_{(1,D)}$ (i.e., the first “row”) and has degree d' and computes the function $g'(y)$. The following claim then completes the proof of the existence of a good μ by setting $w = D$ and $d' = d'$.

▷ **Claim 22.** For formal variables $y \equiv (y_1, \dots, y_w) \equiv (y_{(i,j)})_{(i,j) \in [t] \times [w]}$, let $Q(y)$ be a polynomial of degree $d' \geq 0$ containing only the variables from the first row. Then the degree of $Q(y)$ as a function over $\mathcal{B}(\mathcal{S}, t)^w$ is exactly d' .

The proof of the above claim is deferred to Appendix C.

5 Small-set expansion for spherical noise

In this section, we will prove a small-set expansion theorem for spherical noise, which we have used for (6) in the proof of the small-distance lemma of junta-degree testing. Let $f : \mathbb{Z}_s^n \rightarrow \mathbb{C}$ be arbitrary. For $\nu \in [0, 1]$ the Bernoulli noise operator is defined as $N_\nu f(x) = \mathbb{E}_{y \sim N_\nu(x)} [f(y)] = \mathbb{E}_{y \sim \mathcal{D}_\nu^n} [f(x + y)]$. Similarly, the spherical noise operator is defined for $\nu \in [0, 1]$ such that $\nu n \in \mathbb{Z}$: $S_\nu f(x) = \mathbb{E}_{y \sim S_\nu(x)} [f(y)] = \mathbb{E}_{y \sim \mathcal{E}_\nu} [f(x + y)]$. For the rest of this section, let $\rho \in [0, 1]$ and $\nu = (1 - 1/s)(1 - \rho) \in [0, 1]$; it is easy to check that if each coordinate of x is retained with probability ρ and randomized (uniformly over \mathbb{Z}_s) with probability $1 - \rho$, the resulting string is distributed according to $N_\nu(x)$.

The goal of this section is to show for $s \geq 3$, that we can reduce the problem of small-set expansion for spherical noise to that of Bernoulli noise, for which such a theorem is already known.

► **Theorem 23** (Small-set expansion for spherical noise). *Let $s \geq 3$ and $A \subseteq \mathbb{Z}_s^n$ be such that $\Pr_{x \sim \mathbb{Z}_s^n} [x \in A] = \delta$. Then, for any $\nu \in [1/32, 1]$*

$$\Pr_{\substack{x \sim \mathbb{Z}_s^n \\ y \sim S_\nu(x)}} [x \in A \text{ and } y \in A] \leq 2 \cdot \delta^{1+\lambda}, \tag{8}$$

where $\lambda = \frac{1}{2^{14} \log s}$.

► **Remark.** When the size of the domain s is equal to 2 and $\nu \in [1/32, 31/32]$, the above statement still holds (with the factor 2 replaced with some other constant factor C) as proved by [23] (or Corollary 2.8 in [7]).

Proof of Theorem 23. Let $f : \mathbb{Z}_s^n \rightarrow \mathbb{C}$ be the indicator function of A and consider its Fourier representation as in Definition 13: $f(x) = \sum_{\alpha \in \mathbb{Z}_s^n} \widehat{f}(\alpha) \chi_\alpha(x)$. Then the probability in (8) is equal to

$$\Pr_{\substack{x \sim \mathbb{Z}_s^n \\ y \sim S_\nu(x)}} [x \in A \text{ and } y \in A] = \sum_{\alpha \in \mathbb{Z}_s^n} |\widehat{f}(\alpha)|^2 \mathbb{E}_{y \sim \mathcal{E}_\nu} [\chi_\alpha(y)]. \tag{9}$$

We will show that for any $\alpha \in \mathbb{Z}_s^n$, the quantity $\mathbb{E}_{y \sim \mathcal{E}_\nu} [\chi_\alpha(y)]$ above is upper bounded by $2 \cdot \tilde{\rho}^{|\alpha|}$ for some constant $\tilde{\rho}$. We have

$$\mathbb{E}_{y \sim \mathcal{E}_\nu} [\chi_\alpha(y)] = \mathbb{E}_{y \sim \mathcal{E}_\nu} [\chi_{\alpha_1}(y_1) \cdots \chi_{\alpha_n}(y_n)] = \mathbb{E}_{\substack{I \sim \binom{[n]}{\nu n} \\ \mu \sim \mathbb{Z}_s \setminus \{0\} \\ y \sim 0^I \circ \mu^I}} \left[\prod_{i \in I} \chi_{\alpha_i}(y_i) \prod_{i \notin I} \chi_{\alpha_i}(y_i) \right] \tag{where I and μ are independent}$$

$$\begin{aligned}
 &= \mathbb{E}_{I, \mu, y} \left[\prod_{i \in I} \chi_{\alpha_i}(\mu_i) \right] && \text{(where } I \sim \binom{[n]}{\nu n}, \mu \sim \mathbb{Z}_s \setminus \{0\}, \text{ and } y \sim 0^{\bar{I}} \circ \mu^{\bar{I}}) \\
 &= \mathbb{E}_I \left[\prod_{i \in I} \mathbb{E}_{\mu_i \sim \mathbb{Z}_s \setminus \{0\}} [\chi_{\alpha_i}(\mu_i)] \right]. && (10)
 \end{aligned}$$

Now we note that the inner term

$$\mathbb{E}_{\mu_i \sim \mathbb{Z}_s \setminus \{0\}} [\chi_{\alpha_i}(\mu_i)] = \begin{cases} 1, & \text{if } \alpha_i = 0, \text{ and} \\ \frac{1}{s-1} \left(\sum_{\mu_i \in \mathbb{Z}_s \setminus \{0\}} \chi_{\alpha_i}(\mu_i) \right) = \frac{1}{s-1} (s \mathbb{E}_{\mu_i \sim \mathbb{Z}_s} [\chi_{\alpha_i}(\mu_i)] - 1) = \frac{-1}{s-1} & \text{otherwise.} \end{cases} \quad (11)$$

Therefore, denoting the coordinates of α with non-zero entries by $J \subseteq [n]$, plugging (11) into (10) gives

$$\begin{aligned}
 \mathbb{E}_{y \sim \mathcal{E}_\nu} [\chi_\alpha(y)] &= \mathbb{E}_{I \sim \binom{[n]}{\nu n}} \left[\left(\frac{-1}{s-1} \right)^{|J \cap I|} \right] \leq \mathbb{E}_{I \sim \binom{[n]}{\nu n}} \left[\left(\frac{1}{2} \right)^{|J \cap I|} \right] && \text{(as } s \geq 3) \\
 &\leq \Pr_{I \sim \binom{[n]}{\nu n}} [|J \cap I| < \nu k/2] \cdot 1 + \mathbb{E}_{I \sim \binom{[n]}{\nu n}} \left[\left(\frac{1}{2} \right)^{|J \cap I|} \mid |J \cap I| \geq \nu k/2 \right].
 \end{aligned}$$

Denoting $|J| = \#\alpha$ by k , we observe that $|J \cap I|$ is distributed according to the hypergeometric distribution of k draws (without replacement) from a population of size n and νn many success states. Hence, by a tail bound [16] $\Pr[|J \cap I| < \nu k/2] \leq e^{-\nu^2 k/2}$ and using $\nu \geq 1/32$, we get that $\mathbb{E}_{y \sim \mathcal{E}_\nu} [\chi_\alpha(y)] \leq \Pr_{I \sim \binom{[n]}{\nu n}} [|J \cap I| < \nu k/2] \cdot 1 + \mathbb{E}_{I \sim \binom{[n]}{\nu n}} \left[\left(\frac{1}{2} \right)^{|J \cap I|} \mid |J \cap I| \geq \nu k/2 \right] \leq 2 \cdot \tilde{\rho}^k$ for $\tilde{\rho} := 2^{-2^{-11}}$.

Plugging the above bound in (9), letting $\tilde{\nu} = (1 - 1/s)(1 - \tilde{\rho})$ and $q = 2 + \varepsilon = 2 + \frac{1}{2^{12} \log s}$, we get

$$\Pr_{\substack{x \sim \mathbb{Z}_s^n \\ y \sim S_\nu(x)}} [x \in A \text{ and } y \in A] \leq 2 \sum_{\alpha \in \mathbb{Z}_s^n} \left| \hat{f}(\alpha) \right|^2 \tilde{\rho}^{\#\alpha} = 2 \Pr_{\substack{x \sim \mathbb{Z}_s^n \\ y \sim N_{\tilde{\nu}}(x)}} [x \in A \text{ and } y \in A] \leq 2\delta^{2-2/q},$$

where the last step uses the small-set expansion theorem corresponding to Bernoulli noise (e.g. Theorem 10.25 in [21]). \blacktriangleleft

References

- 1 Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing reed-muller codes. *IEEE Trans. Inf. Theory*, 51(11):4032–4039, 2005. doi:10.1109/TIT.2005.856958.
- 2 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. doi:10.1145/278298.278306.
- 3 Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998. doi:10.1145/273865.273901.
- 4 Vipul Arora, Arnab Bhattacharyya, Noah Fleming, Esty Kelman, and Yuichi Yoshida. Low degree testing over the reals. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22–25, 2023*, pages 738–792. SIAM, 2023. doi:10.1137/1.9781611977554.ch31.
- 5 László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In Cris Koutsougeras and Jeffrey Scott Vitter, editors, *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5–8, 1991, New Orleans, Louisiana, USA*, pages 21–31. ACM, 1991. doi:10.1145/103418.103428.

- 6 László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. *Comput. Complex.*, 1:3–40, 1991. doi:10.1007/BF01200056.
- 7 Mitali Bafna, Srikanth Srinivasan, and Madhu Sudan. Local decoding and testing of polynomials over grids. *Random Struct. Algorithms*, 57(3):658–694, 2020. doi:10.1002/rsa.20933.
- 8 Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter. *SIAM J. Comput.*, 44(5):1287–1324, 2015. doi:10.1137/130929394.
- 9 Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. Some 3cnf properties are hard to test. *SIAM J. Comput.*, 35(1):1–21, 2005. doi:10.1137/S0097539704445445.
- 10 Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal testing of reed-muller codes. In *51th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2010, October 23-26, 2010, Las Vegas, Nevada, USA*, pages 488–497. IEEE Computer Society, 2010. doi:10.1109/FOCS.2010.54.
- 11 Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993. doi:10.1016/0022-0000(93)90044-W.
- 12 Andrej Bogdanov and Gautam Prakriya. Direct sum and partitionability testing over general groups. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 33:1–33:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.33.
- 13 Irit Dinur and Konstantin Golubev. Direct sum testing: The general case. In Dimitris Achlioptas and László A. Végh, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2019, September 20-22, 2019, Massachusetts Institute of Technology, Cambridge, MA, USA*, volume 145 of *LIPIcs*, pages 40:1–40:11. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.40.
- 14 Elad Haramaty, Noga Ron-Zewi, and Madhu Sudan. Absolutely sound testing of lifted codes. *Theory Comput.*, 11:299–338, 2015. doi:10.4086/toc.2015.v011a012.
- 15 Elad Haramaty, Amir Shpilka, and Madhu Sudan. Optimal testing of multivariate polynomials over small prime fields. *SIAM J. Comput.*, 42(2):536–562, 2013. doi:10.1137/120879257.
- 16 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.
- 17 Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. Testing low-degree polynomials over prime fields. *Random Struct. Algorithms*, 35(2):163–193, 2009. doi:10.1002/rsa.20262.
- 18 Tali Kaufman and Dor Minzer. Improved optimal testing results from global hypercontractivity. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 98–109. IEEE, 2022. doi:10.1109/FOCS54457.2022.00017.
- 19 Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM J. Comput.*, 36(3):779–802, 2006. doi:10.1137/S0097539704445615.
- 20 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. doi:10.1145/174130.174138.
- 21 Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. URL: <http://www.cambridge.org/de/academic/subjects/computer-science/algorithmics-complexity-computer-algebra-and-computational-g/analysis-boolean-functions>.
- 22 Øystein Ore. Über höhere kongruenzen. *Norsk Mat. Forenings Skrifter*, 1(7):15, 1922.
- 23 Yury Polyanskiy. Hypercontractivity of spherical averages in hamming space. *SIAM J. Discret. Math.*, 33(2):731–754, 2019. doi:10.1137/15M1046575.
- 24 Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. doi:10.1137/S0097539793255151.

A Existence of the code U

In order to prove Claim 18 we will need the following:

▷ **Claim 24.** There exists a matrix $M \in \mathbb{F}_2^{k \times w}$ such that $U := \{Mz : z \in \mathbb{F}_2^w\} \subseteq \mathbb{F}_2^k$ is of size 2^w and:

- For all $y \neq y' \in U$, we have $k/4 \leq \Delta(y, y') \leq 3k/4$.
- There exists a function $\chi : U \rightarrow \{\pm 1\}$ such that for all $I \in \binom{[k]}{\leq d}$, we have

$$\sum_{y \in U: y^I = 1^I} \chi(y) = 0.$$

Proof. We will show that picking M uniformly at random satisfies both the items with positive probability. For Item 1, it suffices if for all $y \neq 0^k$ in U , $k/4 \leq \#y \leq 3k/4$; that is, for all $z \in \mathbb{F}_2^w \setminus \{0^{d+1}\}$, $k/4 \leq \#Mz \leq 3k/4$. For any fixed $z \neq 0^a$, we note that $y = Mz$ is uniformly distributed over \mathbb{F}_2^k as $M \sim \mathbb{F}_2^{k \times w}$. Hence, by a Chernoff bound, we have $\Pr_M [\#Mz \notin [k/4, 3k/4]] \leq 2e^{-k/24}$. A union bound over all $z \in \mathbb{F}_2^w \setminus \{0^w\}$ gives

$$\Pr_M [\neg(\forall y \neq y' \in U, k/4 \leq \Delta(y, y') \leq 3k/4)] \leq 2^w \cdot 2e^{-k/24} < 1/2.$$

However, it is a known fact that a uniformly chosen rectangular matrix has full rank with probability at least $1/2$. Therefore, with positive probability there must be a matrix M such that it is full rank and Item 1 holds. We fix such an M and prove Item 2.

For $y \in U$, as M is full rank there exists a unique $z \in \mathbb{F}_2^w$ such that $Mz = y$. Then we define

$$\chi(y) := (-1)^{\langle z, \eta \rangle} = (-1)^{z_1 \eta_1 + \dots + z_w \eta_w},$$

where $\eta \in \mathbb{F}_2^w$ is an arbitrary vector such that it is not in the \mathbb{F}_2^w -span of any d rows of M . Such an η always exists as the number of vectors that *can* be expressed as a linear combination of d rows of M is at most

$$\binom{k}{d} 2^d \leq \left(\frac{ek}{d}\right)^d 2^d = (2e\psi s^2)^d < 2^w,$$

the total number of vectors in \mathbb{F}_2^w .

Let $M_1, \dots, M_k \in \mathbb{F}_2^w$ denote the rows of M . For any $I \in \binom{[k]}{\leq d}$ and $y = Mz$, the condition $y^I = 1^I$ is equivalent to: $\langle z, M_i \rangle = 1$ for all $i \in I$. Hence we have

$$\sum_{y \in U: y^I = 1^I} \chi(y) = \sum_{z \in \mathbb{F}_2^w: \forall i \in I, \langle z, M_i \rangle = 1} (-1)^{\langle z, \eta \rangle} \tag{12}$$

As η is linearly independent with $\{M_i\}_{i \in I}$, there exists $\eta' \neq 0^a$ such that $\langle \eta', M_i \rangle = 0$ for all $i \in I$ and $\langle \eta', \eta \rangle = 1$: this is because we can treat these conditions as a system of linear equations over \mathbb{F}_2 .

Note that for any $z \in \mathbb{F}_2^w$, $\langle z, M_i \rangle = 1$ if and only if $\langle z + \eta', M_i \rangle = \langle z, M_i \rangle + \langle \eta', M_i \rangle = 1$. Since $z \neq z + \eta'$, we may partition the summation (12) into buckets of size 2, each bucket corresponding to z and $z + \eta'$ for some z . For each such bucket, the sum is

$$(-1)^{\langle z, \eta \rangle} + (-1)^{\langle z + \eta', \eta \rangle} = (-1)^{\langle z, \eta \rangle} + (-1)^{\langle z, \eta \rangle + \langle \eta', \eta \rangle} = (-1)^{\langle z, \eta \rangle} (1 + (-1)^{\langle \eta', \eta \rangle}) = 0,$$

so the overall sum is also 0. ◁

Using this result, we will prove Claim 18:

Proof of Claim 18. Identifying the 0's (resp. 1's) in \mathbb{F}_2 and \mathbb{Z}_s , we will rephrase the above claim as U being a subset of \mathbb{Z}_s^k instead of \mathbb{F}_2^k : Then, this set $U \subseteq \{0, 1\}^k \subseteq \mathbb{Z}_s^k$ immediately satisfies Item 1 of Claim 18. For Item 2, it suffices to show that for any non-zero junta-degree- d function $P : \mathbb{Z}_s^k \rightarrow \mathcal{G}$,

$$\sum_{y \in U} P(y) \cdot \chi(y) = 0 \tag{13}$$

where $\chi : \{0, 1\}^k \rightarrow \mathbb{Z}$ is from Claim 24. Towards a contradiction, suppose that a junta-degree- d function P has exactly one point in $y^* \in U$ such that $P(y^*) \neq 0$. Then using (13), $0 + \dots + 0 + P(y^*) \cdot \chi(y^*) + 0 + \dots + 0 = 0$ and as $\chi(y^*) = \pm 1$, $P(y^*) = 0$, a contradiction.

To prove (13), we expand P into its junta-polynomial representation:

$$\begin{aligned} \sum_{y \in U} P(y) \cdot \chi(y) &= \sum_{y \in U} \sum_{\substack{a \in \mathbb{Z}_s^k \\ \#a \leq d}} g_a \cdot \left(\prod_{i \in [k]: a_i \neq 0} \delta_{a_i}(y_i) \right) \chi(y) \\ &= \sum_{\substack{a \in \mathbb{Z}_s^k \\ \#a \leq d}} g_a \cdot \left(\sum_{y \in U} \chi(y) \prod_{i \in [k]: a_i \neq 0} \delta_{a_i}(y_i) \right) \end{aligned}$$

For any $a \in \mathbb{Z}_s^k$, letting $I := \{i \in [k] : a_i \neq 0\}$, the inner factor is

$$\sum_{y \in U} \chi(y) \prod_{i \in [k]: a_i \neq 0} \delta_{a_i}(y_i) = \sum_{y \in U: y^I = a^I} \chi(y).$$

Now if a contains any coordinates taking values other than 0 and 1, the above sum is 0 since all the coordinates of $y \in U$ are either 0 or 1. On the other hand, if $a \in \{0, 1\}^k$, then $a^I = 1^I$ and Claim 24 is applicable, again giving a sum of 0. Therefore,

$$\sum_{y \in U} P(y) \cdot \chi(y) = \sum_{\substack{a \in \mathcal{S}^k \\ \#a \leq d}} g_a \cdot \left(\sum_{y \in U} \chi(y) \prod_{i \in [k]: a_i \neq 0} \delta_{a_i}(y_i) \right) = \sum_{\substack{a \in \mathcal{S}^k \\ \#a \leq d}} g_a \cdot 0 = 0. \quad \triangleleft$$

B Proof of the large-distance lemma

Proof of Lemma 16. For this proof, we may assume without loss of generality that $I = [r]$ as relabelling the variables does not affect the probability of a random restriction (i.e., $x_j = \pi_j(x_i)$) being ε_0 -close to junta-degree- d . We will prove the contrapositive: assuming $\delta_d(f_{r-1}) \leq \varepsilon_0$ for more than $k^2/2r(r-1)$ fraction of choices of (i, j, π_j) (call these *bad* restrictions), we will construct a junta-degree- d function P such that $\delta(f_r, P) \leq \varepsilon_1$. Like in [7, 10], the high level idea is to “stitch” together low-junta-degree functions corresponding to the restrictions f_{r-1} (which we shall call $P^{(h)}$) into a low-junta-degree function that is close to f_r .

As there are more than $\frac{k^2}{2r(r-1)}r(r-1)s! = k^2s!/2$ many bad tuples (i, j, π_j) , by pigeon-hole principle, there must be some permutation $\pi : \mathbb{Z}_s \rightarrow \mathbb{Z}_s$ such that the number of bad tuples of the form (i, j, π) is more than $k^2/2$. In fact, we can say something more: Consider the directed graph G_{bad} over vertices $[r]$ with a directed edge (i, j) for each bad tuple (i, j, π) .

41:20 Low-Degree Testing over Grids

As the number of edges in G_{bad} is at least $k^2/2$, by the pigeon-hole principle, we can conclude that there is a *matching* or a *star*⁹ in G_{bad} of size $L := k/4$. For the rest of the proof, we will handle both the cases in parallel as the differences are minor.

Suppose we are in the matching case and the corresponding bad tuples are

$$(i_1, j_1, \pi), \dots, (i_L, j_L, \pi),$$

where $i_1, \dots, i_L, j_1, \dots, j_L$ are all distinct. Let id denote the identity permutation of \mathbb{Z}_s . Consider the function $\tilde{f}_r(x_1, \dots, x_r)$ obtained by replacing the variables x_{i_1}, \dots, x_{i_L} in f_r with $\pi(x_{i_1}), \dots, \pi(x_{i_L})$ respectively. Then, $\delta_d(\tilde{f}_r) = \delta_d(f_r)$ and (i_h, j_h, π) is a bad restriction for f_r if and only if (i_h, j_h, id) is a bad restriction for \tilde{f}_r , for all $h \in [L]$. Moreover, if \tilde{f}_r satisfies $\delta(\tilde{P}, \tilde{f}_r) \leq \varepsilon_1$, then there also exists a junta-degree- d P such that $\delta(P, f_r) \leq \varepsilon_1$ (obtained from \tilde{P} by applying the inverse permutation π^{-1} to x_{i_1}, \dots, x_{i_L}). Therefore, without loss of generality we may assume that $\pi = id$ to construct a junta-degree- d function P such that $\delta(P, f_r) \leq \varepsilon_1$. A similar reduction holds in the star case.

We may further assume w.l.o.g. that the matching case corresponds to the tuples

$$(L+1, 1, id), (L+2, 2, id), \dots, (2L, L, id)$$

and the star case corresponds to

$$(r, 1, id), (r, 2, id), \dots, (r, L, id).$$

For $h \in [L]$, we define

$$R_h := \begin{cases} \{x \in \mathbb{Z}_s^r : x_{L+h} = x_h\} & \text{in the matching case,} \\ \{x \in \mathbb{Z}_s^r : x_h = x_r\} & \text{in the star case.} \end{cases}$$

as the points that agree with the h -th bad restriction (i, j, π) in the matching or star case correspondingly. Let R'_h denote the complement of R_h . Then for any function $P : \mathbb{Z}_s^r \rightarrow \mathcal{G}$,

$$\begin{aligned} \Pr_{x \sim \mathbb{Z}_s^r} [f_r(x) \neq P(x)] &\leq \Pr_x \left[x \notin \bigcup_{h \leq L} R_h \right] \\ &\quad + \sum_{h \leq L} \Pr_x \left[x \in R_h \setminus \bigcup_{h' < h} R_{h'} \right] \cdot \Pr_x \left[f_r(x) \neq P(x) \mid x \in R_h \setminus \bigcup_{h' < h} R_{h'} \right] \end{aligned} \quad (14)$$

To estimate the above probabilities, we note that in both the matching or the star case, $\Pr_{x \sim \mathbb{Z}_s^r} [x \notin \bigcup_{h \leq L} R_h] = \Pr_x [x \in \bigcap_{h \leq L} R'_h] = (1 - \frac{1}{s})^L$, and $\Pr_{x \sim \mathbb{Z}_s^r} [x \in R_h \setminus \bigcup_{h' < h} R_{h'}] = \Pr_x [x \in R_h \cap \bigcap_{h' < h} R'_{h'}] = \frac{1}{s} (1 - \frac{1}{s})^{h-1}$.

For $h \in [L]$, let $f_{r-1}^{(h)} : \mathbb{Z}_s^r \rightarrow \mathcal{G}$ be the restricted function corresponding to the h -th bad tuple, treated as a function of all the r many variables (rather than $r-1$). Let $P^{(h)}$ denote the junta-degree- d function that is of distance at most ε_0 from $f_{r-1}^{(h)}$. We will use the following claim that there is a junta-degree- d function P that agrees with $P^{(h)}$ over R_h , for all h .

▷ **Claim 25.** There exists a junta-degree- d function P such that $P(x) = P^{(h)}(x)$ for all $h \in [L]$ and $x \in R_h$.

⁹ A matching is a set of disjoint edges and a star is a set of edges that share a common start vertex, or a common end vertex.

We defer its proof to the full version and only mention here that the idea is to first show for $h \neq h' \leq L$ that $P^{(h)}|_{h'} = P^{(h')}|_h$, since both these functions agree (with $f_{r-1}^{(h)}$ and $f_{r-1}^{(h')}$) over a “large” subset $R_h \cap R_{h'}$ of their domain. Then one can interpolate the restricted functions into a junta-degree- d function P . Then for such P and any $h \leq L$,

$$\begin{aligned} \Pr_x \left[f_r(x) \neq P(x) \mid x \in R_h \setminus \bigcup_{h' < h} R_{h'} \right] &\leq \frac{\Pr_x [f_r(x) \neq P(x) \mid x \in R_h]}{\Pr [x \in R_h \setminus \bigcap_{h' < h} R_{h'} \mid x \in R_h]} \\ &= \frac{\Pr_x [f_r(x) \neq P(x) \mid x \in R_h]}{\left(1 - \frac{1}{s}\right)^{h-1}} \\ &\leq \left(\frac{s}{s-1}\right)^{h-1} \varepsilon_0. \end{aligned}$$

Then we can bound (14) as $\Pr_{x \sim \mathbb{Z}_s^r} [f_r(x) \neq P(x)] \leq \left(1 - \frac{1}{s}\right)^L + \frac{L\varepsilon_0}{s} \leq \varepsilon_1/2 + \varepsilon_1/2 = \varepsilon_1$ as we can set $\varepsilon_0 := 2s\varepsilon_1/k \geq 1/s^{O(k)}$ and

$$\left(1 - \frac{1}{s}\right)^{k/4} \leq e^{-k/4s} = e^{-\psi sd/4} \leq \frac{1}{2} \left(\frac{1}{2C2^{\lceil \log(8\psi s^2)d \rceil}}\right)^{2^{14} \log s} = \frac{\varepsilon_1}{2}.$$

(for the last inequality, we can take ψ to be a sufficiently large constant)

◀

C Proof of Claim 22

Proof of Claim 22. The proof is by induction on w . The base case $w = 1$ is crucial and it is equivalent to the following claim:

▷ **Claim 26.** For $0 \leq d' \leq s - 1$, the function $f_{d'} : \mathcal{B}(\mathcal{S}, t) \rightarrow \mathbb{F}$ defined as $f_{d'}(z) = z_1^{d'}$ for $z = (z_1, \dots, z_t) \in \mathcal{B}(\mathcal{S}, t)$ has degree exactly d' .

Assuming the above claim to be true, let $w > 1$ be arbitrary. As $d' = 0$ is trivial to handle, we will assume that $d' \geq 1$. Hence, Q contains at least one monomial m of degree d' and containing some variable $y_{(1,j)}$ with individual degree $a \in [s - 1]$. Without loss of generality, suppose $j = w$. Since Claim 26 states that the function $y_{(1,w)}^a$ is linearly independent of degree- $(a - 1)$ functions over $\mathcal{B}(\mathcal{S}, t)$, there exists a function $C : \mathcal{B}(\mathcal{S}, t) \rightarrow \mathbb{F}$ such that for any $f : \mathcal{B}(\mathcal{S}, t) \rightarrow \mathbb{F}$

$$\langle C, f \rangle = \begin{cases} 1 & \text{if } f = y_{(1,w)}^a \text{ i.e., } a\text{-th power of the last coordinate} \\ 0 & \text{if } f \text{ is degree-}(a - 1). \end{cases} \quad (15)$$

Now we decompose Q as a polynomial over variables in the first $w - 1$ columns and coefficients being the monomials over variables in the last column: that is

$$Q(y) = \sum_{\alpha \in [0..a]} Q'_\alpha(y_1, \dots, y_{w-1}) \cdot y_{(1,w)}^\alpha, \quad (16)$$

where y_j represents the variables in the j -th column $Q'_\alpha \neq 0$ has degree $d' - a$. Here, we are using the fact that Q only contains variables from the first row.

Towards a contradiction, suppose there is some degree- $(d' - 1)$ polynomial $R(y)$ such that $Q(y) = R(y)$ for all $y \in \mathcal{B}(\mathcal{S}, t)^w$. We may decompose R as follows:

41:22 Low-Degree Testing over Grids

$$R(y) = \sum_{\alpha \in [0..s-1]^t} R'_\alpha(y_1, \dots, y_{w-1}) \cdot y_w^\alpha$$

where $y_w^\alpha = y_{(1,w)}^{\alpha_1} \cdots y_{(t,w)}^{\alpha_t}$ and for all α , either $R'_\alpha = 0$ or is of degree (as a formal polynomial) at most $d' - 1 - |\alpha|_1$, where $|\alpha|_1 = \alpha_1 + \cdots + \alpha_t$.

Fixing $y_1, \dots, y_{w-1} \in \mathcal{B}(\mathcal{S}, t)$ to arbitrary values and treating $Q(y)$ and $R(y)$ as functions of y_w , we get

$$\begin{aligned} \langle C, Q(y_1, \dots, y_{w-1}, y_w) \rangle &= \left\langle C, \sum_{\alpha \in [0..a]} Q'_\alpha(y_1, \dots, y_{w-1}) \cdot y_{(1,w)}^\alpha \right\rangle \\ &= \sum_{\alpha \in [0..a]} Q'_\alpha(y_1, \dots, y_{w-1}) \cdot \langle C, y_{(1,w)}^\alpha \rangle \\ &= Q'_a(y_1, \dots, y_{w-1}). \end{aligned} \quad (\text{using (15)})$$

Similarly,

$$\begin{aligned} \langle C, R(y_1, \dots, y_{w-1}, y_w) \rangle &= \left\langle C, \sum_{\alpha \in [0..s-1]^t} R'_\alpha(y_1, \dots, y_{w-1}) \cdot y_w^\alpha \right\rangle \\ &= \sum_{\alpha \in [0..s-1]^t: |\alpha|_1 \geq a} R'_\alpha(y_1, \dots, y_{w-1}) \cdot \langle C, y_w^\alpha \rangle \end{aligned} \quad (\text{using (15)})$$

As a polynomial in the variables of y_1, \dots, y_{w-1} , the final expression above is of degree at most $d' - 1 - |\alpha|_1 \leq d' - 1 - a$. However, as a function it is equivalent to Q'_a , which has a strictly higher degree, $d' - a$. This contradicts the induction hypothesis. \triangleleft