Low Degree Testing over the Reals

Vipul Arora[1], Arnab Bhattacharyya[*1], Noah Fleming[†2], Esty Kelman[‡3], and Yuichi Yoshida[§4]

[1]National University of Singapore. {vipul, arnab}@comp.nus.edu.sg.
[2]University of California, San Diego, and Memorial University. nfleming@mun.ca
[3]Hebrew University, and Reichman University. esther.kelman@post.idc.ac.il.
[4]National Institute of Informatics. yyoshida@nii.ac.jp.

April 18, 2022

Abstract

We study the problem of testing whether a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is a polynomial of degree at most d in the *distribution-free* testing model. Here, the distance between functions is measured with respect to an unknown distribution $\mathcal{D}$ over $\mathbb{R}^n$ from which we can draw samples. In contrast to previous work, we do not assume that $\mathcal{D}$ has finite support.

We design a tester that given query access to f, and sample access to $\mathcal{D}$, makes $\mathsf{poly}(d/\varepsilon)$ many queries to f, accepts with probability 1 if f is a polynomial of degree d, and rejects with probability at least $2/3$ if every degree-d polynomial P disagrees with f on a set of mass at least ε with respect to $\mathcal{D}$. Our result also holds under mild assumptions when we receive only a polynomial number of bits of precision for each query to f, or when f can only be queried on rational points representable using a logarithmic number of bits. Along the way, we prove a new stability theorem for multivariate polynomials that may be of independent interest.

1 Introduction

Traditionally, program testing involves running a suspect program on a curated test set and checking the validity of the results. To formalize and quantitatively study this problem, Blum, Luby, and Rubinfeld [BLR93] initiated research on *self-testers*, which check a particular property of the given program by verifying whether the program's output on a random input is consistent with its outputs on other related inputs. Soon afterwards, spurred by connections to the newly emerging areas of interactive proof systems and probabilistically checkable proofs, self-testing blossomed into the general area of *property testing*; see the textbooks [Gol17, BY22] for detailed introductions. Perhaps because of these early connections to complexity theory and coding theory, the standard setup in property testing is to assume that both the domain and range of the function being tested are finite sets.

[*]Supported in part by an NRF Tier 2 grant (MOE2019-T2-1-152). Research partly conducted while visiting the Simons Institute for the Theory of Computing.

[†]Supported by NSERC.

[‡]Supported in part by an Amazon Faculty Research Award to AB, and in part by ERC grant 834735. Research partly conducted while visiting the Simons Institute for the Theory of Computing.

[§]Supported in part by JSPS KAKENHI Grant Number JP17H04676, and 20H05965.

In this work, we return to the roots of property testing and consider testing properties of real-valued functions $f : \mathbb{R}^n \to \mathbb{R}$ with real-valued inputs. Specifically, we focus on the fundamental problem of *low-degree testing* which has been widely and intensely studied in the standard setup. Recall that in the traditional setup we are given query access to a function $f\colon \mathbb{F}^n \to \mathbb{F}$ defined over some finite field $\mathbb{F}$ and a parameter $\varepsilon > 0$. The aim of a property tester for these parameters is to distinguish with probability at least $2/3$ between the case when f is a polynomial of degree at most d, and the case when f is ε-far, i.e., it disagrees with any polynomial P of degree at most d on at least an ε fraction of the domain $\mathbb{F}^n$.

To extend the notion of testing to functions defined over $\mathbb{R}^n$, we need a notion of "ε-farness" in this setting. One approach is to fix a specific distribution $\mathcal{D}$ over $\mathbb{R}^n$, and define f to be ε-far from a property $\mathcal{P}$ if:

$$\inf_{g \in \mathcal{P}} \Pr_{\boldsymbol{x} \sim \mathcal{D}}[f(x) \neq g(x)] > \varepsilon. \tag{1}$$

Indeed, for $\mathcal{D}$ being the standard Gaussian distribution, this is the approach used by most prior work on testing properties of functions over the reals, e.g., testing halfspaces [MORS10b, MORS10a, MORS09, Har19], surface area [Nee14, KNOW14], high-dimensional convexity [CFSS17], linear separators [BBBY12], and linear k-juntas [DMN19]. However, this approach is not entirely satisfactory, as the assumed $\mathcal{D}$ may not be the relevant underlying input distribution.

A different approach is to use the framework of *distribution-free testing*, studied first by Halevy and Kushilevitz [HK07], that does not assume knowledge of $\mathcal{D}$. Instead, it is only assumed that the tester receives sample access to the underlying distribution $\mathcal{D}$, and the goal is to reject when (1) holds. Distribution-free testing has been widely studied for a variety of properties of boolean functions, e.g., monomials [GS09,DR11], juntas [LCS+19], halfspaces [CX16, CP22], and monotonicity [BCS20]. Distribution-free property testing over $\mathbb{R}^n$ is an emerging trend in the field, that has been studied, e.g., for monotonicity [BCS20, HY20], halfspaces [Har19] and polynomial threshold functions [BFPJH21]. Most directly relevant here is the work of Fleming and Yoshida [FY20] where they studied distribution-free testing of linearity of functions $f : \mathbb{R}^n \to \mathbb{R}$.

To further discuss testing real functions, we first formally define distribution-free testing of real functions. For a property $\mathcal{P}$ over real functions, we say that an algorithm is a *tester* for $\mathcal{P}$ if, given query access to a function $f : \mathbb{R}^n \to \mathbb{R}$, and sampling access to an unknown distribution $\mathcal{D}$, and $\varepsilon > 0$, it distinguishes the case that f satisfies $\mathcal{P}$, from the case that f is *ε-far from $\mathcal{P}$ over $\mathcal{D}$*, i.e., for any function $g : \mathbb{R}^n \to \mathbb{R}$ satisfying $\mathcal{P}$,

$$\Pr_{\boldsymbol{x} \sim \mathcal{D}}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})] > \varepsilon$$

holds. We say that a tester is a *one-sided error tester*, if it always accepts functions satisfying $\mathcal{P}$. We also explore testing in the presence of errors. In this context, the early works [ABCG93, GLR+91] introduced the notion of *approximate testing*, which was made more formal by the work of Ergun, Kumar and Rubinfeld [EKR01]. Given two parameters $\alpha < \beta$, in addition to $\varepsilon > 0$, query access to $f : \mathbb{R}^n \to \mathbb{R}$ and sample access to a distribution $\mathcal{D}$, the goal of an approximate tester for a property $\mathcal{P}$ is to distinguish between the following two cases:

- **YES:** There exists $h \in \mathcal{P}$, such that $|f(\boldsymbol{x}) - h(\boldsymbol{x})| < \alpha$ for all $\boldsymbol{x} \in \mathbb{R}^n$.
- **NO:** For every $h \in \mathcal{P}$, $\Pr_{\boldsymbol{x} \sim \mathcal{D}}[|f(\boldsymbol{x}) - h(\boldsymbol{x})| > \beta] > \varepsilon$.

In the **YES** case, we say that f is *pointwise α-close to h*. Here, α should be thought of as a representational limitation, or a round-off/truncation error. For example, $\alpha = 1/\exp(\mathrm{poly}(n))$ can be achieved by storing $\mathrm{poly}(n)$ bits of precision.

1.1 Our Contributions

Our first result gives an exact tester for low-degree that generalizes the result of [FY20]. Note that there is a trivial $\Omega(\max\{d, 1/\varepsilon\})$ lower bound on the complexity of testing degree-d polynomials.

Theorem 1.1. *Let $d \in \mathbb{N}$, and for $L > 0$, suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function that is bounded in the ball $\mathrm{B}(\mathbf{0}, L)$. Given $\varepsilon > 0$, query access to f, and sampling access to an unknown distribution $\mathcal{D}$, there exists a one-sided error, distribution-free, $O(d^5 + \frac{d^2}{\varepsilon}\log\frac{1}{\varepsilon})$-query tester for testing whether f is a degree-d polynomial, or is ε-far from degree-d polynomials over $\mathcal{D}$.*

Some form of the boundedness condition is necessary to test low degree using standard functional equation characterizations. Even for linearity, Hamel [Ham05] showed the existence of functions $f : \mathbb{R} \to \mathbb{R}$ that satisfy the *Cauchy functional equation* $f(x + y) = f(x) + f(y)$ everywhere but are unbounded on *any* measurable set[1]. On the other hand, Cauchy [Cau21] showed that the only continuous solutions to $f(x+y) = f(x) + f(y)$ are the linear maps $f(x) = cx$. Darboux [Dar75] later showed that boundedness on any interval is a weaker condition than continuity that also implies the result of Cauchy. The latter two results were generalized to low-degree polynomials by Fréchet [Fré09] and Ciesielski [Cie59] respectively.

Theorem 1.1 serves as a starting point for our investigation into approximate low-degree testing. In this setting, we give an approximate tester for low-degree polynomials, where the unknown underlying distribution $\mathcal{D}$ is required to be (ε, R)-*concentrated*. We say that a distribution is (ε, R)-concentrated if most of its mass is concentrated in a ball of radius R, that is,

$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)] \geq 1 - \varepsilon.$$

Note that the standard Gaussian distribution is $(0.01, 2\sqrt{n})$-concentrated.

Theorem 1.2. *Let $d \in \mathbb{N}$, $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is bounded in $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, and for $\varepsilon \in (0, 1)$, $R > 0$, let $\mathcal{D}$ be an $(\varepsilon/4, R)$-concentrated distribution. Given $\alpha > 0, \beta \geq 2^{(2n)^{O(d)}} R^d \alpha$, query access to f, and sampling access to $\mathcal{D}$, there is a one-sided error, $O(d^5 + \frac{d^2}{\varepsilon}\log\frac{1}{\varepsilon})$-query tester which, distinguishes between the case when f is pointwise α-close to some degree-d polynomial and the case when, for every degree-d polynomial $h\colon \mathbb{R}^n \to \mathbb{R}$, $\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > \beta] > \varepsilon$.*

Thus, if d is constant, R is polynomial in n, and the tester receives $\mathsf{poly}(n)$ most significant bits of $f(\boldsymbol{p})$ for any query point $\boldsymbol{p}$, the tester accepts when f is a degree-d polynomial, and rejects when f is not pointwise 1-close to a degree-d polynomial on at least an ε fraction of $\mathcal{D}$. In Appendix D, we consider the special case of testing additivity. Here, we give a tester which requires only $O(\log n)$ bits of precision.

The above results assume that the function can be queried on arbitrary points in $\mathbb{R}^n$ which is unrealistic in view of finite precision issues. We also analyze the setting where the tester can evaluate f only on points with finite number of bits of precision and also, the unknown distribution $\mathcal{D}$ is promised to be supported on points with finite number of bits of precision. More precisely, $\mathcal{D}$ is given to be supported on points of the lattice $\mathcal{L} \triangleq \frac{1}{B}\mathbb{Z}^n$, for some parameter B controlling the density of the lattice, and also, f can be queried only on a lattice $\mathcal{L}' \triangleq \frac{1}{B'}\mathbb{Z}^n$ for a bounded B'. This setting models the situation where we only care about the function's behavior on finitely representable inputs, and on such inputs, the function can be evaluated exactly. The goal is to obtain a tester that does not require B' to be very large but still allows B to be large.

[1] In fact, Hamel showed that if f is a non-linear solution, the set $\{(x, f(x))\}$ intersects every neighborhood of every point in $\mathbb{R} \times \mathbb{R}$, and so is clearly unbounded on any measurable set.

Theorem 1.3. *For $d, B, R > 0$, let $B' \geq 16 \max\{n^{5/2+2d} d^{2d}, B^2 R^2/\sqrt{n}\}$ be a multiple of B. Let $\mathcal{L} = \frac{1}{B}\mathbb{Z}^n$ and $\mathcal{L}' = \frac{1}{B'}\mathbb{Z}^n$. Given $\varepsilon > 0$, query access to a function $f : \mathbb{R}^n \to \mathbb{R}$, and sample access to an unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$ supported on $\mathcal{L}$, there is a one-sided error, $O(d^5 + \frac{d^2}{\varepsilon} \log \frac{1}{\varepsilon})$-query tester for testing whether f agrees with a degree-d polynomial on $\mathcal{L}$, or is ε-far from degree-d polynomials over $\mathcal{D}$. The tester queries f on points in $\mathcal{L}'$.*

Note that unlike Theorem 1.2 and Theorem 1.1, our tester for lattices does not make any assumptions about the function f.

1.2 Related Work

Although distribution-free testing (for graph properties) was already defined in an early work on property testing [GGR98], the first distribution-free testers for non-trivial properties appeared much later in the work of Halevy and Kushilevitz [HK07]. Since then, distribution-free testers have been considered for a variety of Boolean functions including low-degree polynomials, dictators, and monotone functions [HK07], k-juntas [HK07, LCS+19, Bsh19, Bel19], conjunctions, decision lists, and linear threshold functions [GS09], monotone and non-monotone monomials [DR11], and monotone conjunctions [GS09, CX16]. The first (partial) distribution-free testing result for functions on the Euclidean space was due to Harms [Har19]: He gave an efficient tester for half spaces over any rotationally invariant distribution. Then, as we mentioned above, Fleming and Yoshida [FY20] gave a tester for linearity of functions over the Euclidean space.

Property testing originated (implicitly, under the name of self-testing) in the work of Blum, Luby, and Rubinfeld [BLR93], who exhibited the famous BLR tester for linearity over $\mathbb{F}_2$. Since then, testers have been developed for higher degree polynomials, such as the famous Rubinfeld Sudan [RS96] and Raz and Safra [RS97] tests for degree-d polynomials over sufficiently large finite fields. One line of work, closely related to ours extended the domain over which these testers worked, culminating in the work of Lipton [Lip89] and Rubinfeld and Sudan [RS92], who gave testers for degree-d polynomials over any finite subset of rationals, where the distance is measured according to the uniform distribution; see [KMS01] for an excellent survey. The main distinguishing features between this paper and the works of [Lip89, RS92] is that (i) we work in the distribution-free setting, (ii) we do not assume that the domain is finite, and (iii) the input function is multivariate.

1.3 Proof Overview

This work significantly extends the framework of Fleming and Yoshida [FY20], who exhibited a constant-query algorithm for testing the linearity of functions over $\mathbb{R}^n$ in the distribution-free setting (when distance is measured according to an arbitrary distribution $\mathcal{D}$); thus, we briefly describe their proof first.

Testing Linearity over the Reals. The tester follows the high-level "self-correct and test" approach of Halevy and Kushilevitz [HK07]. To test whether a given function $f : \mathbb{R}^n \to \mathbb{R}$ is linear, it suffices to construct a *linear* function $g_{\mathsf{lin}} : \mathbb{R}^n \to \mathbb{R}$ such that:

1. If f is indeed a linear function, then $f = g_{\mathsf{lin}}$.
2. For any $\boldsymbol{p} \in \mathbb{R}^n$, we can efficiently query the value of $g_{\mathsf{lin}}(\boldsymbol{p})$ using queries to f.

Indeed, by (1), to test if f is linear, it suffices to estimate the distance between f and g_{lin} (measured according to $\mathcal{D}$), which can be done efficiently by (2).

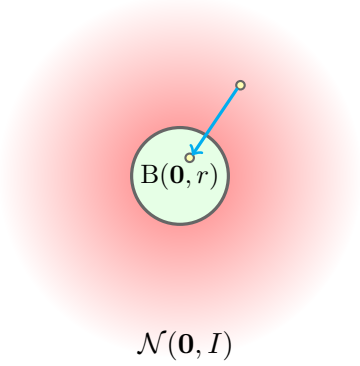

Figure 1: Each point in the Gaussian distribution is projected into the small ball $\mathrm{B}(\mathbf{0}, r)$ at the origin.

To construct g_{lin} they use the standard self-correcting approach pioneered in the Blum, Luby, and Rubinfeld (BLR) test for linearity over $\mathsf{GF}(2)$ [BLR93]. However, this has to be significantly modified. Standard self-correction arguments require that every point in the distribution has equal probability mass, and there is no natural analogue to the uniform distribution over $\mathbb{R}^n$. Instead, they modify the self-correcting argument to work for the standard Gaussian distribution — that is, by evaluating f on points sampled from $\mathcal{N}(\mathbf{0}, I)$, they are able to construct the desired function g_{lin}. Note that even though g_{lin} is constructed using samples from $\mathcal{N}(\mathbf{0}, I)$, in order to test whether f is close to a linear function over the given distribution $\mathcal{D}$, by (1) it suffices to estimate the distance between f and g_{lin} over $\mathcal{D}$. This can be done by sampling sufficiently many points $\boldsymbol{p} \sim \mathcal{D}$ and checking whether $f(\boldsymbol{p}) = g_{\mathsf{lin}}(\boldsymbol{p})$, using (2) in order to evaluate $g_{\mathsf{lin}}(\boldsymbol{p})$.

To circumvent the issue that points have differing probability mass under $\mathcal{N}(\mathbf{0}, I)$, they project every point into a Euclidean ball $\mathrm{B}(\mathbf{0}, r)$ of small radius r at the center of the Gaussian (see Figure 1). Within this ball, every point has approximately the same mass and they are able to perform the self-correction argument. In particular, they define

$$g_{\mathsf{lin}}(\boldsymbol{p}) \triangleq \gamma_{\boldsymbol{p}} \cdot \operatorname*{maj}_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)} \left[f\left(\frac{\boldsymbol{p}}{\gamma_{\boldsymbol{p}}} - \boldsymbol{q} \right) + f(\boldsymbol{q}) \right],$$

where $\gamma_{\boldsymbol{p}} \in \mathbb{R}$ is such that $\boldsymbol{p}/\gamma_{\boldsymbol{p}} \in \mathrm{B}(\mathbf{0}, r)$. That is, $g_{\mathsf{lin}}(\boldsymbol{p})$ is the majority value weighted according to the standard Gaussian distribution. This is essentially the same self-corrected function used in the BLR test, except that each point $\boldsymbol{q}$ is first projected into $\mathrm{B}(\mathbf{0}, r)$.

Finally they argue that, if their tests pass with a sufficiently high probability, then g_{lin} is a linear function, and furthermore, for any $\boldsymbol{p} \in \mathbb{R}^n$, the value of $g_{\mathsf{lin}}(\boldsymbol{p})$ can be recovered with a small number of queries to f.

Exactly Testing Polynomials over the Reals. Our work is a significant generalization of the ideas used in the linearity test so that they may be applied to degree-d polynomials. Given a function $f \colon \mathbb{R}^n \to \mathbb{R}$, we construct a degree-$d$ polynomial $g \colon \mathbb{R}^n \to \mathbb{R}$ such that

1. If f is a degree-d polynomial, then $f = g$.
2. For any $\boldsymbol{p} \in \mathbb{R}^n$, we can efficiently query the value $g(\boldsymbol{p})$ using queries to f.

As in the the case of linear functions, we construct g using samples from the Gaussian distribution. We mitigate the fact that points are weighted non-uniformly, by restricting attention to a small (open) ball $\mathrm{B}(\mathbf{0}, r)$,

defining g within that ball, and then extending outwards. Formally, let $\alpha_i \triangleq (-1)^{i+1}\binom{d+1}{i}$, and for any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ and $\boldsymbol{q} \in \mathbb{R}^n$, let $g_{\boldsymbol{q}}(\boldsymbol{p}) \triangleq \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})$. For points $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, we define g to be

$$g(\boldsymbol{p}) \triangleq \operatorname*{maj}_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)} [g_{\boldsymbol{q}}(\boldsymbol{p})]\,,$$

where the majority is weighted according to $\mathcal{N}(\mathbf{0}, I)$. For points $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$, the value of $g(\boldsymbol{p})$ is defined by interpolating from evaluations of g, on $d+1$ distinct points within $\mathrm{B}(\mathbf{0}, r)$ (this is defined formally in Section 3).

Having thus defined g, we would like to argue that if a certain set of tests pass with sufficiently high probability, then g is a degree-d polynomial. We make this argument in three steps, where each step extends the domain over which we guarantee that g is a polynomial.

1. We show that g is consistent with a degree-d univariate polynomial on any line segment $\mathrm{L}^{\mathrm{B}}_{\boldsymbol{a},\boldsymbol{b}} \triangleq \{\boldsymbol{a} + x\boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r) : x \in \mathbb{R}\}$ within the ball $\mathrm{B}(\mathbf{0}, r)$. To prove this, we generalize the self-correction argument from [RS96] to hold over the ball $\mathrm{B}(\mathbf{0}, r)$ of reals.

2. We show how to stitch together these "local" representations of g on lines into a degree-d *multivariate* polynomial, which is consistent with g within a hypercube contained within $\mathrm{B}(\mathbf{0}, r)$. We describe this step in more detail below.

3. We extend this representation of g within the hypercube to a consistent representation of g as a degree-d polynomial everywhere. This follows by extrapolating g from the small ball $\mathrm{B}(\mathbf{0}, r)$ to all of $\mathbb{R}^n$.

The main innovation is step (2), and therefore we will describe it in more detail. Step (2) is proved in two parts: first, we argue that g can be represented as a polynomial of degree dn; second, we reduce the degree to d.

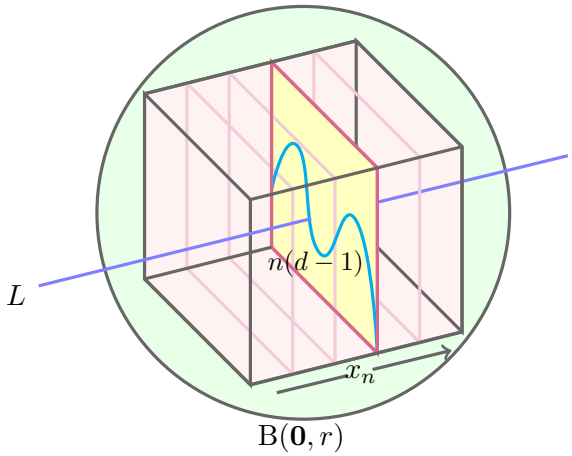

Figure 2: The construction of the degree nd representation of g. $d+1$ slices, parallel to the x_n axis, of the cube are chosen. On each slice g is a degree $n(d-1)$ polynomial (cyan). These degree $n(d-1)$ representations on slices are stitched together along a line L.

To prove the first part, we consider the largest n-dimensional cube that can be inscribed in the ball $\mathrm{B}(\mathbf{0}, r)$. We then discretize the cube by picking $d+1$ slices perpendicular to the x_n axis ($(n-1)$-dimensional

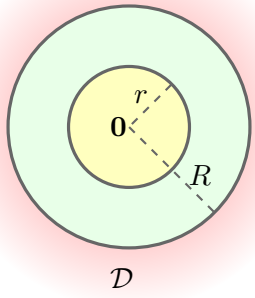

Figure 3: The ball $\mathrm{B}(\mathbf{0}, R)$, containing at least $1 - \varepsilon/4$ of the mass of the $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$, and the ball $\mathrm{B}(\mathbf{0}, r)$ from which g is extrapolated.

sub-cubes), and argue by induction that g can be written as a degree $(n(d-1))$ polynomial on each slice[2] (see Figure 2). To combine these $n(d-1)$-degree polynomials into a degree-dn polynomial, we consider any L line parallel to the x_n axis. This line has exactly one intersection point with each of the $d+1$ slices. By step (1), g restricted to this line is a degree-d *univariate* polynomial. Using this univariate representation of g on the line to interpolate between the $n(d-1)$-degree representations of g on the slices allows us to obtain a representation of g as a degree-dn polynomial within the hypercube.

To reduce the degree of this representation of g from dn to d, we use the fact that (by step (1)) g can be represented as a degree-d univariate polynomial on every line segment within the ball. In particular, we show that for any representation if g as a polynomial of some degree m, there exists a radial line $L_{\mathbf{0},\boldsymbol{b}}$ such that g restricted to this line also has degree m. However, by step (1), g restricted to any line has degree at most d, and this implies that $m \leq d$.

Approximate Testing Polynomials over the Reals. The major new challenge that arises in approximate testing is that we must ensure that our tester accepts all functions that are *pointwise δ-close* to being a polynomial; i.e., we should accept f if there exists a degree-d polynomial $\widehat{f}$ such that for every $\boldsymbol{x} \in \mathbb{R}^n$

$$|f(\boldsymbol{x}) - \widehat{f}(\boldsymbol{x})| \leq \delta.$$

We work in the setting where the unknown distribution $\mathcal{D}$ is known to satisfy the condition that $1 - \varepsilon/4$ fraction of the mass of $\mathcal{D}$ is contained within $\mathrm{B}(\mathbf{0}, R)$ for a given parameter R (see Figure 3).

We begin by constructing a self-corrected function g, except that now, it is in terms of the median[3] instead of the majority. Our analysis then follows the three-step outline mentioned above for the exact case. In the first step, we argue that g approximately satisfies the univariate characterization of degree-d polynomials on every line restricted to $\mathrm{B}(\mathbf{0}, r)$, and hence, g is pointwise close to a low degree univariate polynomial on every such line segment. The last conclusion is due to a theorem of Gajda [Gaj91] from the literature on Hyers-Ulam stability results for functional equations; see the book [HIR12] for a comprehensive survey of this area.

Our main technical contribution comes in the second step of the analysis. We show that being pointwise close to a multivariate low-degree polynomial is approximately a 'lifted' property [GKS13].

[2]The reason that we use a hypercube embedded within the ball — rather than using $d+1$ slices of $\mathrm{B}(\mathbf{0}, r)$ — is that we require each of the $(n-1)$-dimensional polynomials have the same domain. If we took $d+1$ slices of $\mathrm{B}(\mathbf{0}, r)$, this would not be true.

[3]The use of median in the context of approximate testing is not a new idea; see, e.g., [KMS01].

Lemma 1.4. *Let $m \in (0,1], \delta > 0$, and let $h\colon [-m,m]^n \to \mathbb{R}$. If for every line L, there exists a degree-d univariate polynomial $\widehat{h}_{\mathrm{L}}$ that is pointwise δ-close to h_{L} (the restriction of h on the line L), then h is pointwise $((2/m)^{(n^{40d})}\delta)$-close to a degree-$d$ polynomial.*

The proof of Lemma 1.4 is by induction on n, where we show in each step, that (i) the function is pointwise close to a degree-$2d$ polynomial, and then that (ii) the function from step (i) is pointwise close to a degree-d polynomial. Both parts refine the corresponding analysis in the exact case.

- For part (i), we choose $d+1$ hyperplanes $\mathrm{H}_0 \triangleq \{x_n = c_0\}, \ldots, \mathrm{H}_d \triangleq \{x_n = c_d\}$ where $c_0, \ldots, c_d$ are the scaled *Chebyshev nodes*. By induction, there exist degree-d polynomials $\widehat{g}_i$ that are pointwise close to g on H_i. Now, for any line L parallel to the x_n axis, we look at the univariate degree d polynomial g_{L} that g is pointwise close to, and the degree d polynomial $\widehat{g}_{\mathrm{L}}$ that agrees with $\widehat{g}_i$ for each of the intersections between L and H_i. The difference $g_{\mathrm{L}} - \widehat{g}_{\mathrm{L}}$ is small at the Chebyshev nodes, which implies that $g_{\mathrm{L}} - \widehat{g}_{\mathrm{L}}$ is small everywhere inside $[-m,m]^n$. This argument yields a degree-$2d$ polynomial that is pointwise close to g on $[-m,m]^n$.

- We prove a more general result that implies what we need in part (ii).

  Theorem 1.5. *Let $m \in (0,1]$, $n \geq 2$ and p be an n-variate polynomial of total degree at most ℓ, for some $d \leq \ell$. If for every $\boldsymbol{a} \in [-m,m]^n$, the univariate polynomial $p_{\mathbf{0},\boldsymbol{a}}(t) = p(\boldsymbol{a}t)$ which is the restriction of p to the radial line $\mathrm{L}_{\mathbf{0},\boldsymbol{a}}$, is pointwise ε-close to a degree-d univariate polynomial on the interval $t \in [-1,1]$, then p is pointwise η-close to $p^{\leq d}$ (the truncation of p to degree d) on $[-m,m]^n$ for $\eta = 2(2/m)^{2n^{18\ell}}\varepsilon$.*

  In order to prove Theorem 1.5, suppose for the sake of contradiction that $p - p^{\leq d}$ is large at some point in $[-m,m]^n$. By a straightforward argument, this implies that there must be coefficient α_I of a degree $\geq d$ monomial in p which has large magnitude. From this, we would like to conclude that the restriction of p to some radial line $\mathrm{L}_{\mathbf{0},\boldsymbol{a}}$ must not be pointwise close to a degree-d polynomial, and hence we would have a contradiction. Let the restriction of p to this line be defined as

  $$p(\boldsymbol{a}t) = \sum_{k \leq \ell} \gamma_k(\boldsymbol{a}) T_k(t),$$

  where T_k is the kth *Chebyshev polynomial*. It turns out that in order to show $p(\boldsymbol{a}t)$ is not close to a degree-d polynomial, it suffices to show $\gamma_k(\boldsymbol{a})$ is large for some $k > d$.

  The large coefficient α_I of p appears in some coefficient $\gamma_{k^*}(\boldsymbol{a})$ for $k^* > d$[4]. Note that γ_{k^*} is itself a degree-ℓ multivariate polynomial when we consider $\boldsymbol{a}$ as variables. In order to conclude that $\gamma_k(\boldsymbol{a})$ is large for some $\boldsymbol{a}$, we will choose values for $\boldsymbol{a}$ such that γ_{k^*} is a degree-d univariate polynomial (in some variable z) and there is a monomial in γ_{k^*} with a large coefficient; anti-concentration then implies that there is a setting of z which makes γ_{k^*} large. To satisfy this, we want to choose a substitution for $\boldsymbol{a}$ in z such that the monomials under this substitution have exactly the same coefficients as those of $\gamma_{k^*}(\boldsymbol{a})$ (that is, no two monomials collapse to the same monomial).

  Fixing a formal variable z, we set $\boldsymbol{a}$ to be $(z^{y_1}, \ldots, z^{y_n})$ for an integer valued vector $\boldsymbol{y} = (y_1, \ldots, y_n)$, and define $\widetilde{\gamma}_{k^*}(z) = \gamma_{k^*}(z^{y_1}, \ldots, z^{y_n})$. We choose $\boldsymbol{y}$ in such a way that distinct monomials of $\boldsymbol{a}$ in γ_{k^*} lead to distinct powers of z in $\widetilde{\gamma}_{k^*}$; such a $\boldsymbol{y}$ exists due to a probabilistic argument.

[4] In fact, k^* is either $d+1$ or $d+2$.

At this point, we have a univariate polynomial $\widetilde{\gamma}_{k^*}$ that has at least one large coefficient, and we would like to conclude that it has a large value at some point. This is a statement about the *anti-concentration* of the polynomial $\widetilde{\gamma}_{k^*}$. If the largest coefficient were the leading term, then it is well-known that Chebyshev polynomials attain the smallest uniform norm on $[-1, 1]$ among all such polynomials. In our situation, the largest coefficient may not be the leading one; nevertheless, we can show a lower bound on the uniform norm by making a connection to Chebyshev polynomials[5]:

Lemma 1.6. *Let $p(x) = \sum_{i=0}^{d} \alpha_i x^i$ be a degree-d polynomial and let $\eta > 0$. If $|\alpha_i| \geq \eta$ for some $i \geq 1$, then there exists $x \in [-1, 1]$ such that $|p(x)| \geq 2^{-2d^2}\eta$.*

We now return to the main thread of describing the three-step analysis for Theorem 1.2. In the last step, we need to extrapolate our definition of g from within the small ball $\mathrm{B}(\mathbf{0}, r)$ to the bigger ball $\mathrm{B}(\mathbf{0}, R)$, within which the underlying distribution is concentrated. Again, using properties of Chebyshev polynomials, we show that if g is pointwise η-close to a degree-d polynomial in some ball $\mathrm{B}(\mathbf{0}, r')$[6], then its extrapolation is pointwise $(O(R/r'))^d\eta$ pointwise close to a degree-d polynomial in $\mathrm{B}(\mathbf{0}, R)$. After this, the rest of the analysis mirrors the one for the exact case.

Exactly Testing Polynomials over Discrete Domains. For Theorem 1.3, the main complication is that we can no longer evaluate points (even approximately!) on points drawn from $\mathcal{N}(0, I)$. We crucially relied on properties of the Gaussian (e.g., it is stable) for showing the self-correction properties of g in the above results. Instead here, we sample from *discrete Gaussian* distributions on lattices in order to define the self-corrected function g. Discrete Gaussians are a fundamental object of study in lattice cryptography (see, e.g., [MR07, Reg09]). Ours seems to be the first application of discrete Gaussians in a property testing setting.

For a lattice $\mathcal{L}$, the discrete Gaussian $\mathcal{G}(\mathcal{L}, s)$ is proportional to the density function of $\mathcal{N}(0, s)$ on the lattice points. The self-corrected function g is defined as $\mathsf{maj}_{\boldsymbol{q} \sim \mathcal{G}(\mathcal{L}', 1)}[g_{\boldsymbol{q}}(\boldsymbol{p})]$, where $g_{\boldsymbol{q}}$ is the same as in the exact testing analysis over $\mathbb{R}^n$. We perform the same three-step analysis here as above. For the first step, in order to show that g satisfies the degree-d characterization over lattice points, we derive explicit bounds on the TV distance between discrete Gaussians that were implicit in previous literature. For the second step, we follow the argument in the exact case, but we need to ensure that the lattice is large enough so that a nonzero low-degree polynomial is nonzero on at least one lattice point. Finally, in the third step, we extrapolate g from its self-corrected values on lattice points of $\mathcal{L}'$ inside a small ball $\mathrm{B}(\mathbf{0}, r)$ to lattice points of $\mathcal{L}$ on which $\mathcal{D}$ is supported. By the concentration property of $\mathcal{D}$ and from taking $\mathcal{L}'$ fine enough, we can find $d + 1$ lattice points of $\mathcal{L}'$ on any line from the origin to a point in $\mathcal{L} \cap \mathrm{B}(\mathbf{0}, R)$. This suffices for the extrapolation and the rest of the analysis.

1.4 Further Remarks

We leave the question of improving the bounds for the query complexity and the other parameters in Theorem 1.2 and Theorem 1.3 as interesting open problems. Also, it would be very interesting to obtain a separation between the complexities of the exact and approximate testing problems, in terms of query complexity. For the case of $d = 1$, we have an improved analysis that appears in Appendix D.

It is also natural to ask about *tolerant testing* [PRR06] in our setting. This is distinct from approximate testing, because in the completeness case, the function is only required to equal a degree-d polynomial P

[5] Note that (scaled) Chebyshev polynomials are bounded by $2^{1-d}\eta$ within $[-1, 1]$. We leave it open whether the lower bound of $2^{-O(d^2)}$ can be improved to $2^{-O(d)}$. However, for our application, this improvement would not be significant.

[6] In our analysis, we have, and choose r' that are strictly smaller than r.

with some probability over the distribution $\mathcal{D}$ which may be less than 1. Our test should still work under an appropriate choice of parameters, because by the union bound, we can upper bound the probability that one of the queries does not come from P.

Our work also opens the way for investigating the testability of other multivariate functional equations. Is there a general theory that characterizes testability (under natural assumptions) just as there is for finite fields [KS08, BFH+13]?

1.5 Organization

In the following section we discuss some preliminaries used for the exact testing. In Section 3 we give the full proof for the existence of an exact tester for low-degree polynomials, proving Theorem 1.1. Section 4 is devoted to proving Theorem 1.2, giving the approximate tester, wherein in Section 4.1, we give more preliminaries needed for the approximate tester. And Section 5 contains the tester for discrete domains, as specified in Theorem 1.3, wherein in Section 5.1 we give some more preliminaries needed for the discrete case. As the exact tester is the starting point for the other settings, in the later sections we rely on the proofs from Section 3, and show what changes need to be done. Finally, in Appendix D, we prove a sub-case of the approximate tester, where $d = 1$ and show a better result. While in the other appendices we show full proofs of some intermediate lemmata/theorems, that we skipped in the paper for the convenience of the reader.

2 Preliminaries

Here we record some notations and definitions which will be used throughout the paper. For a positive integer n, let $[n] = \{1, 2, \ldots, n\}$. We will reserve non-boldface symbols (such as $a \in \mathbb{R}$) to represent variables and scalars, and we will use boldface (such as $\boldsymbol{a} \in \mathbb{R}^n$) to represent vectors.

For any $S \subseteq \mathbb{R}^n$, we say that $f\colon \mathbb{R}^n \to \mathbb{R}$ is a *polynomial over* S if there exists a degree-d polynomial $g : \mathbb{R}^n \to \mathbb{R}$ such that $f(\boldsymbol{x}) = g(\boldsymbol{x})$ for every $\boldsymbol{x} \in S$. A *line* is a polynomial of the form $\boldsymbol{a} + i\boldsymbol{b}$, where i is a variable, and we will denote by $\mathrm{L}_{\boldsymbol{a},\boldsymbol{b}} \triangleq \{\boldsymbol{a} + i\boldsymbol{b} : i \in \mathbb{R}\}$, the set of points on this line. A *radial line* is a line that passes through the origin; that is, a line of the form $\mathrm{L}_{\boldsymbol{0},\boldsymbol{b}}$ for some $\boldsymbol{b} \in \mathbb{R}^n$. Throughout this paper, it will be convenient to talk about functions restricted to lines. For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$, let $f_{\boldsymbol{a},\boldsymbol{b}} : \mathbb{R} \to \mathbb{R}$ be defined as the restriction of f on $\mathrm{L}_{\boldsymbol{a},\boldsymbol{b}}$, i.e., $f_{\boldsymbol{a},\boldsymbol{b}}(x) = f(\boldsymbol{a} + x\boldsymbol{b})$.

Local Characterization of Degree-d Polynomials. In order to test whether a univariate function f is consistent with a degree-d polynomial, we will use a characterization of degree-d polynomials which is more amenable to this task. This characterization involves inspecting the *finite forward differences* of f, defined as

$$\Delta_h[f](x) \triangleq f(x + h) - f(x), \tag{2}$$

for $h \in \mathbb{R}$. This difference is a linear operator, i.e., for functions f, and g,

$$\Delta_h[f + g](x) = \Delta_h[f](x) + \Delta_h[g](x). \tag{3}$$

Higher order finite forward differences are defined inductively as,

$$\Delta_h^{(m)}[f](x) \triangleq \Delta_h\left[\Delta_h^{(m-1)}[f]\right](x) = (-1)^{m+1} \sum_{i=0}^{m} \alpha_i \cdot f(x + ih), \tag{4}$$

where $\alpha_i \triangleq (-1)^{i+1}\binom{m}{i}$, $m \in \mathbb{Z}_{>1}$, and $\Delta_h^{(1)} = \Delta_h$. Finite forward differences are related to the standard notion of a derivative, and we explain this further in Appendix A.

We will use the following characterization of degree-d polynomials, that follow from well-known results in analysis (see Appendix A for details).

Local Characterization Theorem. *Let $a, b \in \mathbb{R}$ such that $a < b$, and let $g\colon (a,b) \to \mathbb{R}$ be a univariate, bounded function. If for every $x \in (a,b)$ and sufficiently small $h > 0$, such that $a < x < x + (d+1)h < b$, $\Delta_h^{(d+1)}[g](x) = 0$, then g is a degree-d polynomial.*

A discrete variant of this theorem, given in Section 5, will be used for our lattice-based tester.

Sampling from Gaussian Distributions. In order to test that the local characterization holds, we will sample points from the $\boldsymbol{p} \sim \mathcal{N}(\mathbf{0}, \beta I)$ for various values of β. This is possible, given sampling access to $\mathcal{N}(\mathbf{0}, I)$, by multiplying sampled vectors with the respective $\sqrt{\beta}$'s, since $\sqrt{\beta}\boldsymbol{v} \sim \mathcal{N}(\mathbf{0}, \beta I)$, if $\boldsymbol{v} \sim \mathcal{N}(\mathbf{0}, I)$.

In order to generalize our tester to distributions that need not be centered at the origin, but say, at $\boldsymbol{c} \in \mathbb{R}^n$, we can test the local characterization at points sampled from Gaussians that are centered at such $\boldsymbol{c}$'s. This again is possible, given sampling access to $\mathcal{N}(\mathbf{0}, I)$, by translating the sampled vectors by the respective $\boldsymbol{c}$'s, since $\boldsymbol{c} + \boldsymbol{v} \sim \mathcal{N}(\boldsymbol{c}, I)$, if $\boldsymbol{v} \sim \mathcal{N}(\mathbf{0}, I)$.

Throughout this work, we will need to relate points sampled from different Gaussian distributions. For two distributions $\mathcal{D}$ and $\mathcal{D}'$ on the same domain Ω, the total variation distance between them is defined as

$$\mathrm{d_{TV}}(\mathcal{D}, \mathcal{D}') \triangleq \frac{1}{2}\int_{\Omega} |\mathcal{D}(x) - \mathcal{D}'(x)| dx.$$

We will use the following lemma (a proof can be found in [FY20]) to bound the total variation distance between two Gaussian distributions. Let $\|\cdot\|_2$ denote the operator norm on matrices.

Lemma 2.1. *Consider two Gaussian distributions $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ with shared invertible covariance matrices $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$. Then $\mathrm{d_{TV}}(\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})) \le \phi$ holds, if $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2 \le 2\phi/\sqrt{\|\boldsymbol{\Sigma}^{-1}\|_2}$.*

An immediate corollary of Lemma 2.1 is the following:

Lemma 2.2. *For any integer $k \ge 1$, real $r > 0$, and $\boldsymbol{p} \in \mathbb{R}^n$, such that $\|\boldsymbol{p}\|_2 \le r$, it follows that $\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0}, kI), \mathcal{N}(\boldsymbol{p}, kI)) \le kr/2$.*

Proof. Observe that the spectral norm of kI is k, and therefore $\|\boldsymbol{p}\|_2\sqrt{\|kI\|_2} \le kr$. It follows from Lemma 2.1, that $\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0}, kI), \mathcal{N}(\boldsymbol{p}, kI)) \le kr/2$. □

3 Exact Testing

In this section, we develop a distribution-free tester for low-degree polynomials over the $\mathbb{R}^n$, assuming that we can exactly query the input function. Our tester is given in Algorithm 1 and uses the subroutines given in Algorithm 2. The CHARACTERIZATIONTEST checks properties of f which will be sufficient to guarantee that g — the *self-corrected* version of f — is a degree-d polynomial. QUERY-g retrieves the value of $g(\boldsymbol{p})$ for a given point $\boldsymbol{p}$ by running the subroutine QUERY-g-INBALL, which in turn obtains the values of g on points within the small ball $\mathrm{B}(\mathbf{0}, r)$ by evaluating f.

Recall that in Theorem 1.1, f is assumed to be bounded in $\mathrm{B}(\mathbf{0}, L)$, for some $L > 0$. Throughout this section, we assume $L = 2d\sqrt{n}$. This is without loss of generality as we can define $f' : \mathbb{R}^n \to \mathbb{R}$ as $f'(\boldsymbol{x}) = f(\boldsymbol{x}L/(2d\sqrt{n}))$ which is bounded in $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, and the tester can query f' via queries to f. If f is a degree-d polynomial, so is f'. If f is ε-far from degree-d polynomials over a distribution $\mathcal{D}$, so is f' over the distribution $\mathcal{D}'$, where a sample $\boldsymbol{y} \sim \mathcal{D}'$ is generated as $\boldsymbol{y} = \frac{2d\sqrt{n}}{L}\boldsymbol{x}$ where $\boldsymbol{x} \sim \mathcal{D}$.

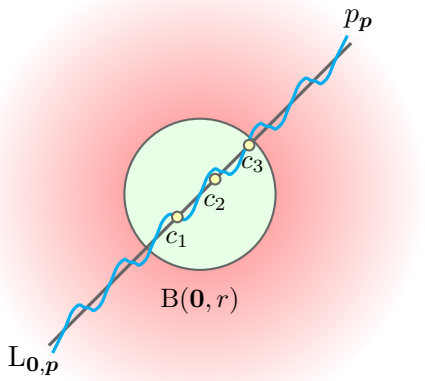

Figure 4: The definition of $g(\boldsymbol{p})$ for $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$. First, a degree-$d$ univariate polynomial $p_{\boldsymbol{p}}$ is defined by the value of g on $d+1$ points $c_i \in \mathrm{B}(\mathbf{0}, r)$ on the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$, such that $p_{\boldsymbol{p}}(c_i) = g(\boldsymbol{p}c_i)$. Then, the value of $g(\boldsymbol{p})$ is defined to be $p_{\boldsymbol{p}}(1)$.

The Self-Corrected Function. As outlined in Section 1.3, by sampling points from the standard Gaussian, we will construct a self-corrected version g of the input function f such that, if our tests (in particular CHARACTERIZATIONTEST in Algorithm 2) pass with sufficiently high probability, then we can guarantee that g is a degree-d polynomial. Let $r = (3d)^{-6}$, and $\mathrm{B}(\mathbf{0}, r)$ be the *open* ball of radius r, centered at the origin. We will guarantee that g is a degree-d polynomial for points $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ first, and then extended the characterization to points outside of this ball. The advantage of restricting our attention to this small ball is that for any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, $\boldsymbol{p} + \boldsymbol{x}$ is approximately distributed as $\boldsymbol{x}$.

We define $g \colon \mathbb{R}^n \to \mathbb{R}$ formally as follows: let $\alpha_i = (-1)^{i+1}\binom{d+1}{i}$, and for any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ and $\boldsymbol{q} \in \mathbb{R}^n$, $g_{\boldsymbol{q}}(\boldsymbol{p}) \triangleq \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})$. The intuition behind $g_{\boldsymbol{q}}(\boldsymbol{p})$ is that it is the value of the univariate, degree-d polynomial at the point $\boldsymbol{p}$, that is uniquely defined by the $d+1$ evaluations $\{f(\boldsymbol{p} + i\boldsymbol{q}) : i \in [d+1]\}$. For points $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, we define the value of g to be

$$g(\boldsymbol{p}) \triangleq \operatorname*{maj}_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)} [g_{\boldsymbol{q}}(\boldsymbol{p})] .$$

For points $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$, we define $g(\boldsymbol{p})$ by interpolating the evaluations of g on points within $\mathrm{B}(\mathbf{0}, r)$ as follows (see Figure 4). Consider the radial line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}} = \{x\boldsymbol{p} : x \in \mathbb{R}\}$ and fix $d+1$ (arbitrary) "distinguished" points along this line $c_0, \ldots, c_d \in \mathbb{R}$ such that $c_i\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ for all i; in Algorithm 2 we choose $c_i = ir/((d+1)\|\boldsymbol{p}|_2)$. Let $p_{\boldsymbol{p}} \colon \mathbb{R}^n \to \mathbb{R}$ be the degree-d, univariate polynomial uniquely defined by these $d+1$ points, such that $p_{\boldsymbol{p}}(c_i) = g(c_i\boldsymbol{p})$, for every $i \in [d+1]$. The value of $g(\boldsymbol{p})$ is defined as $p_{\boldsymbol{p}}(1)$. Note that if g was a degree-d polynomial to begin with, then we would indeed have $p_{\boldsymbol{p}}(1) = g(\boldsymbol{p})$.

The following lemma records the properties of g that will be guaranteed by our tester.

Lemma 3.1. *If* CHARACTERIZATIONTEST *fails with probability at most* $2/3$*, then* g *is a degree-*d *polynomial, and furthermore for any* $\boldsymbol{p} \in \mathbb{R}^n$*,* $g(\boldsymbol{p}) =$ QUERY-$g(\boldsymbol{p})$ *with probability at least* $1 - \frac{\varepsilon}{2}$*.*

We prove the main theorem of this section assuming Lemma 3.1 holds; we restate it next for convenience.

Theorem 1.1. *Let* $d \in \mathbb{N}$*, and for* $L > 0$*, suppose* $f : \mathbb{R}^n \to \mathbb{R}$ *is a function that is bounded in the ball* $\mathrm{B}(\mathbf{0}, L)$*. Given* $\varepsilon > 0$*, query access to* f*, and sampling access to an unknown distribution* $\mathcal{D}$*, there*

Algorithm 1: Distribution-Free Low-Degree Tester

Procedure LOWDEGREETESTER($f, d, \mathcal{D}, \varepsilon$)
  Given : Query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, degree $d \in \mathbb{N}$, sampling access to an unknown distribution $\mathcal{D}$, and farness parameter $\varepsilon > 0$.
  Reject if CHARACTERIZATIONTEST **rejects**;
  $N_1 \leftarrow O(\varepsilon^{-1})$;
  for N_1 *times* **do**
    Sample $\boldsymbol{p} \sim \mathcal{D}$;
    Reject if $f(\boldsymbol{p}) \neq$ QUERY-$g(\boldsymbol{p})$ or if QUERY-$g(\boldsymbol{p})$ **rejects**.
  Accept.

exists a one-sided error, distribution-free, $O(d^5 + \frac{d^2}{\varepsilon}\log\frac{1}{\varepsilon})$-query tester for testing whether f is a degree-d polynomial, or is ε-far from degree-d polynomials over $\mathcal{D}$.

Proof of Theorem 1.1. First we analyze the query complexity. CHARACTERIZATIONTEST performs $O(d^2)$ independent tests, each of which requires $O(d)$ evaluations of f, and is repeated $N_2 = O(d^2)$ times. QUERY-g-INBALL samples $N_2' = O(\log(1/\varepsilon))$ points, each requiring $O(d)$ evaluations of f. QUERY-g picks $O(d)$ points in $\mathrm{B}(\mathbf{0}, r)$ and calls QUERY-g-INBALL on them. LOWDEGREETESTER calls CHARACTERIZATIONTEST once, and then calls QUERY-g, $N_1 = O(1/\varepsilon)$ times. Altogether, our algorithm makes $O(d^5 + \frac{d^2}{\varepsilon}\log(\frac{1}{\varepsilon}))$ queries.

Next, we argue that the tester is correct. If f is a degree-d polynomial, then it accepts with probability 1. Indeed, in this case f restricted to a line $\boldsymbol{p} + i\boldsymbol{q}$ is also a degree-d polynomial, $g = f$, and all of the tests pass with probability 1.

Now, assume that f is ε-far from any degree-d polynomial (according to $\mathcal{D}$). If CHARACTERIZATIONTEST fails with probability at least $2/3$, then we reject with probability at least $2/3$. Otherwise, by Lemma 3.1, g is a degree-d polynomial and so $\Pr_{\boldsymbol{p}\sim\mathcal{D}}[f(\boldsymbol{p}) \neq g(\boldsymbol{p})] > \varepsilon$. The probability that we do not reject in any of the N_1 steps of Algorithm 1 is at most the probability that $f(\boldsymbol{p}) = g(\boldsymbol{p})$ or that QUERY-$g(\boldsymbol{p})$, instead of rejecting, returned some value other than $g(\boldsymbol{p})$. The latter happens with probability at most $\frac{\varepsilon}{2}$ by Lemma 3.1, and so

$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[f(\boldsymbol{p}) = g(\boldsymbol{p}) \vee g(\boldsymbol{p}) \neq \text{QUERY-}g(\boldsymbol{p})] \leq 1 - \varepsilon + \frac{\varepsilon}{2} \leq 1 - \frac{\varepsilon}{2}.$$

Thus, Algorithm 1 accepts with probability at most $(1 - \frac{\varepsilon}{2})^{N_1} < \frac{1}{3}$, by choosing the constant in $N_1 = O(\varepsilon^{-1})$ to be sufficiently large. □

In the remainder of this section we will prove Lemma 3.1. First, in Section 3.1, we show that g agrees with a degree-d, univariate polynomial on every line segment in $\mathrm{B}(\mathbf{0}, r)$. Then, we show that g is consistent with a degree-d, n-variate polynomial within $\mathrm{B}(\mathbf{0}, r)$. Finally, by the fact that for points outside $\mathrm{B}(\mathbf{0}, r)$, g is defined by interpolating evaluations out of $\mathrm{B}(\mathbf{0}, r)$, we show that it is a degree-d, n-variate polynomial on $\mathbb{R}^n$.

3.1 Polynomial Representation on Every Line Within the Ball

We will prove that if CHARACTERIZATIONTEST passes with high probability, then g is consistent with a degree-d polynomial when projected to any line segment that lies within the open ball $\mathrm{B}(\mathbf{0}, r)$ for $r = (3d)^{-6}$.

Algorithm 2: Subroutines

[Recall $\alpha_i \triangleq (-1)^{i+1}\binom{d+1}{i}$.]
Procedure CHARACTERIZATIONTEST
  $N_2 \leftarrow O(d^2)$;
  for N_2 *times* **do**
    for $j \in \{1, \ldots, d+1\}$ **do**
      for $t \in \{0, \ldots, d+1\}$ **do**
        Sample $\boldsymbol{p} \sim \mathcal{N}(\mathbf{0}, j^2(t^2+1)I), \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)$; ▷ [$j^2(t^2+1)$ vs. 1 Test.]
        Reject if $\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0$;
        Sample $\boldsymbol{p} \sim \mathcal{N}(\mathbf{0}, j^2 I), \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I)$; ▷ [$j^2$ vs. t^2+1 Test.]
        Reject if $\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0$;
      Sample $\boldsymbol{p}, \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, j^2 I)$; ▷ [$j^2$ vs. j^2 Test.]
      Reject if $\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0$;
  Accept;
Procedure QUERY-$g(\boldsymbol{p})$
  $r \leftarrow (3d)^{-6}$;
  if $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ **then**
    return QUERY-g-INBALL$(\boldsymbol{p})$;
  for $i \in \{1, \ldots, d+1\}$ **do**
    $c_i \leftarrow ir/((d+1)\|\boldsymbol{p}\|_2)$;
    $v(c_i) \leftarrow$ QUERY-g-INBALL$(c_i\boldsymbol{p})$;
  Let $p_{\boldsymbol{p}} \colon \mathbb{R} \to \mathbb{R}$ be the unique degree-d polynomial such that $p_{\boldsymbol{p}}(c_i) = v(c_i)$ for $i \in [d+1]$;
  return $p_{\boldsymbol{p}}(1)$;
Procedure QUERY-g-INBALL$(\boldsymbol{p})$
  $N_2' \leftarrow O(\log \frac{1}{\varepsilon})$;
  Sample $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N_2'} \sim \mathcal{N}(\mathbf{0}, I)$;
  Reject if there exists $j \in \{2, \ldots, N_2'\}$ such that $\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1) \neq \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_j)$;
  return $\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1)$;

For $\boldsymbol{a}, \boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)$, we will denote by $\mathrm{L}^{\mathrm{B}}_{\boldsymbol{a},\boldsymbol{b}}$ the line segment obtained by restricting the line $\mathrm{L}_{\boldsymbol{a},\boldsymbol{b}}$ to the ball $\mathrm{B}(\mathbf{0}, r)$. The main theorem of this section states that evaluations of g on every point on any line segment within the open ball $\mathrm{B}(\mathbf{0}, r)$, are consistent with a unique, univariate, degree-d polynomial.

Theorem 3.2. *(Polynomial Representation on Lines) If* CHARACTERIZATIONTEST *fails with probability at most* $2/3$*, and* f *is bounded on* $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$*, then for every* $\boldsymbol{a}, \boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)$*, the univariate function* $g_{\boldsymbol{a},\boldsymbol{b}}(x) = g(\boldsymbol{a} + x\boldsymbol{b})$ *defined on points* $x \in \mathrm{L}^{\mathrm{B}}_{\boldsymbol{a},\boldsymbol{b}}$ *is a degree-*d*, univariate polynomial.*

In order to prove this theorem we will need the following auxiliary lemmas.

Lemma 3.3. *If* CHARACTERIZATIONTEST *fails with probability at most* $2/3$*, then for every* $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$*, for all sufficiently small* $h > 0$*, such that* $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ *for every* $i \in [d+1]$*,* $\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q}) = 0$.

Lemma 3.4. *If* f *is bounded on* $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$*, then* g *is bounded on* $\mathrm{B}(\mathbf{0}, r)$.

We prove Theorem 3.2 assuming these lemmas, and prove them afterwards.

Proof of Theorem 3.2. Since f is bounded on $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, by Lemma 3.4, g is bounded on $\mathrm{B}(\mathbf{0}, r)$. Fix some $\boldsymbol{a}, \boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)$. We would like to show that $g_{\boldsymbol{a},\boldsymbol{b}}(x)$ is consistent with a degree-d polynomial on every point x in $\{x \in \mathbb{R} : \boldsymbol{a} + x\boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)\}$; fix such an x. By the Local Characterization Theorem, it suffices to show that for all sufficiently small $h > 0$, satisfying $\boldsymbol{a} + (x + ih)\boldsymbol{b} \in \mathrm{L}^{\mathrm{B}}_{\boldsymbol{a},\boldsymbol{b}}$ for every $i \in [d+1]$,

$$\Delta_h^{(d+1)}[g_{\boldsymbol{a},\boldsymbol{b}}](x) = \sum_{i=0}^{d+1} \alpha_i \cdot g_{\boldsymbol{a},\boldsymbol{b}}(x + ih) = \sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{a} + x\boldsymbol{b} + ih\boldsymbol{b}) = 0.$$

From Lemma 3.3, it follows that $\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q}) = 0$ for every $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ and all sufficiently small $h > 0$, satisfying $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$. Let $\boldsymbol{p} \triangleq \boldsymbol{a} + x\boldsymbol{b}$ and $\boldsymbol{q} \triangleq \boldsymbol{b}$. Observe that $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$, and therefore since $\mathrm{B}(\mathbf{0}, r)$ is an open ball, $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$. Thus,

$$\Delta_h^{(d+1)}[g_{\boldsymbol{a},\boldsymbol{b}}](x) = \sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{a} + x\boldsymbol{b} + ih\boldsymbol{b}) = \sum_{i=0}^{d+1} g(\boldsymbol{p} + ih\boldsymbol{q}) = 0. \qquad \square$$

In the remainder of this subsection we prove Lemma 3.3 and Lemma 3.4. For this, it will be convenient to let ρ denote the *smallest* upper-bound on the probability that each of the tests in the CHARACTERIZATIONTEST failed. That is, for every $j \in [d+1]$ and $t \in \{0, \ldots, d+1\}$, ρ is the smallest value such that

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0}, j^2(t^2+1)I) \\ \boldsymbol{q}\sim\mathcal{N}(\mathbf{0}, I)}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0\right] \leq \rho, \qquad [j^2(t^2+1) \text{ vs. } 1 \text{ Test.}] \tag{5}$$

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0}, j^2 I) \\ \boldsymbol{q}\sim\mathcal{N}(\mathbf{0}, (t^2+1)I)}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0\right] \leq \rho, \qquad [j^2 \text{ vs. } t^2+1 \text{ Test.}] \tag{6}$$

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0}, j^2 I) \\ \boldsymbol{q}\sim\mathcal{N}(\mathbf{0}, j^2 I)}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0\right] \leq \rho. \qquad [j^2 \text{ vs. } j^2 \text{ Test.}] \tag{7}$$

A bound on the rejection probability of CHARACTERIZATIONTEST implies the following bound on ρ.

Claim 3.5. If CHARACTERIZATIONTEST fails with probability at most $2/3$, then ρ is at most $(30d)^{-2}$.

Proof. Each of the tests (5), (6) and (7) are invoked $N_2 = O(d^2)$ in the CHARACTERIZATIONTEST. If any of these tests fail with probability more than $1/(30d)^2$, then CHARACTERIZATIONTEST passes with probability at most $(1 - \frac{1}{(30d)^2})^{O(d^2)} < 1/3$, which contradicts our assumption. $\square$

The proof of Lemma 3.3 will heavily rely on the fact that if ρ is small then $g_{\boldsymbol{q}_1}$ and $g_{\boldsymbol{q}_2}$ agree on points in $\mathrm{B}(\mathbf{0}, r)$ with high probability.

Lemma 3.6. *For every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, and every $t \in \{0, \ldots, d+1\}$,*

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0}, (t^2+1)I) \\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0}, I)}} [g_{\boldsymbol{q}_1}(\boldsymbol{p}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{p})] \leq 4d\rho + 48d^5 r.$$

Proof. Let $t \in \{0, \dots, d+1\}$ and fix some $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$. We will bound the probability that $g_{\boldsymbol{q}_1}(\boldsymbol{p})$ and $q_{\boldsymbol{q}_2}(\boldsymbol{p})$ are different from $\sum_{i=1}^{d+1}\sum_{j=1}^{d+1}\alpha_i\alpha_j \cdot f(\boldsymbol{p}+i\boldsymbol{q}_1+j\boldsymbol{q}_2)$; the lemma will then follow by a union bound.

By definition, $g_{\boldsymbol{q}_2}(\boldsymbol{p}) = \sum_{i=1}^{d+1}\alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q}_2)$. Fixing $i \in [d+1]$, we have

$$\begin{aligned}
&\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}[f(\underbrace{\boldsymbol{p}+i\boldsymbol{q}_1}_{\triangleq\boldsymbol{m}}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{p}+i\boldsymbol{q}_1)] = \Pr_{\substack{\boldsymbol{m}\sim\mathcal{N}(\boldsymbol{p},i^2(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\Big[f(\boldsymbol{m}) \neq \sum_{j=1}^{d+1}\alpha_j \cdot f(\boldsymbol{m}+j\boldsymbol{q}_2)\Big]\\
&\leq \Pr_{\substack{\boldsymbol{m}\sim\mathcal{N}(\mathbf{0},i^2(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\Big[\sum_{j=0}^{d+1}\alpha_j \cdot f(\boldsymbol{m}+j\boldsymbol{q}_2) \neq 0\Big] + 2\,\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0}, i^2(t^2+1)I), \mathcal{N}(\boldsymbol{p}, i^2(t^2+1)I))\\
&\leq \rho + i^2(t^2+1)r \leq \rho + 20d^4r. \qquad \text{(By (5) and Lemma 2.2)}
\end{aligned}$$

By a similar calculation, for every $j \in [d+1]$, we have that

$$\begin{aligned}
&\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}[f(\underbrace{\boldsymbol{p}+j\boldsymbol{q}_2}_{\triangleq\boldsymbol{m}}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{p}+j\boldsymbol{q}_2)] = \Pr_{\substack{\boldsymbol{m}\sim\mathcal{N}(\boldsymbol{p},j^2I)\\ \boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)}}\Big[f(\boldsymbol{m}) \neq \sum_{i=1}^{d+1}\alpha_i \cdot f(\boldsymbol{m}+i\boldsymbol{q}_1)\Big]\\
&\leq \Pr_{\substack{\boldsymbol{m}\sim\mathcal{N}(\mathbf{0},j^2I)\\ \boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)}}\left[\sum_{i=0}^{d+1}\alpha_1 \cdot f(\boldsymbol{m}+i\boldsymbol{q}_1) \neq 0\right] + 2\,\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0}, j^2I), \mathcal{N}(\boldsymbol{p}, j^2I))\\
&\leq \rho + j^2r \leq \rho + 4d^2r. \qquad \text{(By (6) and Lemma 2.2)}
\end{aligned}$$

Taking a union bound over $i \in [d+1]$ and $j \in [d+1]$ respectively, it follows that

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[\underbrace{\sum_{i=1}^{d+1}\alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q}_1)}_{=g_{\boldsymbol{q}_1}(\boldsymbol{p})} \neq \sum_{i=1}^{d+1}\sum_{j=1}^{d+1}\alpha_i\alpha_j \cdot f((\boldsymbol{p}+i\boldsymbol{q}_1)+j\boldsymbol{q}_2)\right] \leq (d+1)(\rho+20d^4r),$$

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[\underbrace{\sum_{j=1}^{d+1}\alpha_j \cdot f(\boldsymbol{p}+j\boldsymbol{q}_2)}_{=g_{\boldsymbol{q}_2}(\boldsymbol{p})} \neq \sum_{j=1}^{d+1}\sum_{i=1}^{d+1}\alpha_i\alpha_j \cdot f((\boldsymbol{p}+j\boldsymbol{q}_2)+i\boldsymbol{q}_1)\right] \leq (d+1)(\rho+4d^2r).$$

The first inequality is at most $2d\rho + 40d^5r$, while the second is at most $2d\rho + 8d^3r$. Thus, by a union bound over the two previous inequalities we can conclude that

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}[g_{\boldsymbol{q}_1}(\boldsymbol{p}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{p})] \leq 4d\rho + 48d^5r. \qquad \square$$

The next corollary follows immediately by instantiating the parameters in the previous lemma.

Corollary 3.7. *If* CHARACTERIZATIONTEST *fails with probability at most* $2/3$*, then for every* $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ *and every* $t \in \{0, \dots, d+1\}$*,*

$$\Pr_{\boldsymbol{q}\sim\mathcal{N}(\mathbf{0},(t^2+1)I)}[g(\boldsymbol{p}) \neq g_{\boldsymbol{q}}(\boldsymbol{p})] \leq \frac{1}{7d}.$$

Proof. Observe that for any $t \in \{0, \dots, d+1\}$, and any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$,

$$\Pr_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I)}[g(\boldsymbol{p}) \neq g_{\boldsymbol{q}}(\boldsymbol{p})] \leq \Pr_{\boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, I)}[g(\boldsymbol{p}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{p})] + \Pr_{\substack{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I) \\ \boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, I)}}[g_{\boldsymbol{q}_1}(\boldsymbol{p}) \neq g_{\boldsymbol{q}}(\boldsymbol{p})].$$

By Lemma 3.6, this is at most $2(4d\rho + 48d^5 r)$, where for the first term, we have used the fact that $g_{\boldsymbol{q}_1}(\boldsymbol{p})$ is defined as the majority of $\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)$. By Claim 3.5 and by our choice of $r = (3d)^{-6}$, this probability is at most $1/(7d)$. □

We are now ready to prove Lemma 3.3.

Proof of Lemma 3.3. Fix $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$, and let $h > 0$ be such that $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$; note that h exists as $\mathrm{B}(\mathbf{0}, r)$ is an *open* ball containing $\boldsymbol{p}$. We will argue that the following hold simultaneously with non-zero probability over $\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)$:

$$\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q}) = \sum_{i=0}^{d+1} \alpha_i \cdot g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + ih\boldsymbol{q}), \tag{8}$$

$$\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + j\boldsymbol{q}_1 + i(h\boldsymbol{q} + j\boldsymbol{q}_2)) = 0, \text{ for every } j \in [d+1]. \tag{9}$$

We will complete the proof assuming that these bounds hold. Fix any $\boldsymbol{q}_1, \boldsymbol{q}_2$ satisfying both (8), and (9). Then,

$$\begin{aligned}
\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q}) &= \sum_{i=0}^{d+1} \alpha_i \cdot g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + ih\boldsymbol{q}) && \text{(By (8))} \\
&= \sum_{i=0}^{d+1} \alpha_i \left(\sum_{j=1}^{d+1} \alpha_j \cdot f(\boldsymbol{p} + ih\boldsymbol{q} + j(\boldsymbol{q}_1 + i\boldsymbol{q}_2)) \right) \\
&= \sum_{j=1}^{d+1} \alpha_j \left(\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + j\boldsymbol{q}_1 + i(h\boldsymbol{q} + j\boldsymbol{q}_2)) \right) \\
&= \sum_{j=1}^{d+1} \alpha_j \cdot 0 = 0. && \text{(By (9))}
\end{aligned}$$

Next, we argue that (8) and (9) hold separately with sufficiently high probability.

$$\begin{aligned}
\Pr_{\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)}[(8)] &= \Pr_{\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)} \left[\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q}) = \sum_{i=0}^{d+1} \alpha_i \cdot g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + ih\boldsymbol{q}) \right] \\
&\geq \Pr_{\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)} [g(\boldsymbol{p} + ih\boldsymbol{q}) = g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + ih\boldsymbol{q}), \ \ \forall i \in \{0, \dots, d+1\}] \\
&= 1 - \Pr_{\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)} [\exists i \in \{0, \dots, d+1\} : g(\boldsymbol{p} + ih\boldsymbol{q}) \neq g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + ih\boldsymbol{q})] \\
&\geq 1 - \sum_{i=0}^{d+1} \Pr_{\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)} [g(\boldsymbol{p} + ih\boldsymbol{q}) \neq g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + ih\boldsymbol{q})] && \text{(By union bound)}
\end{aligned}$$

$$= 1 - \sum_{i=0}^{d+1} \Pr_{\boldsymbol{m}\sim\mathcal{N}(\mathbf{0},(i^2+1)I)} [g(\boldsymbol{p}+ih\boldsymbol{q}) \neq g_{\boldsymbol{m}}(\boldsymbol{p}+ih\boldsymbol{q})] \qquad \text{(Letting } \boldsymbol{m} \triangleq \boldsymbol{q}_1 + i\boldsymbol{q}_2\text{)}$$

$$\geq 1 - \frac{d+2}{7d} > \frac{1}{2}. \qquad \text{(Applying Corollary 3.7, as } \boldsymbol{p}+ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0},r)\text{)}$$

For (9), fix some $j \in [d+1]$, then

$$\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\underbrace{\boldsymbol{p}+j\boldsymbol{q}_1}_{\triangleq \boldsymbol{z}_1} + i(\underbrace{h\boldsymbol{q}+j\boldsymbol{q}_2}_{\triangleq \boldsymbol{z}_2})) \neq 0\right] = \Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{N}(\boldsymbol{p},j^2I)\\ \boldsymbol{z}_2\sim\mathcal{N}(h\boldsymbol{q},j^2I)}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{z}_1+i\boldsymbol{z}_2) \neq 0\right]$$

$$\leq \Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{N}(\mathbf{0},j^2I)\\ \boldsymbol{z}_2\sim\mathcal{N}(\mathbf{0},j^2I)}} \left[\sum_{i=0}^{d+1} \alpha_i f(\boldsymbol{z}_1+i\boldsymbol{z}_2) \neq 0\right] + 2(\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0},j^2I),\mathcal{N}(\boldsymbol{p},j^2I)) + \mathrm{d_{TV}}(\mathcal{N}(\mathbf{0},j^2I),\mathcal{N}(h\boldsymbol{q},j^2I)))$$

$$\leq \rho + j^2 r + j^2 h r \leq \rho + 8d^2 r. \qquad \text{(By (7) and Lemma 2.2)}$$

By a union bound over all $j \in [d+1]$,

$$\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}[(9)] = \Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)} \left[\forall j \in [d+1], \sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p}+j\boldsymbol{q}_1+i(h\boldsymbol{q}+j\boldsymbol{q}_2)) = 0\right] \geq 1-(2d\rho+16d^3r),$$

which is at least $2/3$ by our choice of $r = (3d)^{-6}$ and Claim 3.5. A final union bound over (8), and (9) concludes that both hold simultaneously with non-zero probability. □

Finally, in order to conclude that g is indeed a polynomial by using Local Characterization Theorem on lines within $\mathrm{B}(\mathbf{0},r)$, we will argue that g is bounded in $\mathrm{B}(\mathbf{0},r)$

Proof of Lemma 3.4. It suffices to prove $g_{\boldsymbol{q}}(\boldsymbol{p})$ is bounded for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0},r)$, and every $\boldsymbol{q} \in \mathbb{R}^n$ such that $g_{\boldsymbol{q}}(\boldsymbol{p}) = g(\boldsymbol{p})$. By Corollary 3.7, $g(\boldsymbol{p}) = g_{\boldsymbol{q}}(\boldsymbol{p})$ with probability at least $1 - 1/7d$ for $\boldsymbol{q} \sim \mathcal{N}(\mathbf{0},I)$. By [BHK20, Theorem 2.9], at least 99% of the mass in $\mathcal{N}(\mathbf{0},I)$ lies in the annulus $|r - \sqrt{n}| \leq 20$. Therefore, we can conclude that $g(\boldsymbol{p})$ agrees with $g_{\boldsymbol{q}}(\boldsymbol{p})$ for $\boldsymbol{q}$ satisfying $|\|\boldsymbol{q}\|_2 - \sqrt{n}| \leq 20$. Note that $g_{\boldsymbol{q}}(\boldsymbol{p})$ depends only on $\{f(\boldsymbol{p}+i\boldsymbol{q})\}_{i=0}^{d+1}$, and $\max_i\{\|\boldsymbol{p}+i\boldsymbol{q}\|_2\} \leq (d+2)\max\{\|\boldsymbol{p}\|_2, \|\boldsymbol{q}\|_2\} \leq (d+2)(\sqrt{n}+20)$. Thus, if f is bounded on $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, then g is bounded on $\mathrm{B}(\mathbf{0},r)$. □

3.2 Polynomial Representation Within a Hypercube

Let $m \in \mathbb{R}$ be be the largest value, such that the hypercube $[-m,m]^n$ is strictly contained within the open ball $\mathrm{B}(\mathbf{0},r)$; in particular, $r/(2\sqrt{n}) \leq m < r/\sqrt{n}$. We show that if the conditions of Theorem 3.2 are met, then g is consistent with a degree-d multivariate polynomial on $[-m,m]^n$. This is done in two steps; first, in Lemma 3.8 we show that g is consistent with a finite bounded degree polynomial. Then, in Lemma 3.9, we show that this degree can be reduced to d.

Let $\boldsymbol{e}_i$ denote the ith standard basis vector, defined as $\boldsymbol{e}_{i,j} = 0$ if $j \neq i$ and $\boldsymbol{e}_{i,i} = 1$.

Lemma 3.8. *(Local to Global) Let $m > 0$, and let $h\colon [-m,m]^n \to \mathbb{R}$. If for every $i \in [n]$, and $\boldsymbol{a} \in [-m,m]^n$ such that $a_i = 0$, the restriction of h to the line segment $\mathrm{L}_{\boldsymbol{a},\boldsymbol{e}_i}$, the univariate function $h_{\boldsymbol{a},\boldsymbol{e}_i}\colon [-m,m] \to \mathbb{R}$ is consistent with a degree-d univariate polynomial on the interval $[-m,m]$, then h is consistent with an n-variate polynomial of degree at most dn.*

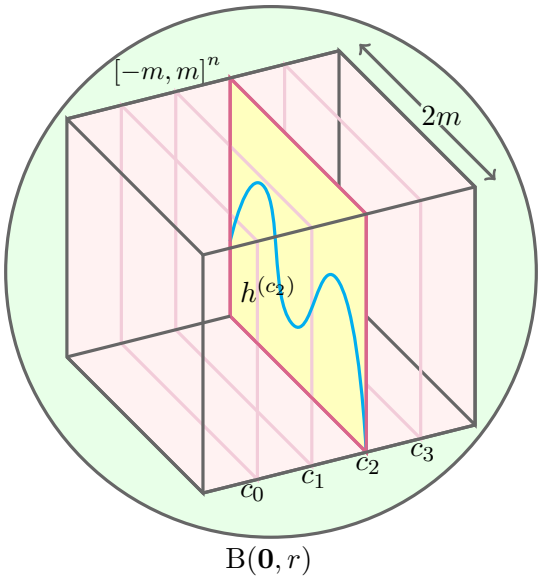

Figure 5: The construction of the polynomial $h(a_1,\ldots,a_{n-1})$. $d+1$ slices of the cube $[-m,m]^n$ are chosen, where the ith slice corresponds to setting $x_n = t = c_i$. The picture depicts setting $t = c_2$ and thus $\delta_{c_i}(t) = 0$ for all $i \neq 2$ and $\delta_{c_2}(t) = 1$, selecting the polynomial representation $h^{(c_2)}$ of h on the 2nd slice.

Proof. We will show that h is a degree-dn polynomial by induction on the dimension n. For the base case when $n = 1$, we have that $h = h_{\mathbf{0},1}$ and therefore is of degree d by assumption.

Assume the statement is true for dimension $n-1$. Let $c \in [-m,m]$ and define $h^{(c)} \colon [-m,m]^{n-1} \to \mathbb{R}$ as

$$h^{(c)}(x_1,\ldots,x_{n-1}) \triangleq h(x_1,\ldots,x_{n-1},c).$$

We will argue that $h^{(c)}$ satisfies the conditions of Lemma 3.8: Fix $i \in [n-1]$, and $\boldsymbol{a} \in [-m,m]^{n-1}$ with $a_i = 0$, and define $\boldsymbol{a}^{\uparrow} = (a_1,\ldots,a_{n-1},c) \in [-m,m]^n$ to be an extension of $\boldsymbol{a}$ to dimension n. By assumption, $h_{\boldsymbol{a}^{\uparrow},\boldsymbol{e}_i} \colon [-m,m] \to \mathbb{R}$ is a degree-d polynomial. For every $x \in [-m,m]$, we have

$$h_{\boldsymbol{a}^{\uparrow},\boldsymbol{e}_i}(x) = h(a_1,\ldots,a_{i-1},x,a_{i+1},\ldots,a_{n-1},c) = h^{(c)}(\boldsymbol{a}+x\boldsymbol{e_i}) = h^{(c)}_{\boldsymbol{a},\boldsymbol{e_i}}(x),$$

and so $h^{(c)}_{\boldsymbol{a},\boldsymbol{e_i}}(x)$ is a degree-d polynomial on the domain $[-m,m]$. Thus, by the inductive hypothesis we can conclude that $h^{(c)} \colon [-m,m]^{n-1} \to \mathbb{R}$ is a degree-$d(n-1)$ multivariate polynomial.

It remains to show that h is a degree-dn polynomial. Let $c_0, c_1, \ldots, c_d \in [-m,m]$ be any $d+1$ distinct values. Denote by δ_{c_i} the unique degree-d polynomial satisfying

$$\delta_{c_i}(c_j) \triangleq \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases} \tag{10}$$

Using δ_{c_i}, we will show that h can be written as a polynomial of degree dn.

First, we claim that for every fixed $\boldsymbol{a} \in [-m,m]^{n-1}$ and variable t,

$$h(a_1,\ldots,a_{n-1},t) = \sum_{i=0}^{d} \delta_{c_i}(t) h^{(c_i)}(a_1,\ldots,a_{n-1}).$$

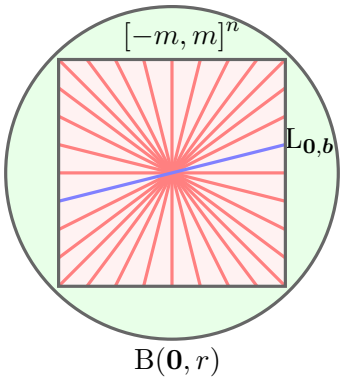

Figure 6: The radial lines $\mathrm{L}_{\mathbf{0},\boldsymbol{b}}$ within the hypercube $[-m,m]^n$ in two dimensions.

To see this, observe that $h^{(c_i)}(a_1,\ldots,a_{n-1})$ is a constant and therefore $\delta_{c_i}(t)h^{(c_i)}(a_1,\ldots,a_{n-1})$ is a degree-d polynomial. Thus, $\sum_{i=0}^{d}\delta_{c_i}(t)h^{(c_i)}(a_1,\ldots,a_{n-1})$ and $h(a_1,\ldots,a_{n-1},t)$ are degree-d polynomials (the latter is by assumption). Furthermore, these degree-d polynomials agree on the $d+1$ distinct points $c_0,\ldots,c_d$ and therefore they must be equal. As this equality holds for every $(\boldsymbol{a},t)\in[-m,m]^n$, it follows that for every $\boldsymbol{x}\in[-m,m]^n$,

$$h(x_1,\ldots,x_n)=\sum_{i=0}^{d}\underbrace{\delta_{c_i}(x_n)}_{\text{degree } d}\underbrace{h^{(c_i)}(x_1,\ldots,x_{n-1})}_{\text{degree } d(n-1)},$$

which is a degree dn representation of h. □

Lemma 3.9. *(Degree Reduction) Let $\alpha\in\mathbb{N}, m>0$, and $h\colon\mathbb{R}^n\to\mathbb{R}$ be a multivariate polynomial of finite degree α. If for every radial line segment in the cube $[-m,m]^n$, the restriction of h to that line segment is consistent with a polynomial of degree at most d, then $\alpha\le d$.*

Proof. Fix some $\boldsymbol{b}\in[-m,m]^n$ and consider the radial line $\mathrm{L}_{\mathbf{0},\boldsymbol{b}}$. The n-variate polynomial h, restricted to this line, $h_{\mathbf{0},\boldsymbol{b}}(x)=h(x\boldsymbol{b})$ for x such that $x\boldsymbol{b}\in[-m,m]^n$, can be written as

$$h(x\boldsymbol{b})=\sum_{k=0}^{\alpha}\sum_{i_1+\cdots+i_n=k}c_{i_1,\ldots,i_n}\prod_{j=1}^{n}(xb_j)^{i_j}=\sum_{k=0}^{\alpha}\Bigg(\sum_{i_1+\cdots+i_n=k}c_{i_1,\ldots,i_n}\prod_{j=1}^{n}b_j^{i_j}\Bigg)x^k,$$

which is a univariate degree-α polynomial in x. Consider the coefficient c_α of x^α in $h(x\boldsymbol{b})$ as a function of $\boldsymbol{b}$,

$$c_\alpha(\boldsymbol{b})\triangleq\sum_{i_1+\cdots+i_n=\alpha}c_{i_1,\ldots,i_n}\prod_{j=1}^{n}b_j^{i_j},$$

this is a n-variate polynomial of degree α in the variables $\boldsymbol{b}$. Note that $c_\alpha\not\equiv 0$, as otherwise h would have degree less than α. Fix some $\boldsymbol{b}=\boldsymbol{b}^*\in[-m,m]^n$ such that $c_\alpha(\boldsymbol{b}^*)\neq 0$, such a point exists since c_α has finite number of roots, and view x as the only variable; as $c_\alpha(\boldsymbol{b}^*)\neq 0$, $h_{\mathbf{0},\boldsymbol{b}^*}$ is a univariate polynomial of degree α. However, by assumption $h_{\mathbf{0},\boldsymbol{b}^*}(x)$ has degree at most d, and hence $\alpha\le d$. □

3.3 Polynomial Representation Everywhere

We are now ready to prove that g is a degree-d polynomial over $\mathbb{R}^n$.

Lemma 3.10. *If* CHARACTERIZATIONTEST *fails with probability at most* $2/3$*, then* g *is a degree-*d*,* n*-variate polynomial.*

Proof. Consider the largest n-dimensional hypercube $\mathrm{H} \triangleq [-m, m]^n$ that can be inscribed in the open ball $\mathrm{B}(\mathbf{0}, r)$. By Theorem 3.2, g restricted to any line segment $\mathrm{L}^{\mathrm{B}}_{\boldsymbol{p},\boldsymbol{q}} = \mathrm{L}_{\boldsymbol{p},\boldsymbol{q}} \cap \mathrm{B}(\mathbf{0}, r)$ within the ball $\mathrm{B}(\mathbf{0}, r)$ is consistent with a univariate degree-d polynomial, and therefore the same holds for g restricted to any line segment $\mathrm{L}^{\mathrm{H}}_{\boldsymbol{p},\boldsymbol{q}}$, as $\mathrm{H} \subset \mathrm{B}(\mathbf{0}, r)$.

By Lemma 3.8 and Lemma 3.9, we can conclude that $g(\boldsymbol{x})\colon [-m, m]^n \to \mathbb{R}$ is consistent with a polynomial of degree at most d within H. Hence, for every $\boldsymbol{\alpha} \in \mathbb{N}^n$ such that $\|\boldsymbol{\alpha}\|_1 \leq d$, there exists $c_{\boldsymbol{\alpha}} \in \mathbb{R}$, such that for every $\boldsymbol{x} \in \mathrm{H}$, we can write

$$g(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^n : \|\boldsymbol{\alpha}\|_1 \leq d} c_{\boldsymbol{\alpha}} \prod_{j=1}^{n} x_j^{\alpha_j}. \tag{11}$$

Next, we argue that g is also consistent with this polynomial representation for every point within $\mathrm{B}(\mathbf{0}, r)$. By Theorem 3.2, for any $\boldsymbol{y} \in \mathrm{B}(\mathbf{0}, r) \setminus \mathrm{H}$ and $x \in \mathbb{R}$, for which $x\boldsymbol{y} \in \mathrm{B}(\mathbf{0}, r)$, it follows that $g(x\boldsymbol{y})$ has a unique representation as a univariate polynomial. This polynomial must be consistent with (11) on any point $x'\boldsymbol{y} \in \mathrm{H}$, with $x' \in \mathbb{R}$. As these are both polynomials (agreeing on at least $(d+1)$ points), it follows that both polynomial must be consistent on any point on the line segment $\mathrm{L}^{\mathrm{B}}_{\mathbf{0},\boldsymbol{y}}$. As we know that g is consistent with the univariate representation within $\mathrm{B}(\mathbf{0}, r)$, it follows that the representation (11) holds for $g(\boldsymbol{y})$ for any $\boldsymbol{y} \in \mathrm{B}(\mathbf{0}, r)$.

It remains to argue that g is consistent with this degree-d polynomial representation everywhere. Recall that we defined $g(\boldsymbol{p})$ for $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$, by extrapolating its representation within $\mathrm{B}(\mathbf{0}, r)$ along line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$, to obtain a representation of $g(x\boldsymbol{p})$ as a degree-d (univariate) polynomial.

Thus, g is consistent with a degree-d, n-variate polynomial over $\mathbb{R}^n$.

□

Finally, having established Lemma 3.10, we are ready to prove Lemma 3.1, restated here for convenience.

Lemma 3.1. *If* CHARACTERIZATIONTEST *fails with probability at most* $2/3$*, then* g *is a degree-*d *polynomial, and furthermore for any* $\boldsymbol{p} \in \mathbb{R}^n$*,* $g(\boldsymbol{p}) =$ QUERY-$g(\boldsymbol{p})$ *with probability at least* $1 - \frac{\varepsilon}{2}$*.*

Proof of Lemma 3.1. Suppose that CHARACTERIZATIONTEST fails with probability at most $2/3$, then by Lemma 3.10, g is a degree-d polynomial. It remains to bound the probability that $g(\boldsymbol{p}) \neq$ QUERY-$g(\boldsymbol{p})$ for $\boldsymbol{p} \in \mathbb{R}^n$. To query g on a point $\boldsymbol{p} \in \mathbb{R}^n$, QUERY-$g(\boldsymbol{p})$ call QUERY-g-INBALL$(\boldsymbol{p})$ if $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ or otherwise it attempts to obtain $d+1$ distinct points on the line segment $\mathrm{L}^{\mathrm{B}}_{\mathbf{0},\boldsymbol{p}}$ using QUERY-g-INBALL$(\cdot)$ for each and then interpolate g along this line. For each of these $d+1$ points $\boldsymbol{s}$, QUERY-g-INBALL$(\boldsymbol{s})$ samples an additional N_2' points $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N_2'} \sim \mathcal{N}(\mathbf{0}, I)$, and checks whether

$$\sum_{i \in [d+1]} \alpha_i \cdot f(\boldsymbol{s} + i\boldsymbol{q}_1) = \sum_{i \in [d+1]} \alpha_i \cdot f(\boldsymbol{s} + i\boldsymbol{q}_j),$$

for all $j \in [N_2']$; it fails if any of these checks fail. Note that by the definition of $g_{\boldsymbol{q}}$, this is equivalent to checking whether $g_{\boldsymbol{q}_1}(\boldsymbol{s}) \neq g_{\boldsymbol{q}_j}(\boldsymbol{s})$. By Corollary 3.7 the probability that $g_{\boldsymbol{q}_1}(\boldsymbol{s}) \neq g(\boldsymbol{s})$ is at most $1/(7d)$,

since $\boldsymbol{s} \in \mathrm{B}(\mathbf{0}, r)$. The probability that QUERY-g-INBALL($\boldsymbol{s}$) returns an incorrect value is the probability that $g(\boldsymbol{s}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{s}) = g_{\boldsymbol{q}_j}(\boldsymbol{s})$ for every $\boldsymbol{q}_j$, which is at most $(7d)^{-N_2'} \leq \frac{\varepsilon}{4d}$ by choosing $N_2' = O(\log(1/\varepsilon))$. As QUERY-$g(\boldsymbol{p})$ evaluate at most $d+1$ points using QUERY-g-INBALL$(\cdot)$, the probability that these points are all evaluated correctly, is at least $1 - \varepsilon/2$. □

4 Approximately Testing Polynomials

In this section, we generalize our polynomial tester to be robust against noise. Given query access to a function $f\colon \mathbb{R}^n \to \mathbb{R}$ bounded on the ball $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, and sampling access to an unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$, and constants $\alpha, \varepsilon > 0, \beta \geq \alpha$, a *point-wise approximate tester* for degree-d polynomials is an algorithm that distinguishes between the following two cases with probability at least $2/3$:

- **Yes Case:** There exists a degree-d polynomial $h\colon \mathbb{R}^n \to \mathbb{R}$ such that for every $\boldsymbol{x} \in \mathbb{R}^n$,

$$|f(\boldsymbol{x}) - h(\boldsymbol{x})| \leq \alpha;$$

- **No Case:** For any degree-d polynomial $h\colon \mathbb{R}^n \to \mathbb{R}$,

$$\Pr_{\boldsymbol{x}\sim\mathcal{D}}\left[|f(\boldsymbol{x}) - h(\boldsymbol{x})| > \beta\right] > \varepsilon.$$

An alternative interpretation of this model is as follows: we would like to design a low-degree tester for a function $f^*\colon \mathbb{R}^n \to \mathbb{R}$; however, on every $\boldsymbol{p} \in \mathbb{R}^n$, we are only able to obtain "noisy" evaluations of $f(\boldsymbol{p})$ within an accuracy of up to α. We represent this by giving the tester query-access to a function $f\colon \mathbb{R}^n \to \mathbb{R}$, such that for every $\boldsymbol{p} \in \mathbb{R}^n$,

$$|f^*(\boldsymbol{p}) - f(\boldsymbol{p})| \leq \alpha.$$

This setup is quite natural, and captures the setting in which we are only able to observe a small number of bits of precision of the evaluations of $f^*(\boldsymbol{p})$. The main theorem of this section is the following.[7]

Theorem 1.2. *Let $d \in \mathbb{N}$, $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is bounded in $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, and for $\varepsilon \in (0,1), R > 0$, let $\mathcal{D}$ be an $(\varepsilon/4, R)$-concentrated distribution. Given $\alpha > 0, \beta \geq 2^{(2n)^{O(d)}} R^d \alpha$, query access to f, and sampling access to $\mathcal{D}$, there is a one-sided error, $O(d^5 + \frac{d^2}{\varepsilon}\log\frac{1}{\varepsilon})$-query tester which, distinguishes between the case when f is pointwise α-close to some degree-d polynomial and the case when, for every degree-d polynomial $h\colon \mathbb{R}^n \to \mathbb{R}$, $\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > \beta] > \varepsilon$.*

Our self-corrected function g will be the same as the self-corrected function in the exact case, with one small twist: We use the median rather than the majority, as the median is more robust to errors.

The Self-Corrected Function. Let r be sufficiently small ($r = (4d)^{-6}$ suffices). We first define our self-corrected function for the points $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ as the (weighted) median value of $g_{\boldsymbol{q}}(\boldsymbol{p}) \triangleq \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})$, weighted according to the probability of $\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)$. That is,

$$g(\boldsymbol{p}) \triangleq \operatorname*{med}_{\boldsymbol{q}\sim\mathcal{N}(\mathbf{0},I)}[g_{\boldsymbol{q}}(\boldsymbol{p})].$$

[7] We note that it is possible to relax the condition on f to be bounded in $\mathrm{B}(\mathbf{0}, L)$ for some known $L > 0$. This then leads to β being dependent on L as well. To avoid complicating the parameters, we have chosen to present the less general theorem here.

For points $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$ we define the value of g by extrapolating it from within the ball $\mathrm{B}(\mathbf{0}, r)$ along the radial line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$. To do so, we will interpolate a univariate polynomial on the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$ using the evaluation of g on $d+1$ points in $\mathrm{B}(\mathbf{0}, r)$. For our analysis, it will be convenient to take these points to be $c_0, \ldots, c_d$, where[8] $c_i \triangleq (r/\|\boldsymbol{p}\|_2)\cos(\pi(i+1/2)/(d+1))$. Let $p_{\boldsymbol{p}}$ be the unique univariate degree-d polynomial such that $p_{\boldsymbol{p}}(c_i) = g(\boldsymbol{p}c_i)$ for all i. Then, we define $g(\boldsymbol{p}) \triangleq p_{\boldsymbol{p}}(1)$.

Our tester is given in Algorithm 3, with subroutines in Algorithm 4.

Algorithm 3: Low-Degree Approximate Tester

Procedure LowDegreeApproxTester($f, d, \mathcal{D}, \alpha, \varepsilon, R$)
  Given : Query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, a degree $d \in \mathbb{N}$, sampling access to an unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$, a noise parameter $\alpha > 0$, and a farness parameter $\varepsilon > 0$.
  $\delta \leftarrow 2^{d+1}\alpha$;
  $r \leftarrow (4d)^{-6}$;
  Reject if ApproxCharacterizationTest **rejects**;
  for $N_3 \leftarrow O(\varepsilon^{-1})$ *times* **do**
    Sample $\boldsymbol{p} \sim \mathcal{D}$;
    if $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$ **then**
      Reject if $|f(\boldsymbol{p}) -$ ApproxQuery-$g(\boldsymbol{p})| > 2 \cdot 2^{(2n)^{45d}} R^d \delta$, or if ApproxQuery-$g(\boldsymbol{p})$ **rejects**.
  Accept.

Bridging the gap between Median and Majority. The following lemma will allow us to port the techniques that we used in Section 3, where g was defined as a *majority* over the standard gaussian, to our setting where g is defined as a median. This lemma gives sufficient conditions for the median of any distribution to be close to a random element.

Lemma 4.1. *Let Ω be a sample space, $g : \Omega \to \mathbb{R}$ and $\mathcal{D}$ be a distribution over Ω. For any $\eta \in [0, 1/4]$, $\delta \in \mathbb{R}$, if $\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{D}}[|g(\boldsymbol{q}_1) - g(\boldsymbol{q}_2)| < \delta] > 1 - \eta$, then $\Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(\boldsymbol{q}_1)| < \delta] > 1 - 4\eta$, where $g_{\mathsf{med}} = \mathsf{med}_{\boldsymbol{q}\sim\mathcal{D}}\{g(\boldsymbol{q})\}$.*

The proof is given in Appendix E.

4.1 Preliminaries on Chebyshev Polynomials

Our proof will heavily rely on properties of the Chebyshev polynomials (of the first kind), which we recall next; further details on Chebyshev polynomials can be found in [MH02]. Denote by $T_d(x)$, the d-th Chebyshev polynomial. T_d is a degree-d polynomial and has d roots $\widehat{c}_i \triangleq \cos(\pi(i+1/2)/d)$ for $i \in \{0, \ldots, d-1\}$ in the interval $[-1, 1]$, known as *Chebyshev nodes*. On the interval $[-1, 1]$, the extrema of the Chebyshev polynomials are either -1 or 1, and thus we have

$$x \in [-1, 1] \implies |T_d(x)| \leq 1. \tag{12}$$

Chebyshev polynomials form a basis of polynomials, and in particular satisfy the following orthogonality properties.

[8] These are the Chebyshev nodes of the $(d+1)$-st Chebyshev polynomial, scaled to lie on $\mathrm{L}_{\mathbf{0},\boldsymbol{p}} \cap \mathrm{B}(\mathbf{0}, r)$, as in Section 4.1.

Algorithm 4: Approximate Subroutines

[Recall $\alpha_i \triangleq (-1)^{i+1}\binom{d+1}{i}$ and $\delta = 2^{d+1}\alpha$.]

Procedure ApproxCharacterizationTest

  $N_4 \leftarrow O(d^2)$;

  for N_4 *times* **do**

    for $j \in \{1, \ldots, d+1\}$ **do**

      for $t \in \{0, \ldots, d+1\}$ **do**

        Sample $\boldsymbol{p} \sim \mathcal{N}(\mathbf{0}, j^2(t^2+1)I), \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)$; ▷ [$j^2(t^2+1)$ vs. 1 Test.]

        Reject if $|\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})| > \delta$;

        Sample $\boldsymbol{p} \sim \mathcal{N}(\mathbf{0}, j^2 I), \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I)$; ▷ [$j^2$ vs. t^2+1 Test.]

        Reject if $|\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})| > \delta$;

      Sample $\boldsymbol{p}, \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, j^2 I)$; ▷ [$j^2$ vs. j^2 Test.]

      Reject if $|\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})| > \delta$;

  Accept;

Procedure ApproxQuery-$g(\boldsymbol{p})$

  if $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ **then**

    return ApproxQuery-g-InBall$(\boldsymbol{p})$;

  for $i \in \{0, 1, \ldots, d\}$ **do**

    $c_i \leftarrow \frac{r}{\|\boldsymbol{p}\|_2} \cos\left(\frac{\pi(i+1/2)}{d+1}\right)$;

    $v(c_i) \leftarrow$ ApproxQuery-g-InBall$(c_i\boldsymbol{p})$;

  Let $p_{\boldsymbol{p}} \colon \mathbb{R} \to \mathbb{R}$ be the unique degree-d polynomial such that $p_{\boldsymbol{p}}(c_i) = v(c_i)$ for $i \in \{0, \ldots, d\}$;

  return $p_{\boldsymbol{p}}(1)$;

Procedure ApproxQuery-g-InBall$(\boldsymbol{p})$

  $N_4' \leftarrow O(\log \frac{1}{\varepsilon})$;

  Sample $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N_4'} \sim \mathcal{N}(\mathbf{0}, I)$;

  Reject if there exists $j \in \{2, \ldots, N_4'\}$ such that
  $|\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1) - \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_j)| > 2^{d+2}\delta$;

  return $\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1)$;

Orthogonality. The polynomials T_d are orthogonal with respect to the weight function $w(x) \triangleq (1-x^2)^{-1/2}$ on the interval $[-1, 1]$. Formally,

$$\int_{-1}^{1} T_n(x)\, T_m(x)\, \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \begin{cases} 0 & \text{if } n \neq m, \\ \pi & \text{if } n = m = 0, \\ \pi/2 & \text{if } n = m \neq 0. \end{cases} \tag{13}$$

Discrete orthogonality. The polynomials T_d are also discretely orthogonal:

$$\sum_{k=0}^{d} T_i(\widehat{c}_k)\, T_j(\widehat{c}_k) = \begin{cases} 0 & \text{if } \ i \neq j, \\ d+1 & \text{if } \ i = j = 0, \\ \frac{d+1}{2} & \text{if } \ i = j \neq 0, \end{cases} \tag{14}$$

where $d \geq \max(i, j)$, and the $\widehat{c}_k$ are the $d+1$ Chebyshev nodes of T_{d+1}.

The following lemma will be useful throughout our proof, and follows in a straightforward fashion from properties of Chebyshev polynomials.

Lemma 4.2. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a univariate polynomial of degree at most d. And let $\widehat{c}_0, \ldots, \widehat{c}_d$ be the Chebyshev nodes of T_{d+1}. If $|f(\widehat{c}_k)| \leq \varepsilon$ for every $k \in \{0, \ldots, d\}$, then for every $x \in [-1, 1]$*

$$|f(x)| \leq \sqrt{2}(d+1)\varepsilon.$$

Proof. Let $f(x) = \sum_{i=0}^{d} \alpha_i T_i(x)$ be the Chebyshev expansion of f. Since for every k, we have $f(\widehat{c}_k)^2 \leq \varepsilon^2$,

$$\sum_{k=0}^{d} f(\widehat{c}_k)^2 \leq (d+1)\varepsilon^2.$$

On the other hand,

$$\begin{aligned}
\sum_{k=0}^{d} f(\widehat{c}_k)^2 &= \sum_{k=0}^{d} \left(\sum_{i=0}^{d} \alpha_i T_i(\widehat{c}_k) \right)^2 \\
&= \sum_{k=0}^{d} \sum_{i,j=0}^{d} \alpha_i \alpha_j T_i(\widehat{c}_k) T_j(\widehat{c}_k) \\
&= \sum_{i,j=0}^{d} \alpha_i \alpha_j \sum_{k=0}^{d} T_i(\widehat{c}_k) T_j(\widehat{c}_k) \\
&\geq \frac{d+1}{2} \sum_{i=0}^{d} \alpha_i^2. && \text{(By (14))}
\end{aligned}$$

Combining the above bounds, we have that $|\alpha_i| \leq \sqrt{2}\varepsilon$ for every $i \in \{0, \ldots, d\}$. Thus, for every $x \in [-1, 1]$,

$$|f(x)| = \left| \sum_{i=0}^{d} \alpha_i T_i(x) \right| \leq \sum_{i=0}^{d} |\alpha_i| |T_i(x)| \leq (d+1)\sqrt{2}\varepsilon. \qquad \text{(by (12))}$$

□

By scaling the Chebyshev nodes, we can obtain the following corollary, which is a scaled version of Lemma 4.2 to any given interval, rather than $[-1, 1]$.

Corollary 4.3. *Let f be a univariate degree-d polynomial, let $m \in \mathbb{R}_{>0}$, and $c_k \triangleq m \cos\left(\frac{\pi}{d+1}(k+1/2)\right)$ for $k \in \{0, \ldots, d\}$ be the Chebyshev nodes of T_{d+1} scaled to the interval $[-m, m]$. If $|f(c_k)| \leq \varepsilon$ for every $k \in \{0, \ldots, d\}$, then for any $x \in [-m, m]$,*

$$|f(x)| \leq \sqrt{2}(d+1)\varepsilon.$$

Proof. In the proof of Lemma 4.2 we represent $f(x) = \sum_{i=0}^{d} \alpha_i T_i(x/m)$ as a linear combination of the Chebyshev polynomials with the back-scaled variable. The other parts of the proof are the same. □

4.2 Correctness of the Approximate Polynomial Tester

In the remainder of this section we will argue the correctness of our tester (Theorem 1.2). The next lemma records the properties of g that it guarantees.

Lemma 4.4. *Let $r = (4d)^{-6}$, $\delta = 2^{d+1}\alpha$, as set in Algorithm 4, and $R > r$. If* ApproxCharacterizationTest *fails with probability at most $2/3$, then g is pointwise $2^{(2n)^{45d}} R^d\delta$-close to a degree-d polynomial in $\mathrm{B}(\mathbf{0}, R)$. Furthermore, for every point $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$* ApproxQuery-$g(\boldsymbol{p})$ *well approximates $g(\boldsymbol{p})$ with high probability, that is,*

$$\Pr\left[|g(\boldsymbol{p}) - \textsc{ApproxQuery-}g(\boldsymbol{p})| \le (12R/r)^d 2^{d+4}\delta\right] \ge 1 - \frac{\varepsilon}{4}.$$

We prove the main theorem of this section assuming that Lemma 4.4 holds.

Proof of Theorem 1.2. If f is point-wise α-close to a degree-d polynomial h, then for any $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$,

$$\left|\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})\right| = \left|\sum_{i=0}^{d+1} \alpha_i \cdot (f(\boldsymbol{p} + i\boldsymbol{q}) - h(\boldsymbol{p} + i\boldsymbol{q}))\right| \le \sum_{i=0}^{d+1} |\alpha_i| \cdot \alpha \le 2^{d+1}\alpha = \delta.$$

Thus, ApproxCharacterizationTest always passes, and ApproxQuery-$g(\boldsymbol{p})$ returns a value that is $2^{d+3}\delta$-close to $g(\boldsymbol{p})$, without rejecting with probability 1, and Algorithm 3 always accepts. To see this observe, for any $\boldsymbol{p}, \boldsymbol{q}_1, \boldsymbol{q}_j \in \mathbb{R}^n$,

$$\begin{aligned}
|g_{\boldsymbol{q}_1}(\boldsymbol{p}) - g_{\boldsymbol{q}_j}(\boldsymbol{p})| &= \Big|\sum_{i=1}^{d+1} \alpha_i f(\boldsymbol{p} + i\boldsymbol{q}_1) - \sum_{i=1}^{d+1} \alpha_i f(\boldsymbol{p} + i\boldsymbol{q}_j)\Big| = \Big|\sum_{i=0}^{d+1} \alpha_i f(\boldsymbol{p} + i\boldsymbol{q}_1) - \sum_{i=0}^{d+1} \alpha_i f(\boldsymbol{p} + i\boldsymbol{q}_j)\Big| \\
&= \Big|\sum_{i=0}^{d+1} \alpha_i (f(\boldsymbol{p} + i\boldsymbol{q}_1) - h(\boldsymbol{p} + i\boldsymbol{q}_1)) - \sum_{i=0}^{d+1} \alpha_i (f(\boldsymbol{p} + i\boldsymbol{q}_j) - h(\boldsymbol{p} + i\boldsymbol{q}_j))\Big| \\
&\le 2\sum_{i=0}^{d+1} |\alpha_i (f(\boldsymbol{p} + i\boldsymbol{q}_1) - h(\boldsymbol{p} + i\boldsymbol{q}_1))| \le 2\sum_{i=0}^{d+1} |\alpha_i| \cdot \alpha = 2^{d+2}\alpha < 2^{d+2}\delta.
\end{aligned}$$

So, by Lemma 4.1, we may claim $\Pr_{\boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, I)}[|g(\boldsymbol{p}) - \textsc{ApproxQuery-}g(\boldsymbol{p})| < 2^{d+2}\delta] = 1$, where $\textsc{ApproxQuery-}g(\boldsymbol{p}) \triangleq g_{\boldsymbol{q}_1}(\boldsymbol{p})$, by ApproxQuery-$g$-InBall($\boldsymbol{p}$), and $g(\boldsymbol{p}) = \mathsf{med}_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)}\{g_{\boldsymbol{q}}(\boldsymbol{p})\}$.

Next, we show that if f is β-far from all degree-d polynomials, for $\beta \triangleq 2 \cdot 2^{(2n)^{45d}} R^d\delta$, then Algorithm 3 rejects with probability at least $2/3$. Let $\delta_1 \triangleq 2^{(2n)^{45d}} R^d\delta$, and $\delta_2 \triangleq (12R/r)^d 2^{d+4}\delta$. If ApproxCharacterizationTest fails with probability at least $2/3$, then we reject f with probability at least $2/3$. Otherwise, by Lemma 4.4, g is pointwise δ_1-close in $\mathrm{B}(\mathbf{0}, R)$ to some degree-d polynomial H, and for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$, $\Pr[|g(\boldsymbol{p}) - \textsc{ApproxQuery-}g(\boldsymbol{p})| > \delta_2] < \frac{\varepsilon}{4}$. Hence, $\Pr_{\boldsymbol{p} \sim \mathcal{D}}[|f(\boldsymbol{p}) - g(\boldsymbol{p})| > \beta - \delta_1] > \varepsilon$, noting $\beta - \delta_1 \ge \delta_1$.

The probability that we do not reject in any of the N_3 steps of Algorithm 3 is at most the probability that either $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$, for every sampled point $\boldsymbol{p}$, or $|f(\boldsymbol{p}) - g(\boldsymbol{p})| \le \delta_1$, or that ApproxQuery-$g(\boldsymbol{p})$ returned a value that is δ_2-far from $g(\boldsymbol{p})$ (instead of rejecting). The first event happens with probability at most $\frac{\varepsilon}{4}$, while the last happens with probability at most $\frac{\varepsilon}{4}$ by Lemma 4.4. Thus,

$$\Pr_{\boldsymbol{p} \sim \mathcal{D}}[\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, R) \vee |f(\boldsymbol{p}) - g(\boldsymbol{p})| \le \delta_1 \vee |g(\boldsymbol{p}) - \textsc{Query-}g(\boldsymbol{p})| > \delta_2] \le \frac{\varepsilon}{4} + 1 - \varepsilon + \frac{\varepsilon}{4} \le 1 - \frac{\varepsilon}{2},$$

and Algorithm 3 accepts with probability at most $\left(1 - \frac{\varepsilon}{2}\right)^{N_3} < \frac{1}{3}$ for sufficiently large $N_3 = O(1/\varepsilon)$.

Finally, the bound on the query complexity of the tester follows the same argument, as in the exact case, for Algorithm 1, noting that for sampled points $\boldsymbol{p} \sim \mathcal{D}$ that don't fall in $\mathrm{B}(\mathbf{0}, R)$, LOWDEGREEAPPROXTESTER makes no queries to f, and thus matches the $O(d^5 + \frac{d^2}{\varepsilon}\log(\frac{1}{\varepsilon}))$ query complexity of the LOWDEGREETESTER. □

In the remainder of this section, we will prove Lemma 4.4. This will be done in three steps, similar to the proof outline for Lemma 3.1. First, we show that g is pointwise close a univariate polynomial of degree d on every line segment in $\mathrm{B}(\mathbf{0}, r)$. Then, we show that g is pointwise close to a degree-d, n-variate polynomial within $\mathrm{B}(\mathbf{0}, r)$. Finally, by the fact that g is defined by interpolating evaluations out of $\mathrm{B}(\mathbf{0}, r)$, we show that it is pointwise close to a degree-d, n-variate polynomial on $\mathbb{R}^n$.

4.3 Polynomial Approximation on Every Line Within the Ball

First, we will argue that g is approximately consistent with a degree d polynomial on every line within the ball $\mathrm{B}(\mathbf{0}, r)$. The following is an approximate analogue of Theorem 3.2.

Theorem 4.5. *If* APPROXCHARACTERIZATIONTEST *fails with probability at most* $2/3$*, and* f *is bounded on* $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$*, then for every* $\boldsymbol{a}, \boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)$*, the univariate function* $g_{\boldsymbol{a},\boldsymbol{b}}(x) = g(\boldsymbol{a} + x\boldsymbol{b})$ *defined on points* $x \in \mathrm{L}^{\mathrm{B}}_{\boldsymbol{a},\boldsymbol{b}}$ *is pointwise* $2^{15d^2} \cdot \delta$*-close to a degree-*d*, univariate polynomial.*

The main technical tool in the proof of this theorem will be the following corollary of a result[9] from [Gaj91], which guarantees that any bounded function f defined on a line segment, which has small $(d+1)$-st order finite forward differences, is point-wise close to a degree-d polynomial, on that line segment.

Theorem 4.6. *Let* $x_0 \in \mathbb{R}, d \in \mathbb{N}, \phi, a \in (0, \infty)$*, and a bounded function* $f : (x_0 - a, x_0 + a) \to \mathbb{R}$*, such that for all* $x \in (x_0 - a, x_0 + a)$*, and* $h \in (-a, a)$*, with* $x + (d+1)h \in (x_0 - a, x_0 + a)$*,* $|\Delta_h^{(d+1)}[f](x)| \leq \phi$*. Then, there exists a degree-*d *polynomial* $g : \mathbb{R} \to \mathbb{R}$*, such that for every* $x \in (x_0 - a, x_0 + a)$*,* $|f(x) - g(x)| \leq 2^{8d^2}\phi$*.*

Thus, in order to prove an approximate analogue of Theorem 3.2, it suffices to show that the self-corrected function g satisfies the conditions of Theorem 4.6; i.e., along every line the $(d+1)$st order finite differences of the restriction of g to these lines are small, which will occupy the remainder of this subsection.

Let ρ denote the bound of the probability that each of the tests in the APPROXCHARACTERIZATIONTEST fails. That is, for every $j \in \{1, \ldots, d+1\}$ and $t \in \{0, \ldots, d+1\}$:

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0}, j^2(t^2+1)I) \\ \boldsymbol{q}\sim\mathcal{N}(\mathbf{0}, I)}} \left[\left| \sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \right| > \delta \right] \leq \rho. \qquad [j^2(t^2+1) \text{ vs. } 1 \text{ Test.}] \tag{15}$$

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0}, j^2 I) \\ \boldsymbol{q}\sim\mathcal{N}(\mathbf{0}, (t^2+1)I)}} \left[\left| \sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \right| > \delta \right] \leq \rho. \qquad [j^2 \text{ vs. } t^2 + 1 \text{ Test.}] \tag{16}$$

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0}, j^2 I) \\ \boldsymbol{q}\sim\mathcal{N}(\mathbf{0}, j^2 I)}} \left[\left| \sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \right| > \delta \right] \leq \rho. \qquad [j^2 \text{ vs. } j^2 \text{ Test.}] \tag{17}$$

Following the same argument as in Claim 3.5, we first bound ρ:

[9] Stated in Appendix B as Theorem B.1.

Claim 4.7. If ApproxCharacterizationTest fails with probability at most $2/3$, then ρ is at most $(30d)^{-2}$.

Then, we prove an approximate version of Lemma 3.6 (which lower bounded collision probabilities), via an identical argument, the proof of which can be found in Appendix B:

Lemma 4.8. *For every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, and every $t \in \{0, \dots, d+1\}$,*

$$\Pr_{\substack{\boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, (t^2+1)I) \\ \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)}} \left[|g_{\boldsymbol{q}_1}(\boldsymbol{p}) - g_{\boldsymbol{q}_2}(\boldsymbol{p})| > 2^{d+2}\delta \right] \leq 4d\rho + 48d^5 r.$$

An immediate corollary is the following.

Corollary 4.9. *If* ApproxCharacterizationTest *fails with probability at most $2/3$, then for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ and every $t \in \{0, \dots, d+1\}$,*

$$\Pr_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I)} [|g(\boldsymbol{p}) - g_{\boldsymbol{q}}(\boldsymbol{p})| > 2^{d+3}\delta] < \frac{1}{7d}.$$

Proof. By Claim 4.7, ρ at at most $(30d)^{-2}$. Observe that for any $t \in \{0, \dots, d+1\}$,

$$\begin{aligned}
&\Pr_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I)} [|g(\boldsymbol{p}) - g_{\boldsymbol{q}}(\boldsymbol{p})| > 2^{d+3}\delta] \\
&\leq \Pr_{\boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, I)} [|g(\boldsymbol{p}) - g_{\boldsymbol{q}_1}(\boldsymbol{p})| > 2^{d+2}\delta] + \Pr_{\substack{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, (t^2+1)I) \\ \boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, I)}} [|g_{\boldsymbol{q}_1}(\boldsymbol{p}) - g_{\boldsymbol{q}}(\boldsymbol{p})| > 2^{d+2}\delta] \\
&\leq 5(4d\rho + 48d^5 r). \qquad \text{(By Lemma 4.1, } \because g(\boldsymbol{p}) = \mathsf{med}_{\boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, I)}\{g_{\boldsymbol{q}}(\boldsymbol{p})\}\text{, and Lemma 4.8)}
\end{aligned}$$

By choosing $r = (4d)^{-6}$ and with $\rho \leq (30d)^{-2}$, we get $5(4d\rho + 48d^5 r) \leq 1/(7d)$. □

Next, we prove the approximate analogue of Lemma 3.3, (which showed that the $(d+1)$st order finite differences of g's restrictions to all lines in $\mathrm{B}(\mathbf{0}, r)$ vanish) via an identical argument, and the proof of which can also be found in Appendix B.

Lemma 4.10. *If* ApproxCharacterizationTest *fails with probability at most $2/3$, then for every $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ and sufficiently small $h \in \mathbb{R}$, such that $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$, we have $|\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q})| \leq 2^{2d+5}\delta$.*

We are now ready to prove Theorem 4.5.

Proof of Theorem 4.5. First note that since f is bounded on $\mathrm{B}(\mathbf{0}, 2d\sqrt{n})$, by the same argument as in Lemma 3.4, g is bounded on $\mathrm{B}(\mathbf{0}, r)$. Next, fix some $\boldsymbol{a}, \boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)$; we would like to show that $g_{\boldsymbol{a},\boldsymbol{b}}(x)$ is pointwise close a unique degree-d polynomial for every point x in $\{x \in \mathbb{R} : \boldsymbol{a} + x\boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r)\}$; fix such an x. By Theorem 4.6, it suffices to show that for all sufficiently small $h \in \mathbb{R}$, such that $\boldsymbol{a} + (x+ih)\boldsymbol{b} \in \mathrm{L}^{\mathrm{B}}_{\boldsymbol{a},\boldsymbol{b}}$ for every $i \in [d+1]$,

$$|\Delta_h^{(d+1)}[g_{\boldsymbol{a},\boldsymbol{b}}](x)| = \left| \sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{a} + x\boldsymbol{b} + ih\boldsymbol{b}) \right| \leq 2^{7d^2}\delta.$$

By Lemma 4.10, we have that for every $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ and sufficiently small $h \in \mathbb{R}$, such that $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$, $|\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + ih\boldsymbol{q})| \leq 2^{2d+5}\delta$. Let $\boldsymbol{p} \triangleq \boldsymbol{a} + x\boldsymbol{b}$ and $\boldsymbol{q} \triangleq \boldsymbol{b}$. Since $\mathrm{B}(\mathbf{0}, r)$ is an open ball containing $\boldsymbol{p}$, and $\boldsymbol{q}$, we have $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$. Thus,

$$|\Delta_h^{(d+1)}[g_{\boldsymbol{a},\boldsymbol{b}}](x)| = \left| \sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{a} + x\boldsymbol{b} + ih\boldsymbol{b}) \right| = \left| \sum_{i=0}^{d+1} g(\boldsymbol{p} + ih\boldsymbol{q}) \right| \leq 2^{2d+5}\delta \leq 2^{7d^2}\delta. \qquad \square$$

4.4 Polynomial Approximation Within the Hypercube

Let $0 < m \leq 1$ be a large value such that the hypercube $[-m, m]^n$ is contained within $\mathrm{B}(\mathbf{0}, r)$; setting $m = r/(2\sqrt{n})$ suffices. We will prove that the self-corrected function g is close to a degree-d polynomial on $[-m, m]^n$. The following lemma is the approximate analogue of Lemma 3.8, and Lemma 3.9 combined into one.

Lemma 4.11. *Let $m \in (0, 1], \delta > 0$, and let $h\colon [-m, m]^n \to \mathbb{R}$. If for every line L, the restriction of h to this line h_{L}, is pointwise δ-close to a degree-d polynomial $\widehat{h}_{\mathrm{L}}$, then h is pointwise $((2/m)^{n^{40d}}\delta)$-close to a degree-d, n-variate polynomial.*

The proof of Lemma 4.11 is by induction. At each inductive step we build a degree-$2d$ polynomial and then reduce it to degree d using the following Lemma, the proof of which is in the of which is deferred until the following subsection.

Theorem 1.5. *Let $m \in (0, 1]$, $n \geq 2$ and p be an n-variate polynomial of total degree at most ℓ, for some $d \leq \ell$. If for every $\boldsymbol{a} \in [-m, m]^n$, the univariate polynomial $p_{\mathbf{0},\boldsymbol{a}}(t) = p(\boldsymbol{a}t)$ which is the restriction of p to the radial line $\mathrm{L}_{\mathbf{0},\boldsymbol{a}}$, is pointwise ε-close to a degree-d univariate polynomial on the interval $t \in [-1, 1]$, then p is pointwise η-close to $p^{\leq d}$ (the truncation of p to degree d) on $[-m, m]^n$ for $\eta = 2(2/m)^{2n^{18\ell}}\varepsilon$.*

Proof of Lemma 4.11. We will show that h is pointwise close to an n-variate degree-d polynomial H_n by induction on the dimension n. Set $\delta_n \triangleq (2/m)^{n^{40d}}\delta$. For the base case, when $n = 1$, we have that $h = h_{0,1}$ is pointwise δ-close to a univariate polynomial $\widehat{h}_{0,1}$ of degree d by assumption, so we let $H_1 = \widehat{h}_{0,1}$ and $\delta_1 = (2/m)\delta \geq \delta$.

Assume that the statement is true for $n - 1$, with $\delta_{n-1} = (2/m)^{(n-1)^{40d}}\delta$. For any $c \in [-m, m]$, define $h^{(c)}\colon [-m, m]^{n-1} \to \mathbb{R}$ as

$$h^{(c)}(x_1, \ldots, x_{n-1}) \triangleq h(x_1, \ldots, x_{n-1}, c).$$

We will argue that $h^{(c)}$ is pointwise δ_{n-1}-close to an $(n-1)$-variate polynomial of total degree at most d. Fix $i \in [n-1]$, and $\boldsymbol{a} \in [-m, m]^{n-1}$ with $a_i = 0$, and let $\boldsymbol{a}^{\uparrow} = (a_1, \ldots, a_{n-1}, c) \in [-m, m]^n$ to be an extension of $\boldsymbol{a}$ to dimension n. As well, let $\boldsymbol{e}_i$ denote the ith standard basis vector. By assumption, $h_{\boldsymbol{a}^{\uparrow},\boldsymbol{e}_i}\colon [-m, m] \to \mathbb{R}$ is pointwise δ-close to some univariate degree-d polynomial, which we will denote by $\widehat{h}_{\boldsymbol{a}^{\uparrow},\boldsymbol{e}_i}$. For every $x \in [-m, m]$, we have

$$h_{\boldsymbol{a}^{\uparrow},\boldsymbol{e}_i}(x) = h(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_{n-1}, c) = h^{(c)}(\boldsymbol{a} + x\boldsymbol{e}_i) = h^{(c)}_{\boldsymbol{a},\boldsymbol{e}_i}(x),$$

and so $h^{(c)}_{\boldsymbol{a},\boldsymbol{e}_i}(x)$ is δ-close to $\widehat{h}_{\boldsymbol{a}^{\uparrow},\boldsymbol{e}_i}$ on $[-m, m]$. Thus, by the induction hypothesis, $h^{(c)}\colon [-m, m]^{n-1} \to \mathbb{R}$ is pointwise δ_{n-1}-close to an $(n-1)$-variate polynomial of total degree at most d, which we will denote by $H^{(c)}_{n-1}$.

It remains to show that h is pointwise δ_n-close to an n-variate polynomial H_n of total degree at most d on $[-m, m]^n$. Let $c_0, \ldots, c_d \in [-m, m]$ be the scaled Chebyshev nodes $c_i \triangleq m\cos\left(\frac{\pi}{2}(i + 1/2)/(d + 1)\right)$. Let $\delta_{c_i}\colon \mathbb{R} \to \mathbb{R}$ be the unique degree-d polynomial which satisfies

$$\delta_{c_i}(c_j) = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Using δ_{c_i}, we build a degree at most $2d$ polynomial

$$H(x_1, \ldots, x_n) \triangleq \sum_{i=0}^{d} \underbrace{\delta_{c_i}(x_n)}_{\text{degree } d} \underbrace{H^{(c_i)}_{n-1}(x_1, \ldots, x_{n-1})}_{\text{degree } \leq d}.$$

Next, we argue that H is pointwise close to h. Fix some $\boldsymbol{b} = (b_1, \ldots, b_{n-1}) \in [-m, m]^{n-1}$ and let $\boldsymbol{b}^{\uparrow} = (b_1, \ldots, b_{n-1}, 0)$ be an extension of $\boldsymbol{b}$ to dimension n. Consider the following two univariate functions in the variable t. The first function is

$$h_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(t) = h(b_1, \ldots, b_{n-1}, t),$$

which by assumption is pointwise δ-close to a univariate degree-d polynomial $\widehat{h}_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(t)$. The second is the polynomial H with the first $n-1$ variables fixed to $\boldsymbol{b}$,

$$H(b_1, \ldots, b_{n-1}, t) = \sum_{i=0}^{d} \delta_{c_i}(t) H_{n-1}^{(c_i)}(b_1, \ldots, b_{n-1}).$$

Since the $H_{n-1}^{(c_i)}(b_1, \ldots, b_{n-1})$ are constants in t, $H_n(b_1, \ldots, b_{n-1}, t)$ is a univariate polynomial of degree d.

Observe that for $c_0, \ldots c_d$,

$$\begin{aligned} |\widehat{h}_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(c_i) - H(b_1, \ldots, b_{n-1}, c_i)| &\leq |\widehat{h}_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(c_i) - h_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(c_i)| + |h^{(c_i)}(\boldsymbol{b}) - H(b_1, \ldots, b_{n-1}, c_i)| \\ &\leq \delta + \delta_{n-1}, \end{aligned}$$

where the first inequality follows because $h_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(c_i) = h^{(c_i)}(\boldsymbol{b})$ and the second follows by the inductive hypothesis, since $H(b_1, \ldots, b_{n-1}, c_i) = H_{n-1}^{(c_i)}(b_1, \ldots, b_n)$ by definition.

Applying Corollary 4.3 to the error function $e(t) = \widehat{h}_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(t) - H(b_1, \ldots, b_{n-1}, t)$, we have that for every $t \in [-m, m]$, the difference between the two degree-d polynomials is at most

$$|\widehat{h}_{\boldsymbol{b}^{\uparrow}, \boldsymbol{e_n}}(t) - H(b_1, \ldots, b_{n-1}, t)| \leq \sqrt{2}(d+1)(\delta + \delta_{n-1}).$$

Since this is true for every $\boldsymbol{b} \in [-m, m]^{n-1}$ and $t \in [-m, m]$, we have that for every $\boldsymbol{x} \in [-m, m]^n$

$$|H(\boldsymbol{x}) - h(\boldsymbol{x})| \leq |H(\boldsymbol{x}) - \widehat{h}_{(x_1, \ldots, x_{n-1}), \boldsymbol{e}_n}(x_n)| + |\widehat{h}_{(x_1, \ldots, x_{n-1}), \boldsymbol{e}_n}(x_n) - h(\boldsymbol{x})| \leq \sqrt{2}(d+1)(\delta + \delta_{n-1}) + \delta.$$

Note that for every $\boldsymbol{a} \in [-m, m]^n$, the restriction $H_{\boldsymbol{0}, \boldsymbol{a}}$ on the radial line $\mathrm{L}_{\boldsymbol{0}, \boldsymbol{a}}$ is a univariate polynomial which is pointwise $(10d\delta_{n-1})$-close to the degree d univariate polynomial $\widehat{h}_{\boldsymbol{0}, \boldsymbol{a}}$ on points in the cube $[-m, m]^n$, since $|H_{\boldsymbol{0}, \boldsymbol{a}}(t) - \widehat{h}_{\boldsymbol{0}, \boldsymbol{a}}(t)| \leq |H_{\boldsymbol{0}, \boldsymbol{a}}(t) - h_{\boldsymbol{0}, \boldsymbol{a}}(t)| + |h_{\boldsymbol{0}, \boldsymbol{a}}(t) - \widehat{h}_{\boldsymbol{0}, \boldsymbol{a}}(t)| \leq \sqrt{2}(d+1)(\delta + \delta_{n-1}) + 2\delta$ for every $t \in [-1, 1]$.

Applying Theorem 1.5 on H, we have that $H_n \triangleq H^{\leq d}$ is pointwise $(20d(2/m)^{2n^{36d}}\delta_{n-1})$-close to H on the cube $[-m, m]^n$. Thus, for every $\boldsymbol{x} \in [-m, m]^n$, we have

$$\begin{aligned} |h(\boldsymbol{x}) - H_n(\boldsymbol{x})| &\leq |h(\boldsymbol{x}) - H(\boldsymbol{x})| + |H(\boldsymbol{x}) - H_n(\boldsymbol{x})| \\ &\leq (9d\delta_{n-1}) + (20d(2/m)^{2n^{36d}}\delta_{n-1}) \\ &\leq 30d(2/m)^{2n^{36d}}\delta_{n-1} \\ &\leq \delta_n. \end{aligned}$$

□

4.4.1 Proof of Theorem 1.5

Consider the monic Chebyshev polynomials $\widetilde{T}_n(x) \triangleq 2^{1-n}T_n(x)$, with $\|\widetilde{T}_n(x)\|_\infty = 2^{1-n}$ on the interval $x \in [-1, 1]$. Then, by the extremal property that Chebyshev polynomials have the minimum maximal absolute value among all monic polynomials of the same degree on the interval $[-1, 1]$, we have the following fact and the subsequent lemma.

Fact 4.12. *For every monic polynomial $p(t)$ of degree $d \geq 1$ there exists $x \in [-1,1]$ such that $|p(x)| \geq 2^{1-d}$.*

Corollary 4.13. *Let $m \leq 1$ and $p(t)$ is a monic polynomial of degree $d \geq 1$. Then, there exists $x \in [-m,m]$ such that $|p(x)| \geq m^d 2^{1-d}$.*

Proof. Let $p(x) = x^d + \sum_{i=0}^{d-1} \alpha_i x^i$, and note that $p(xm) = (xm)^d + + \sum_{i=0}^{d-1} \alpha_i (xm)^i$. Then $p(xm)/m^d$ is a degree-d *monic* polynomial. Thus, by Fact 4.12, there exists $x \in [-1,1]$ such that $|p(xm)/m^d| \geq 2^{1-d}$. That is, there is $x \in [-m,m]$ such that $|p(x)| \geq m^d 2^{1-d}$. □

Lemma 1.6. *Let $p(x) = \sum_{i=0}^{d} \alpha_i x^i$ be a degree-d polynomial and let $\eta > 0$. If $|\alpha_i| \geq \eta$ for some $i \geq 1$, then there exists $x \in [-1,1]$ such that $|p(x)| \geq 2^{-2d^2}\eta$.*

Proof. Let ℓ be the largest index such that $|\alpha_\ell| \geq 2^{(-\sum_{k=1}^{\ell} k)}\eta$; ℓ exists since $|\alpha_i| \geq \eta \geq 2^{(-\sum_{k=1}^{i} k)}\eta$. Note that by the maximality of ℓ, for every $j > \ell$, $|\alpha_j| < 2^{(-\sum_{k=1}^{j} k)}\eta$. Therefore,

$$\frac{p(x)}{\alpha_\ell} = \sum_{j=0}^{\ell} \frac{\alpha_j}{\alpha_\ell} x^j + \sum_{j=\ell+1}^{d} \frac{\alpha_j}{c_\alpha} x^j = \frac{p^{\leq \ell}(x)}{\alpha_\ell} + \frac{p^{>\ell}(x)}{\alpha_\ell}.$$

Observe that $p^{\leq \ell}(x)/\alpha_\ell$ is a monic polynomial of degree at most ℓ, and thus by Fact 4.12 there exists $x \in [-1,1]$ such that $|p^{\leq \ell}(x)|/|\alpha_\ell| \geq 2^{1-\ell}$. On the other hand, for every $x \in [-1,1]$ we have that

$$\frac{|p^{>\ell}(x)|}{|\alpha_\ell|} \leq \sum_{j=\ell+1}^{d} \frac{|\alpha_j|}{|\alpha_\ell|} \leq \sum_{j=\ell+1}^{d} 2^{-\left(\sum_{k=\ell+1}^{j} k\right)} \leq \sum_{j>\ell} 2^{-j} \leq 2^{-\ell}.$$

Altogether this implies that there is some $x \in [-1,1]$ such that

$$\frac{|p(x)|}{|\alpha_\ell|} \geq \frac{|p^{\leq \ell}(x)|}{|\alpha_\ell|} - \frac{|p^{>\ell}(x)|}{|\alpha_\ell|} \geq 2^{1-\ell} - 2^{-\ell} = 2^{-\ell},$$

and it follows that $|p(x)| \geq 2^{-\ell}|\alpha_\ell| \geq 2^{-\ell}(2^{-\sum_{k=1}^{\ell} k})\eta = (2^{-\ell(\ell+3)/2})\eta \geq \eta 2^{-2d^2}$. □

Corollary 4.14. *Fix $\eta > 0$, $m < 1$, and let $p(x) = \sum_{i=0}^{d} \alpha_i x^i$ be a degree-d polynomial. If $|\alpha_i| \geq \eta$ for some $i \in [d]$, then there exists $x \in [-m,m]$ such that $|p(x)| \geq 2^{-2d^2} m^d \eta$.*

Proof. Writing $p(x) = p^{\leq \ell}(x) + p^{>\ell}(x)$ as in Lemma 1.6, and noticing $p^{\leq \ell}(x)/\alpha_\ell$ is a monic, degree ℓ polynomial, we invoke Corollary 4.13 to claim, there exists $x \in [-m,m]$ such that $|p^{\leq \ell}(x)|/|\alpha_\ell| \geq 2^{1-\ell} m^\ell$. While simultaneously, we have for every $x \in [-m,m]$:

$$\frac{|p^{>\ell}(x)|}{|\alpha_\ell|} \leq \sum_{j=\ell+1}^{d} \frac{|\alpha_j||m|^j}{|\alpha_\ell|} \leq \sum_{j=\ell+1}^{d} 2^{-\left(\sum_{k=\ell+1}^{d} k\right)} m^j \leq \sum_{j>\ell} 2^{-j} m^j \leq m^\ell \sum_{j>\ell} 2^{-j} \leq 2^{-\ell} m^\ell.$$

Therefore, there exists $x \in [-m,m]$ such that $|p(x)|/|\alpha_\ell| \geq 2^{-\ell} m^\ell$, and we have $|p(x)| \geq 2^{-\ell} m^\ell |\alpha_\ell| \geq 2^{-2d^2} m^d \eta$. □

For vectors $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$, denote by $\langle \boldsymbol{y}, \boldsymbol{z} \rangle = \sum_{j \in [n]} y_j z_j$ the standard inner product between them.

Lemma 4.15. *For $k \geq n^{8d}$ there exists $\boldsymbol{y} \in \{0, \ldots, k\}^n$ such that for any $\boldsymbol{z}^{(1)} \neq \boldsymbol{z}^{(2)} \in \{0, \ldots, d\}^n$ satisfying $\sum_{j=1}^{n} z_j^{(i)} \leq 2d$ for $i \in \{1,2\}$, it holds that $\langle \boldsymbol{y}, \boldsymbol{z}^{(1)} \rangle \neq \langle \boldsymbol{y}, \boldsymbol{z}^{(2)} \rangle$.*

Proof. Let $\mathcal{Z} = \{\boldsymbol{z} \in \{-d, \ldots, 0, \ldots, d\}^n \colon \|\boldsymbol{z}\|_0 \leq 4d\}$, where $\|\cdot\|_0$ gives the number of non-zero coordinates. Note that $\boldsymbol{z}^{(1)} - \boldsymbol{z}^{(2)} \in \mathcal{Z}$. Thus, it suffices to show that there exists $\boldsymbol{y}$ such that for any $\boldsymbol{z} \in \mathcal{Z}$, if $\langle \boldsymbol{y}, \boldsymbol{z} \rangle = 0$ then $\boldsymbol{z} = \boldsymbol{0}$. Suppose $\boldsymbol{z} \neq \boldsymbol{0}$, and let ℓ be such that $z_\ell \neq 0$. Sample $\boldsymbol{y}$ uniformly from $\{0, \ldots, k\}^n$. Then,

$$\Pr_{\boldsymbol{y}}[\langle \boldsymbol{y}, \boldsymbol{z} \rangle = 0] = \Pr\left[y_\ell = -\frac{1}{z_\ell} \sum_{j \neq \ell} y_j z_j\right] \leq \frac{1}{k}.$$

By a union bound over all $\boldsymbol{z} \neq \boldsymbol{0}$,

$$\Pr_{\boldsymbol{y}}[\exists \boldsymbol{z} : \langle \boldsymbol{y}, \boldsymbol{z} \rangle = 0] \leq \frac{|\mathcal{Z}| - 1}{k} < \frac{\binom{n}{4d}(2d+1)^{4d}}{k} \leq \frac{n^{8d}}{k}.$$

Thus, choosing $k \geq n^{8d}$, there exists $\boldsymbol{y}$ such that for every $\boldsymbol{0} \neq \boldsymbol{z} \in \mathcal{Z}$ it holds $\langle \boldsymbol{y}, \boldsymbol{z} \rangle \neq 0$. □

Let us introduce some notation. For an n-variate polynomial $p(\boldsymbol{x}) = \sum_{I \in \mathbb{N}^n} \alpha_I \prod_{j \in [n]} x_j^{I_j}$, let

$$p^{\leq d}(\boldsymbol{x}) \triangleq \sum_{I \in \mathbb{N}^n : \|I\|_1 \leq d} \alpha_I \prod_{j \in [n]} x_j^{I_j}$$

be the truncation of p to degree d.

Fact 4.16. *Let $p(\boldsymbol{x}) = \sum_{I \in \mathbb{N}^n} \alpha_I \prod_{j \in [n]} x_j^{I_j}$ be an n-variate polynomial. Then, for every point $\boldsymbol{x} \in [-1, 1]^n$,*

$$|p(\boldsymbol{x}) - p^{\leq d}(\boldsymbol{x})| \leq \sum_{I \in \mathbb{N}^n : \|I\|_1 > d} |\alpha_I|.$$

Proof. Observe, that for every $\boldsymbol{x} \in [-1, 1]^n$, for every j, $|x_j| \leq 1$, and hence

$$|p(\boldsymbol{x}) - p^{\leq d}(\boldsymbol{x})| = \Big| \sum_{I \in \mathbb{N}^n : \|I\|_1 > d} \alpha_I \prod_{j \in [n]} x_j^{I_j} \Big| \leq \sum_{I \in \mathbb{N}^n : \|I\|_1 > d} |\alpha_I| \prod_{j \in [n]} |x_j|^{I_j} \leq \sum_{I \in \mathbb{N}^n : \|I\|_1 > d} |\alpha_I|. \qquad \square$$

Corollary 4.17. *Let $p(\boldsymbol{x}) = \sum_{I \in \mathbb{N}^n} \alpha_I \prod_{j \in [n]} x_j^{I_j}$ be a polynomial of total degree $\ell \geq d$. If for every $I \in \mathbb{N}^n$ such that $\|I\|_1 > d$, we have $|\alpha_I| \leq \eta$, then p is pointwise $\tilde{\eta}$-close to $p^{\leq d}$ on $[-1, 1]^n$, where $\tilde{\eta} = \eta |\{I \mid d < \|I\|_1 \leq \ell\}| \leq \eta (n+1)^\ell$.*

We are now ready to prove Theorem 1.5.

Proof of Theorem 1.5. Let

$$p(\boldsymbol{x}) = \sum_{I \in \mathbb{N}^n : \|I\|_1 \leq \ell} \alpha_I \prod_{j \in [n]} x_j^{I_j}.$$

Assume by contradiction that p is not pointwise η-close to $p^{\leq d}$ on $[-m, m]^n$. Then, by Corollary 4.17 there exists $\tilde{I}$ such that $\|\tilde{I}\|_1 > d$ and $|\alpha_{\tilde{I}}| > (n+1)^{-\ell} \eta$. Fix $\boldsymbol{a} = (a_1, \ldots, a_n) \in [-m, m]^n$, then the restriction of p to the line $L_{(\boldsymbol{0}, \boldsymbol{a})}$ is

$$p_{\boldsymbol{0}, \boldsymbol{a}}(t) = \sum_{I \in \mathbb{N}^n : \|I\|_1 \leq \ell} \alpha_I \prod_{j \in [n]} a_j^{I_j} t^{\|I\|_1} = \sum_{r=0}^{\ell} \left(\sum_{I \in \mathbb{N}^n : \|I\|_1 = r} \alpha_I \prod_{j \in [n]} a_j^{I_j} \right) t^r.$$

By the Fourier-Chebyshev expansion, we can write each monomial $t^r = \sum_{k=0}^{r} \beta_{k,r} T_k(t)$, where

$$\beta_{k,r} = \begin{cases} 0 & \text{if } k \not\equiv r \bmod 2, \\ 2^{1-r}\binom{r}{(r-k)/2} & \text{if } k \equiv r \bmod 2, \text{ and } k \neq 0, \\ 2^{-r}\binom{r}{(r-k)/2} & \text{if } k = 0, \text{ and } r \equiv 0 \bmod 2. \end{cases}$$

which gives

$$\begin{aligned} p_{\mathbf{0},\boldsymbol{a}}(t) &= \sum_{r=0}^{\ell} \left(\sum_{I \in \mathbb{N}^n : \|I\|_1 = r} \alpha_I \prod_{j \in [n]} a_j^{I_j} \right) \sum_{k=0}^{r} \beta_{k,r} T_k(t) \\ &= \sum_{r=0}^{\ell} \sum_{k=0}^{r} \left(\sum_{I \in \mathbb{N}^n : \|I\|_1 = r} \beta_{k,r} \alpha_I \prod_{j \in [n]} a_j^{I_j} \right) T_k(t) \\ &= \sum_{k=0}^{\ell} \underbrace{\left(\sum_{r=k}^{\ell} \sum_{I \in \mathbb{N}^n : \|I\|_1 = r} \beta_{k,r} \alpha_I \prod_{j \in [n]} a_j^{I_j} \right)}_{\triangleq q_k(\boldsymbol{a})} T_k(t). \end{aligned}$$

Let $q_k(\boldsymbol{a})$ be the coefficient of $T_k(t)$ in the previous expansion and let $\tilde{r} = \|\tilde{I}\|_1$. Note that by the values of the coefficients $\beta_{k,r}$, we have $\alpha_{\tilde{I}}$ appears either in q_{d+1} or in q_{d+2} depending on the parity of $\tilde{r}$; let $i = d+1$ or $i = d+2$ be such that $i \equiv \tilde{r} \bmod 2$. Thus, the coefficient of the monomial $\prod_{j \in [n]} a_j^{\tilde{I}_j}$ in q_i is

$$\beta_{i,\tilde{r}} a_{\tilde{I}} = 2^{1-\tilde{r}} \binom{\tilde{r}}{(\tilde{r}-i)/2} \alpha_{\tilde{I}} \geq 2^{-\tilde{r}} \alpha_{\tilde{I}} \geq (2n+2)^{-\ell} \eta.$$

Using this, we will derive a contradiction to the following claim.

Claim 4.18. For all $\boldsymbol{a} \in [-m, m]^n$, $|q_i(\boldsymbol{a})| \leq \sqrt{2}\varepsilon$.

We defer the proof of Claim 4.18 until later and complete the proof first. As $\|I\|_1 \leq \ell$ for every I, let $\boldsymbol{y} \in \{0, \ldots, n^{8\ell}\}^n$ be given by Lemma 4.15 and consider the univariate polynomial $\widetilde{q}_i(z) \triangleq q_i(z^{\boldsymbol{y}_1}, z^{\boldsymbol{y}_2} \ldots, z^{\boldsymbol{y}_n})$ in z. By the guarantee of Lemma 4.15, for any $I \neq I'$ with $\|I\|_1, \|I'\|_1 \leq \ell$, it holds that $\langle \boldsymbol{y}, I \rangle \neq \langle \boldsymbol{y}, I' \rangle$, and thus the coefficients of $\widetilde{q}_i(z)$ are exactly the same as coefficients of q_i (that is, no two monomials become the same after the substitution of $z^{\boldsymbol{y}_i}$). Therefore, there exists a coefficient in $\widetilde{q}_i$ which is at least $(2n+2)^{-\ell}\eta$. On the other hand, since $\langle \boldsymbol{y}, I \rangle \leq \|I\|_1 n^{8\ell} \leq \ell n^{8\ell}$, the degree of $\widetilde{q}_i$ is at most $\ell n^{8\ell} \leq n^{9\ell}$, and thus Corollary 4.14 implies that there is some $z \in [-m, m]$, such that $|\widetilde{q}_i(z)| \geq 2^{-2n^{18\ell}} m^{n^{9\ell}} \eta \geq (\frac{2}{m})^{-2n^{18\ell}} \eta > \sqrt{2}\varepsilon$. This contradicts Claim 4.18. □

Proof of Claim 4.18. Let $\widehat{p}_{\mathbf{0},\boldsymbol{a}}(t)$ be the univariate degree-d polynomial which is pointwise ε-close to $p_{\mathbf{0},\boldsymbol{a}}(t)$ on $t \in [-1, 1]$, and let its Fourier-Chebyshev expansion be $\widehat{p}_{\mathbf{0},\boldsymbol{a}}(t) = \sum_{k=0}^{d} \gamma_k T_k(t)$. Consider the error polynomial

$$e(t) \triangleq p_{\mathbf{0},\boldsymbol{a}}(t) - \widehat{p}_{\mathbf{0},\boldsymbol{a}}(t) = \sum_{k=0}^{\ell} (q_k(\boldsymbol{a}) - \gamma_k) T_k(t),$$

where we define $\gamma_k \triangleq 0$ for $k > d$. Note that since $p_{\mathbf{0},\boldsymbol{a}}$ and $\widehat{p}_{\mathbf{0},\boldsymbol{a}}$ are ε-close on $[-1, 1]$, $|e(t)| \leq \varepsilon$ for all $t \in [-1, 1]$. Letting $w(t) \triangleq (1-t^2)^{-1/2}$ be the Chebyshev weight function, and noting that $\int_{-1}^{1} w(t) dt = \pi$,

we have,

$$\begin{aligned}
\varepsilon^2\pi &\geq \int_{-1}^{1} e^2(t)w(t)dt \\
&= \int_{-1}^{1} \left(\sum_{k=0}^{\ell}(q_k(\boldsymbol{a}) - \gamma_k)T_k(t)\right)^2 w(t)dt \\
&= \sum_{k=0}^{\ell}(q_k(\boldsymbol{a}) - \gamma_k)^2 \int_{-1}^{1} T_k(t)T_k(t)w(t)dt \\
&\geq \frac{\pi}{2}\sum_{k=0}^{\ell}(q_k(\boldsymbol{a}) - \gamma_k)^2 \\
&\geq \frac{\pi}{2}\left(q_i(\boldsymbol{a})\right)^2,
\end{aligned}$$

where the first steps by the orthogonality of Chebyshev polynomials (13), and the final inequality follows because $i > d$ and so $\gamma_i = 0$. Rearranging, we conclude that $|q_i(\boldsymbol{a})| \leq \sqrt{2}\varepsilon$. □

4.4.2 Extrapolation

In this section we show that if g is pointwise close to a degree d polynomial then within $\mathrm{B}(\mathbf{0}, r)$, then it must be pointwise close to a degree-d polynomial within a bigger ball $\mathrm{B}(\mathbf{0}, R)$.

Lemma 4.19. *Let $R > r' > 0$ be any real numbers. If g is pointwise η-close to a degree-d polynomial in $\mathrm{B}(\mathbf{0}, r')$, then g is pointwise $(12R/r')^d\eta$-close to a degree-d polynomial on all points in $\mathrm{B}(\mathbf{0}, R)$.*

Proof. Let $H\colon \mathbb{R}^n \to \mathbb{R}$ be the degree-d polynomial which is η-close to g on $\mathrm{B}(\mathbf{0}, r')$. We will argue that for any $\boldsymbol{x} \in \mathrm{B}(\mathbf{0}, R)$,

$$|g(\boldsymbol{x}) - H(\boldsymbol{x})| \leq (12R/r')^d\eta.$$

If $\boldsymbol{x} \in \mathrm{B}(\mathbf{0}, r')$, then this holds by assumption, so we consider the case when $\boldsymbol{x} \notin \mathrm{B}(\mathbf{0}, r')$. Recall that we define the value of g on points $\boldsymbol{x} \notin \mathrm{B}(\mathbf{0}, r')$ by pretending that it *is* a degree-d polynomial and using $d+1$ points in $\mathrm{B}(\mathbf{0}, r')$ to extrapolate its value along radial lines from within the ball. In particular, let $c_0, \ldots, c_d$ be the Chebyshev nodes $c_i \triangleq (r'/\|\boldsymbol{x}\|_2)\cos\left(\frac{\pi}{d+1}(i+1/2)\right)$, scaled so that they lie within $\mathrm{L}_{\mathbf{0},\boldsymbol{x}} \cap \mathrm{B}(\mathbf{0}, r')$. Then, the value of $g(\boldsymbol{x})$ for $\boldsymbol{x} \notin \mathrm{B}(\mathbf{0}, r')$ is defined by interpolating a degree-d univariate polynomial $p_{\boldsymbol{x}}$ such that $p_{\boldsymbol{x}}(c_i) = g(\boldsymbol{x}c_i)$ for i, and then the value of $g(\boldsymbol{x})$ is defined as $p_{\boldsymbol{x}}(1)$.

Thus, in order to bound the distance between $g(\boldsymbol{x})$ and $H(\boldsymbol{x})$, it suffices to bound the distance between $p_{\boldsymbol{x}}(t)$ and $H_{(\mathbf{0},\boldsymbol{x})}(t)$ for $t = 1$. Consider the error polynomial $e(t) \triangleq p_{\boldsymbol{x}}(t) - H_{(\mathbf{0},\boldsymbol{x})}(t)$. As e is a polynomial of degree at most d, we can consider its Fourier-Chebyshev expansion,

$$e(t) = \sum_{i=0}^{d} \alpha_i T_i(t\|\boldsymbol{x}\|_2/r'),$$

where $T_i(t\|\boldsymbol{x}\|_2/r')$ is the ith Chebyshev polynomial with the Chebyshev nodes back-scaled to the interval $[-1, 1]$. By assumption, $e(c_i) \leq \eta$ for each i, which allows us (by the same argument as in Lemma 4.2 and Corollary 4.3) to bound the coefficients $|\alpha_i| \leq \sqrt{2}\eta$. The ith Chebyshev polynomial involves at most $i+1$ terms, each of which are of degree at most i and has coefficients of value at most 2^i, and therefore

$|T_i(\|\boldsymbol{x}\|_2/r')| \le (i+1)2^i(\|\boldsymbol{x}\|_2/r')^i$. Altogether, this allows us to bound the value of the error polynomial on $t=1$ by

$$|e(1)| \le \sqrt{2}\eta(d+1)^2(2\|\boldsymbol{x}\|_2/r')^d \le (12R/r')^d\eta,$$

where the second inequality is by $\sqrt{2}(d+1)^2 \le 6^d$ for every $d \ge 1$, and the last holds as $\boldsymbol{x} \in \mathrm{B}(\mathbf{0}, R)$. Since $g(\boldsymbol{x}) = p_{\boldsymbol{x}}(1)$, we have that the distance between $g(\boldsymbol{x})$ and $H(\boldsymbol{x})$ is at most $(12R/r')^d\eta$. □

4.5 Approximate Polynomial Representation in a Large Ball

We now prove the approximate analogue of Lemma 3.10 which showed g is a degree-d polynomial over $\mathbb{R}^n$.

Lemma 4.20. *Let $r = (4d)^{-6}$ and $R > r$. If* ApproxCharacterizationTest *fails with probability at most $2/3$, then g is point-wise $2^{(2n)^{45d}}R^d\delta$-close to a degree-d, n-variate polynomial on all points in $\mathrm{B}(\mathbf{0}, R)$.*

Proof. By Theorem 4.5, g restricted to any line segment $\mathrm{L}^{\mathrm{B}}_{\boldsymbol{p},\boldsymbol{q}} = \mathrm{L}_{\boldsymbol{p},\boldsymbol{q}} \cap \mathrm{B}(\mathbf{0}, r)$ is point-wise $2^{15d^2}\delta$-close to a unique univariate degree-d polynomial. Applying Lemma 4.11 (with $m = r/(2\sqrt{n})$), we have that g is pointwise $2^{15d^2}(4\sqrt{n}/r)^{n^{40d}}\delta$-close to a degree d, n-variate polynomial on every point in the hypercube $\mathrm{H} = [-m, m]^n$, contained within $\mathrm{B}(\mathbf{0}, r)$. We then consider a smaller ball $\mathrm{B}(\mathbf{0}, r')$ of radius $r' = m$, contained within H. By Lemma 4.19, it follows that g is point-wise $2^{15d^2}(4\sqrt{n}/r)^{n^{40d}}(24\sqrt{n}R/r)^d\delta \le 2^{(2n)^{45d}}R^d\delta$-close to a degree-$d$, n-variate polynomial in $\mathrm{B}(\mathbf{0}, R)$. □

Finally, we are ready to prove the main lemma of this section.

Proof of Lemma 4.4. Suppose that ApproxCharacterizationTest fails with probability at most $2/3$, then by Lemma 4.20, g is pointwise $2^{(2n)^{45d}}R^d\delta$-close to a degree-d polynomial in $\mathrm{B}(\mathbf{0}, R)$. It remains to bound $\Pr[|g(\boldsymbol{p}) - \text{ApproxQuery-}g(\boldsymbol{p})| > (12R/r)^d 2^{d+4}\delta]$, where $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$. In the **YES** case, f is point-wise α-close to a degree-d polynomial h, and so for any $\boldsymbol{p}, \boldsymbol{q}$

$$|\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q})| = |\sum_{i=0}^{d+1} \alpha_i \cdot (f(\boldsymbol{p}+i\boldsymbol{q}) - h(\boldsymbol{p}+i\boldsymbol{q}))| \le \sum_{i=0}^{d+1} |\alpha_i| \cdot \alpha \le 2^{d+1}\alpha = \delta.$$

Therefore, ApproxCharacterizationTest always passes, and ApproxQuery-$g(\boldsymbol{p})$ returns a value that is $2^{d+3}\delta$-close to $g(\boldsymbol{p})$. Assume that f is not a degree-d polynomial. To query g on a point $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$, ApproxQuery-$g(\boldsymbol{p})$ attempts to obtain $d+1$ points on the line segment $\mathrm{L}^{\mathrm{B}}_{\mathbf{0},\boldsymbol{p}}$ and then interpolate g along this line. For these points $\boldsymbol{s} \in \{c_k\boldsymbol{p}\}_{k=0}^d$, ApproxQuery-$g$-InBall($\boldsymbol{s}$) samples an additional $N'_4 = O(\log(1/\varepsilon))$ points $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N'_4} \sim \mathcal{N}(\mathbf{0}, I)$, and checks whether

$$\Big|\sum_{i\in[d+1]} \alpha_i \cdot f(\boldsymbol{s}+i\boldsymbol{q}_1) - \sum_{i\in[d+1]} \alpha_i \cdot f(\boldsymbol{s}+i\boldsymbol{q}_j)\Big| \le 2^{d+2}\delta,$$

for all $j \in [N'_4]$; it rejects if any of these checks fail. By the definition of $g_{\boldsymbol{q}}$, this is equivalent to checking whether $|g_{\boldsymbol{q}_1}(\boldsymbol{s}) - g_{\boldsymbol{q}_j}(\boldsymbol{s})| > 2^{d+2}\delta$; by Lemma 4.8, this occurs with probability at most $1/(7d)$, since $\boldsymbol{s} \in \mathrm{B}(\mathbf{0}, r)$. The probability that ApproxQuery-g-InBall($\boldsymbol{s}$) doesn't reject, yet $|g(\boldsymbol{s}) -$ ApproxQuery-g-InBall($\boldsymbol{s}$)$| > 2^{d+4}\delta$, is the probability that: $|g(\boldsymbol{s}) - g_{\boldsymbol{q}_1}(\boldsymbol{s})| > 2^{d+4}\delta$, and $|g_{\boldsymbol{q}_1}(\boldsymbol{s}) - g_{\boldsymbol{q}_j}(\boldsymbol{s})| \le 2^{d+3}\delta$, (and therefore $|g(\boldsymbol{s}) - g_{\boldsymbol{q}_j}(\boldsymbol{s})| > 2^{d+3}\delta$) for every $\boldsymbol{q}_j$. By Corollary 4.9, this probability is at most $(7d)^{-N'_4} < 2^{-N'_4}/d^{N'_4} \le \varepsilon/(4(d+1))$, where the final inequality follows by choosing $N'_4 = O(\log(1/\varepsilon))$.

As APPROXQUERY-$g(\boldsymbol{p})$ approximately recovers the value of g on points $\{c_i\boldsymbol{p}\}_{i=0}^{d}$, we have that for every $i \in \{0,\dots,d\}$,

$$\Pr[|g_{\mathbf{0},\boldsymbol{p}}(c_i) - \text{APPROXQUERY-}g\text{-INBALL}(c_i\boldsymbol{p})| > 2^{d+4}\delta] \le \frac{\varepsilon}{4(d+1)}.$$

Thus, by Lemma 4.19 and a union bound over i,

$$\Pr[|g(\boldsymbol{p}) - \text{APPROXQUERY-}g(\boldsymbol{p})| > (12R/r)^d 2^{d+4}\delta] = \Pr[|g(\boldsymbol{p}) - g_{\mathbf{0},\boldsymbol{p}}(1)| > (12R/r)^d 2^{d+4}\delta]$$
$$\le \sum_{i=0}^{d} \Pr[|g_{\mathbf{0},\boldsymbol{p}}(c_i) - \text{APPROXQUERY-}g\text{-INBALL}(c_i\boldsymbol{p})| > 2^{d+4}\delta] \le \frac{\varepsilon}{4}. \qquad \square$$

5 Exact Testing over Discrete Domains

In this section we show that the test for degree-d polynomials from Section 3 can be modified to work for (sufficiently dense) discrete domains. The main theorem of this section is as follows:

Theorem 1.3. *For $d, B, R > 0$, let $B' \ge 16\max\{n^{5/2+2d}d^{2d}, B^2R^2/\sqrt{n}\}$ be a multiple of B. Let $\mathcal{L} = \frac{1}{B}\mathbb{Z}^n$ and $\mathcal{L}' = \frac{1}{B'}\mathbb{Z}^n$. Given $\varepsilon > 0$, query access to a function $f : \mathbb{R}^n \to \mathbb{R}$, and sample access to an unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$ supported on $\mathcal{L}$, there is a one-sided error, $O(d^5 + \frac{d^2}{\varepsilon}\log\frac{1}{\varepsilon})$-query tester for testing whether f agrees with a degree-d polynomial on $\mathcal{L}$, or is ε-far from degree-d polynomials over $\mathcal{D}$. The tester queries f on points in $\mathcal{L}'$.*

The key idea behind our tester is to define the self-corrected function g relative to a *discretized* Gaussian distribution defined over $\mathcal{L}$.

Definition 5.1. Given a lattice $\mathcal{L}$, and any $s > 0$, the *discrete Gaussian* $\mathcal{G}(\mathcal{L}, s)$ is the probability distribution over $\mathcal{L}$ such that the probability of drawing $\boldsymbol{x} \in \mathcal{L}$ is $\propto \tau_s(\boldsymbol{x}) \triangleq \exp(-\pi\|\boldsymbol{x}\|^2/s^2)$. (If unspecified, $s = 1$.)

That we are able to efficiently sample from a discrete Gaussian is guaranteed by the following lemma.

Lemma 5.2 (Lemma 2.3 in [BLP+13])**.** *There is a probabilistic polynomial time algorithm that given a positive integer B and parameter $r \ge \Omega(\sqrt{\log n}/B)$, outputs a sample distributed according to $\mathcal{G}(\frac{1}{B}\mathbb{Z}^n, r)$.*

At a high-level, the design of our tester will follow the same strategy as the design of our exact tester from Section 3, with several modifications to handle the lattice $\mathcal{L}$. From our unknown function f, we will define a self-corrected function g such that we have query access to g, and such that if our tests pass with sufficiently high probability then g is a degree-d polynomial on $\mathcal{L}$, and equals f on $\mathcal{L}$ if f is itself a degree-d polynomial.

As before, we define g on points within a small ball $\mathrm{B}(\mathbf{0}, r)$; for points $\boldsymbol{p} \in \mathcal{L} \setminus \mathrm{B}(\mathbf{0}, r)$ we will define their value by extrapolating the value of g within $\mathrm{B}(\mathbf{0}, r)$ by choosing $d+1$ points within $\mathrm{B}(\mathbf{0}, r) \cap \mathcal{L}$ along the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$, using them to interpolating a degree-d univariate polynomial $p_{\boldsymbol{p}}$, and then using $p_{\boldsymbol{p}}$ to define the value of $g(\boldsymbol{p})$ (see Figure 4). In order to certify that g is indeed a degree-d polynomial, we will use the following variant of the Local Characterization Theorem; a proof of which is given in Appendix A.

Discrete Local Characterization Theorem. *Fix $a > 0$, $M \ge d+1$, and let $S \triangleq \{\frac{ia}{M} : i \in \mathbb{N}\}$. If $f : [0, a] \to \mathbb{R}$ is a univariate function such that $\Delta_{\frac{a}{M}}^{(d+1)}[f](x) = 0$, for every $x \in S \cap [0, a]$ satisfying $x + (d+1)\frac{a}{M} \in [0, a]$, then f agrees with a degree-d polynomial over the points in $S \cap [0, a]$.*

However, for an arbitrary point $\boldsymbol{p} \in \mathcal{L} \setminus \mathrm{B}(\mathbf{0}, r)$ there may not be $d+1$ points on the line segment $\mathrm{L}_{\mathbf{0},\boldsymbol{p}} \cap \mathcal{L} \cap \mathrm{B}(\mathbf{0}, r)$ — this can occur if $\boldsymbol{p}$ is sufficiently far away from $\mathbf{0}$ — and thus we cannot define $p_{\boldsymbol{p}}$. To remedy this, we make two modifications. First, we assume that our distribution is $(\varepsilon/4, R)$-concentrated — that $1 - \varepsilon/4$ fraction of the mass of the unknown distribution $\mathcal{D}$ is in $\mathrm{B}(\mathbf{0}, R)$. We define g only on points within $\mathrm{B}(\mathbf{0}, R)$ and we will not test whether f differs from a degree-d polynomial outside of $\mathrm{B}(\mathbf{0}, R)$; as these point constitute only a small $\varepsilon/4$ portion of $\mathcal{D}$, which we can simply fold into the error of our tester. Second, in order to ensure that for any point $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$, g is defined on at least $d+1$ points on the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$ within $\mathrm{B}(\mathbf{0}, r)$, we define g on a finer lattice $\mathcal{L}' \triangleq \frac{r}{(d+1)R}\mathcal{L} = \frac{r}{BR(d+1)}\mathbb{Z}^n$ within $\mathrm{B}(\mathbf{0}, r)$.

The Self-Corrected Function. Let $R > r > 0$, and let our (unknown) $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$ be supported over a given lattice $\mathcal{L} = \frac{1}{B}\mathbb{Z}^n$. Let $\mathcal{L}' \triangleq \frac{r}{(d+1)R}\mathcal{L}$ be a refinement of $\mathcal{L}$. We define the self-corrected function g, whose domain is $\mathcal{L} \cup (\mathcal{L}' \cap B(\mathbf{0}, r))$, as follows. Let $\alpha_i \triangleq (-1)^{i+1}\binom{d+1}{i}$, and for any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r) \cap \mathcal{L}'$, and $\boldsymbol{q} \in \mathcal{L}$, let $g_{\boldsymbol{q}}(\boldsymbol{p}) \triangleq \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q})$. For any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r) \cap \mathcal{L}'$, we define

$$g(\boldsymbol{p}) \triangleq \operatorname*{maj}_{\boldsymbol{q} \sim \mathcal{G}(\mathcal{L},1)} [g_{\boldsymbol{q}}(\boldsymbol{p})].$$

For points $\boldsymbol{p} \in (B(\mathbf{0}, R) \setminus \mathrm{B}(\mathbf{0}, r)) \cap \mathcal{L}$ we define the value of $g(\boldsymbol{p})$ by interpolating a degree-d univariate polynomial along the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$ as follows: Let $c_0, \ldots, c_d \in \mathbb{R}$ be $d+1$ "distinguished" points (arbitrary, but fixed) on the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}}$ within $\mathrm{B}(\mathbf{0}, r) \cap \frac{r}{R(d+1)}\mathcal{L}$; in Algorithm 6 we choose $c_i = ir/((d+1)\|\boldsymbol{p}\|_2)$ and note that these points lie within $\mathcal{L}'$. Let $p_{\boldsymbol{p}}$ be the unique univariate polynomial such that $p_{\boldsymbol{p}}(c_i) = g(c_i\boldsymbol{p})$ for every $i \in [d+1]$. We define $g(\boldsymbol{p}) \triangleq p_{\boldsymbol{p}}(1)$.

Our tester is given in Algorithm 5, with corresponding subroutines in Algorithm 6.

Algorithm 5: Low-Degree Discrete Tester

Procedure DISCRETELOWDEGREETESTER($f, d, \mathcal{D}, \varepsilon, R, B$)

Given : Query access to f, a degree $d \in \mathbb{N}$, sampling access to an $(\varepsilon/4, R)$-concentrated unknown distribution $\mathcal{D}$ supported over the lattice $\mathcal{L} \triangleq \frac{1}{B}\mathbb{Z}^n$, where ε is the farness parameter, and B is the density parameter.

Reject if DISCRETECHARACTERIZATIONTEST **rejects**;

for $N_5 \leftarrow O(\varepsilon^{-1})$ *times* **do**

  Sample $\boldsymbol{p} \sim \mathcal{D}$;

  if $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$ **then**

    Reject if $f(\boldsymbol{p}) \neq$ DISCRETEQUERY-$g(\boldsymbol{p})$ or if DISCRETEQUERY-$g(\boldsymbol{p})$ **rejects**.

Accept.

In the remainder of this section we will prove Theorem 1.3. However, before we are able to do so, we require several structural results about Lattices and discrete Gaussians, which will occupy the next subsection.

5.1 Preliminaries on Lattices and Discrete Gaussians

First, we recall that many of the properties of Gaussian distributions translate over to their discrete variants.

Fact 5.3 (Fact 2 in [AGHS13])**.** *Suppose $\mathcal{L}$ is a lattice, and $s > 0$ is a parameter. If $\boldsymbol{x}$ is distributed as $\mathcal{G}(\mathcal{L}, s)$, then for any integer t, $t\boldsymbol{x}$ is distributed as $\mathcal{G}(t\mathcal{L}, ts)$.*

Algorithm 6: Discrete Subroutines

Procedure DISCRETECHARACTERIZATIONTEST
  $N_6 \leftarrow O(d^2)$;
  for N_6 *times* **do**
    for $j \in \{1, \ldots, d+1\}$ **do**
      for $t \in \{0, \ldots, d+1\}$ **do**
        Sample $\boldsymbol{p} \sim \mathcal{G}(j\mathcal{L}, j\sqrt{t^2+1}), \boldsymbol{q} \sim \mathcal{G}(\mathcal{L}, 1)$; ▷ [$j^2(t^2+1)$ vs. 1 Test.]
        Reject if $\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0$;
        Sample $\boldsymbol{p} \sim \mathcal{G}(j\mathcal{L}, j), \boldsymbol{q} \sim \mathcal{G}(\mathcal{L}, \sqrt{t^2+1})$; ▷ [$j^2$ vs. (t^2+1) Test.]
        Reject if $\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0$;
      Sample $\boldsymbol{p}, \boldsymbol{q} \sim \mathcal{G}(j\mathcal{L}, j)$; ▷ [$j^2$ vs. j^2 Test.]
      Reject if $\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}) \neq 0$;
  Accept;
Procedure DISCRETEQUERY-$g(\boldsymbol{p})$
  $r \leftarrow d\sqrt{n}/(2B)$;
  if $\boldsymbol{p} \in \mathrm{B}(\boldsymbol{0}, r)$ **then**
    return DISCRETEQUERY-g-INBALL$(\boldsymbol{p})$;
  for $i \in [d+1]$ **do**
    $c_i \leftarrow \frac{ir}{(d+1)\|\boldsymbol{p}\|_2}$;
    $v(c_i) \leftarrow$ DISCRETEQUERY-g-INBALL$(c_i\boldsymbol{p})$;
  Let $p_{\boldsymbol{p}} \colon \mathbb{R} \to \mathbb{R}$ be the unique degree-d polynomial such that $p_{\boldsymbol{p}}(i) = v(c_i)$ for $i \in [d+1]$;
  return $p_{\boldsymbol{p}}(1)$;
Procedure DISCRETEQUERY-g-INBALL$(\boldsymbol{p})$
  $N_6' \leftarrow O(\log \frac{1}{\varepsilon})$;
  Sample $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N_6'} \sim \mathcal{G}(\frac{r}{R(d+1)}\mathcal{L}, 1)$;
  Reject if there exists $j \in \{2, \ldots, N_6'\}$ such that $\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1) \neq \sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_j)$;
  return $\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1)$;

We implicitly use this fact to sample random vectors from scaled discrete gaussians in Algorithm 6. Next, we record a bound on the total variation distance between two Gaussians which is analogous to Lemma 2.1. To state the theorem, we need the following smoothing parameter defined in [MR07] as a parameter of a lattice with the following property: if one picks a noise vector from a Gaussian distribution with radius at least as large as the smoothing parameter, and reduces the noise vector modulo the fundamental parallelopiped of the lattice, then the resulting distribution is very close to uniform.

Definition 5.4. For a lattice $\mathcal{L} \subseteq \mathbb{R}^n$ and a parameter $\vartheta > 0$, the *smoothing parameter* $\eta_\vartheta(\mathcal{L})$ is the smallest $s > 0$ such that:

$$\tau_{1/s}\left(\mathcal{L}^* \setminus \{\boldsymbol{0}\}\right) \triangleq \sum_{\boldsymbol{x} \in \mathcal{L}^* \setminus \{\boldsymbol{0}\}} \tau_{1/s}(\boldsymbol{x}) = \sum_{\boldsymbol{x} \in \mathcal{L}^* \setminus \{\boldsymbol{0}\}} \exp\left(-\pi \|\boldsymbol{x}\|^2 s^2\right) \leq \vartheta,$$

where τ is defined in Definition 5.1, and $\mathcal{L}^* \triangleq \{\boldsymbol{x} \in \mathrm{span}(\mathcal{L}) : \forall \boldsymbol{y} \in \mathcal{L}, \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{Z}\}$ is the *dual lattice* of $\mathcal{L}$.

Observation 5.5. For any lattice $\mathcal{L}$, parameter $\vartheta > 0$, and $i \in \mathbb{Z}_{\geq 2}$, $(i\mathcal{L})^* = \frac{1}{i}\mathcal{L}^*$, and

$$\eta_\vartheta(i\mathcal{L}) = \min_{s>0}\left\{ \sum_{\boldsymbol{x}\in(i\mathcal{L})^*\setminus\{\mathbf{0}\}} \exp\left(-\pi\|\boldsymbol{x}\|^2 s^2\right)\right\} = \min_{s>0}\left\{ \sum_{\boldsymbol{x}\in\mathcal{L}^*\setminus\{\mathbf{0}\}} \exp\left(-\pi\|\boldsymbol{x}\|^2 s^2/i^2\right)\right\} \leq i^2\eta_\vartheta(\mathcal{L}).$$

The next theorem follows from (a quantified version of) Theorem 3.3 of [MP13], which we prove in Appendix C. Since ϑ is taken as a negligible function of n, so for small k, $4k\vartheta \ll 1$. Later, while invoking it, we will set $k \leq d+2$, which satisfies the restriction on k, and ϑ contained therein.

Theorem 5.6. *Let $k \in \mathbb{Z}_{>0}$, $\vartheta \in \mathbb{R}_{>0}$ be such that $k \leq 1/(4\vartheta)$. Let $\mathcal{L} \subseteq \mathbb{R}^n$ be a lattice, and let $\boldsymbol{s} \in \mathbb{R}^k_{>0}$ and $\boldsymbol{z} \in \mathbb{Z}^k$ be such that $s_i \geq \sqrt{2}\|\boldsymbol{z}\|_\infty\eta_\vartheta(\mathcal{L})$ for every $i \in [k]$. If $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ are sampled independently from $\boldsymbol{y}_i \sim \mathcal{G}(\mathcal{L}, s_i)$ then the distribution of $\sum_{i=1}^k z_i\boldsymbol{y}_i$ is $4k\vartheta$-close in total variation distance to $\mathcal{G}(\gcd(\boldsymbol{z})\mathcal{L}, \sqrt{\sum_{i=1}^k (z_i s_i)^2})$.*

Next, we recall a simple bound on the total variation distance incurred by shifting the center of a discrete Gaussian. Denote by $\mathrm{erf}(\cdot)$, the Gaussian error function.

Lemma 5.7 (Remark from Lemma 6 of [AGHS13])**.** *For any full-rank lattice $\mathcal{L} \subseteq \mathbb{R}^n$, $\vartheta \in (0, 1/2)$, $c > 1$ and a parameter s such that $s > (1+2c)\eta_\vartheta(\mathcal{L})$, the following holds: The total variation distance between $\mathcal{G}(\mathcal{L}, s)$ and $\mathcal{G}(\boldsymbol{v}+\mathcal{L}, s)$ for any $\boldsymbol{v} \in \mathcal{L}$ is at most*

$$\frac{\mathrm{erf}(q)}{\mathrm{erf}(qc)} \cdot \frac{1+\vartheta}{1-\vartheta},$$

where $q = \|\boldsymbol{v}\|_2^2\sqrt{\pi}/s$.

By combining Lemma 5.7 and Observation 5.5 we obtain the following corollary, which bounds the distance between two discrete Gaussians.

Corollary 5.8. *Let $i \in [d+1]$, t be a non-negative integer, $\vartheta \in (0, 1/2)$, $r \leq d\sqrt{\eta_\vartheta(\mathcal{L})}$, and let $\mathcal{L} = \frac{1}{B}\mathbb{Z}^n$. Then, for any $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r) \cap (d+1)!\mathcal{L}$,*

$$\mathrm{d_{TV}}\left(\mathcal{G}(\boldsymbol{p}+i\mathcal{L}, it), \mathcal{G}(i\mathcal{L}, it)\right) \leq 98d^2\eta_\vartheta(\mathcal{L}).$$

Proof. To bound the total variation distance, we aim to apply Lemma 5.7. To do so, we need to choose c such that $it > (1+2c)\eta_\vartheta(i\mathcal{L})$. Observe that

$$(1+2c)\eta_\vartheta(i\mathcal{L}) \leq (1+2c)i^2\eta_\vartheta(\mathcal{L}) \leq (1+2c)(d+1)^2\eta_\vartheta(\mathcal{L}) < 12cd^2\eta_\vartheta(\mathcal{L}),$$

where the first inequality follows by Observation 5.5. Thus, letting $c \triangleq t(12d^2\eta_\vartheta(\mathcal{L}))^{-1}$ we have $(1+2c)\eta_\vartheta(i\mathcal{L}) < t \leq it$. Applying Lemma 5.7,

$$\mathrm{d_{TV}}(\mathcal{G}(\boldsymbol{p}+i\mathcal{L}, it), \mathcal{G}(i\mathcal{L}, it)) \leq \frac{(1+\vartheta)\mathrm{erf}(\|\boldsymbol{p}\|_2^2\sqrt{\pi}/it)}{(1-\vartheta)\mathrm{erf}(c\|\boldsymbol{p}\|_2^2\sqrt{\pi}/it)}.$$

Note that because $\mathrm{erf}(x) \triangleq \Pr_{y\sim\mathcal{N}(0,1/2)}[y \in [-x, x]]$ and the PDF of $\mathcal{N}(0, 1/2)$ is $\frac{1}{\sqrt{\pi}}e^{-x^2}$, we have $\frac{2x}{\sqrt{\pi}}e^{-x^2} \leq \mathrm{erf}(x) \leq \frac{2x}{\sqrt{\pi}}$. This means that for any $q > 0$, $\frac{e^{-q^2}}{c} \leq \frac{\mathrm{erf}(q)}{\mathrm{erf}(qc)} \leq \frac{e^{c^2q^2}}{c}$, and so if $q \leq e/c$, it holds that $\frac{\mathrm{erf}(q)}{\mathrm{erf}(qc)} \leq \frac{e}{c}$. Because $\boldsymbol{p} \in B(\mathbf{0}, r)$, $q \triangleq \|\boldsymbol{p}\|_2^2\sqrt{\pi}/it \leq r^2\sqrt{\pi}/it \leq e/c$ by our choice of c and r. Thus,

$$\frac{(1+\vartheta)\mathrm{erf}(\|\boldsymbol{p}\|_2^2\sqrt{\pi}/it)}{(1-\vartheta)\mathrm{erf}(c\|\boldsymbol{p}\|_2^2\sqrt{\pi}/it)} \leq \frac{1+\vartheta}{1-\vartheta}(e/c) \leq 3e/c = 98d^2\eta_\vartheta(\mathcal{L}),$$

where the second inequality follows because $\vartheta \in (0, 1/2)$. □

5.1.1 Correctness of the Discrete Low Degree Tester

We are now ready to prove Theorem 1.3. In fact, we prove a more general result, which handles lattices parameterized by their smoothness parameter. We state this generalization next.

Theorem 5.9. *Let $R > r > 0$ and $B', d > 0$ satisfy $rB' > (d+1)!n^{3/2+d}$. Let $\vartheta \le (192d)^{-1}$ be such that $\eta_\vartheta(\frac{1}{B'}\mathbb{Z}^n) \le (6d)^{-4}$, and $r \le d\sqrt{\eta_\vartheta(\frac{1}{B'}\mathbb{Z}^n)}$. Let $\mathcal{L} = \frac{R(d+1)}{rB'}\mathbb{Z}^n$, $\mathcal{L}' = \frac{1}{B'}\mathbb{Z}^n$, $\varepsilon > 0$, and $f : \mathbb{R}^n \to \mathbb{R}$. There is a one-sided error, $O(d^5 + \frac{d^2}{\varepsilon}\log\frac{1}{\varepsilon})$-query tester for testing whether f agrees with a degree-d polynomial on $\mathcal{L}$ with respect to an $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$ supported over $\mathcal{L}$.*

Theorem 1.3 follows by letting $B = \frac{r}{R(d+1)}B'$, setting $\vartheta = 2^{-n}$, $r = \frac{(d+1)n^{1/4}}{4\sqrt{B'}}$, and $B' > 16n^{5/2+2d}\cdot d^{2d}$. To see that the inequalities in the statement of Theorem 5.9 are satisfied, we use the following bound which can be found in [MR07]:

$$\frac{\sqrt{n/\pi}}{B} \le \eta_{2^{-n}}\left(\frac{1}{B}\mathbb{Z}^n\right) \le \frac{\sqrt{n}}{B}.$$

The following lemma records the properties of g which are guaranteed by our tester.

Lemma 5.10. *Assume that the conditions of Theorem 5.9 hold. If* DiscreteCharacterizationTest *fails with probability at most $2/3$, then g consistent with a degree-d polynomial within $\mathrm{B}(\mathbf{0}, R) \cap \mathcal{L}$, and furthermore for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R) \cap \mathcal{L}$, $g(\boldsymbol{p}) =$* DiscreteQuery-$g(\boldsymbol{p})$ *with probability at least $1 - \frac{\varepsilon}{4}$.*

We prove the main theorem assuming that this lemma holds.

Proof of Theorem 5.9. The proof follows the same argument as the proof of Theorem 1.1, with two small changes. First, we use Lemma 5.10 in place of Lemma 3.1. Second, since we only test points within $\mathrm{B}(\mathbf{0}, R)$, we err on those points which are not. However, since $\mathcal{D}$ is $(\varepsilon/4, R)$-concentrated, the probability mass of these points is at most $\varepsilon/4$, and this is folded into the error probability of our tester.

□

5.2 Polynomial Representation on Lines Within a Small Ball

We turn now to proving Lemma 5.10; this will be done over the following three subsections. First, we prove that g is consistent with a degree-d polynomial on every line within $\mathrm{B}(\mathbf{0}, r)$.

Theorem 5.11. *(Polynomial representation on lines) Let $B, r > 0$ be such that $(rB/(d+1)!n^{3/2+d})^n > 1$. Let $\vartheta \le (192d)^{-1}$ and $\mathcal{L} = \frac{1}{B}\mathbb{Z}^n$ be a lattice such that $\eta_\vartheta(\mathcal{L}) \le (6d)^{-4}$. If* DiscreteCharacterizationTest *fails with probability at most $2/3$, then for any $\boldsymbol{a}, \boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r) \cap (d+1)!\mathcal{L}$, there is a degree-$d$ univariate polynomial which is consistent with $g_{\boldsymbol{a},\boldsymbol{b}}(x) = g(\boldsymbol{a} + x\boldsymbol{b})$ on every point x such that $\boldsymbol{a} + x\boldsymbol{b} \in \mathrm{B}(\mathbf{0}, r) \cap (d+1)!\mathcal{L}$.*

The proof of Theorem 5.11 follows by exactly the same argument as the proof of Theorem 3.2, using the Discrete Local Characterization Theorem, and Lemma 5.12 in place of Lemma 3.3.

Lemma 5.12. *Let $B, r > 0$ be such that $(rB/(d+1)!n^{3/2+d})^n > 1$. Let $\vartheta \le (64(d+2))^{-1}$ and $\mathcal{L} = \frac{1}{B}\mathbb{Z}^n$ be a lattice such that $\eta_\vartheta(\mathcal{L}) \le (6d)^{-4}$. If* DiscreteCharacterizationTest *fails with probability at most $2/3$, then for every $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r) \cap (d+1)!\mathcal{L}$ such that $\boldsymbol{p} + i\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r) \cap (d+1)!\mathcal{L}$ for every $i \in [d+1]$, we have $\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + i\boldsymbol{q}) = 0$.*

In the remainder of this subsection, we will prove Lemma 5.12.

Let ρ denote the bound of the probability that each of the tests in the DISCRETECHARACTERIZATIONTEST fails. That is, for every $j \in \{1, \dots, d+1\}$ and $t \in \{0, \dots, d+1\}$, the following are bounded:

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{G}(j\mathcal{L}, j\sqrt{t^2+1}) \\ \boldsymbol{q}\sim\mathcal{G}(\mathcal{L},1)}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q}) \neq 0\right] \leq \rho. \qquad [j^2(t^2+1) \text{ vs. } 1 \text{ Test.}] \tag{18}$$

$$\Pr_{\substack{\boldsymbol{p}\sim\mathcal{G}(j\mathcal{L}, j) \\ \boldsymbol{q}\sim\mathcal{G}(\mathcal{L},\sqrt{t^2+1})}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q}) \neq 0\right] \leq \rho. \qquad [j^2 \text{ vs. } (t^2+1) \text{ Test.}] \tag{19}$$

$$\Pr_{\boldsymbol{p},\boldsymbol{q}\sim\mathcal{G}(j\mathcal{L}, j)} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q}) \neq 0\right] \leq \rho. \qquad [j^2 \text{ vs. } j^2 \text{ Test.}] \tag{20}$$

We first bound ρ, by an identical argument, used earlier in Claim 3.5 to bound ρ in the exact case:

Claim 5.13. If DISCRETECHARACTERIZATIONTEST fails with probability at most $2/3$, then $\rho \leq (4d)^{-2}$.

Then we bound the probability that $g_{\boldsymbol{q}_1}$ and $g_{\boldsymbol{q}_2}$ differ, in the intersection of $\mathrm{B}(\mathbf{0}, r)$ with $(d+1)!\mathcal{L}$:

Lemma 5.14. *Let $\mathcal{L}$ denote the lattice $\frac{1}{B}\mathbb{Z}^n$. Let $\eta_\vartheta \in (0, 1/2)$, and $r \leq d\sqrt{\eta_\vartheta(\mathcal{L})}$. Then, for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r) \cap (d+1)!\mathcal{L}$ and every $t \in S \triangleq \{\sqrt{i^2+1} : i \in \{0, \dots, d+1\}\}$,*

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{G}(\mathcal{L},t) \\ \boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}} [g_{\boldsymbol{q}_1}(\boldsymbol{p}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{p})] \leq 2(d+1)\left(\rho + 196d^2\eta_\vartheta(\mathcal{L})\right).$$

Proof. We follow the proof of Lemma 3.6. Fix $\boldsymbol{p}$, t as in the statement of the lemma. For $i \in [d+1]$, will bound the following probability

$$\begin{aligned}
&\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{G}(\mathcal{L},t) \\ \boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}} \left[f(\underbrace{\boldsymbol{p}+i\boldsymbol{q}_1}_{\triangleq \boldsymbol{m}}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{p}+i\boldsymbol{q}_1)\right] = \Pr_{\substack{\boldsymbol{m}\sim\mathcal{G}(\boldsymbol{p}+i\mathcal{L}, it) \\ \boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}} \left[f(\boldsymbol{m}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{m})\right], \\
&\leq \Pr_{\substack{\boldsymbol{m}\sim\mathcal{G}(i\mathcal{L}, it) \\ \boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}} \left[\sum_{j=0}^{d+1} \alpha_i \cdot f(\boldsymbol{m}+j\boldsymbol{q}_2) \neq 0\right] + 2\,\mathrm{d_{TV}}(\mathcal{G}(\boldsymbol{p}+i\mathcal{L}, it), \mathcal{G}(i\mathcal{L}, it)), \quad \text{(By definition of } g_{\boldsymbol{q}_2}(\boldsymbol{m})) \\
&\leq \rho + 196d^2\eta_\vartheta(\mathcal{L}),
\end{aligned}$$

where the bound on the second term follows from Corollary 5.8, and the first term is bounded by the rejection probability ρ, by (18). Note that we can indeed apply Corollary 5.8, since $\boldsymbol{p} \in (d+1)!\mathcal{L}$ guarantees that $\boldsymbol{p} \in i\mathcal{L}$ for every $i \in [d+1]$.

By the same argument as above, with (18) replaced by (19), for any $j \in [d+1]$, we can bound

$$\begin{aligned}
&\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{G}(\mathcal{L},t) \\ \boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}} \left[f(\underbrace{\boldsymbol{p}+j\boldsymbol{q}_2}_{\triangleq \boldsymbol{m}}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{p}+j\boldsymbol{q}_2)\right] = \Pr_{\substack{\boldsymbol{m}\sim\mathcal{G}(\boldsymbol{p}+j\mathcal{L}, j) \\ \boldsymbol{q}_1\sim\mathcal{G}(\mathcal{L},t)}} \left[f(\boldsymbol{m}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{m})\right], \\
&\leq \Pr_{\substack{\boldsymbol{m}\sim\mathcal{G}(j\mathcal{L}, j) \\ \boldsymbol{q}_1\sim\mathcal{G}(\mathcal{L},t)}} \left[\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{m}+i\boldsymbol{q}_1) \neq 0\right] + 2\,\mathrm{d_{TV}}(\mathcal{G}(\boldsymbol{p}+j\mathcal{L}, j), \mathcal{G}(j\mathcal{L}, j)), \quad \text{(By definition of } g_{\boldsymbol{q}_1}(\boldsymbol{m})) \\
&\leq \rho + 196d^2\eta_\vartheta(\mathcal{L}).
\end{aligned}$$

Taking a union bound over all $i, j \in [d+1]$ gives

$$\Pr_{\substack{\boldsymbol{q}_1 \sim \mathcal{G}(\mathcal{L},t) \\ \boldsymbol{q}_2 \sim \mathcal{G}(\mathcal{L},1)}} \left[\underbrace{\sum_{i=1}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1)}_{= g_{\boldsymbol{q}_1}(\boldsymbol{p})} \neq \sum_{i=1}^{d+1} \sum_{j=1}^{d+1} \alpha_i \alpha_j \cdot f((\boldsymbol{p} + i\boldsymbol{q}_1) + j\boldsymbol{q}_2) \right] \leq (d+1) \left(\rho + 196 d^2 \eta_\vartheta(\mathcal{L}) \right),$$

$$\Pr_{\substack{\boldsymbol{q}_1 \sim \mathcal{G}(\mathcal{L},t) \\ \boldsymbol{q}_2 \sim \mathcal{G}(\mathcal{L},1)}} \left[\underbrace{\sum_{j=1}^{d+1} \alpha_j \cdot f(\boldsymbol{p} + j\boldsymbol{q}_2)}_{= g_{\boldsymbol{q}_2}(\boldsymbol{p})} \neq \sum_{j=1}^{d+1} \sum_{i=1}^{d+1} \alpha_i \alpha_j \cdot f((\boldsymbol{p} + j\boldsymbol{q}_2) + i\boldsymbol{q}_1) \right] \leq (d+1) \left(\rho + 196 d^2 \eta_\vartheta(\mathcal{L}) \right).$$

Thus, by a final union bound, we can conclude that

$$\Pr_{\substack{\boldsymbol{q}_1 \sim \mathcal{G}(\mathcal{L},t) \\ \boldsymbol{q}_2 \sim \mathcal{G}(\mathcal{L},1)}} [g_{\boldsymbol{q}_1}(\boldsymbol{p}) \neq g_{\boldsymbol{q}_2}(\boldsymbol{p})] \leq 2(d+1) \left(\rho + 196 d^2 \eta_\vartheta(\mathcal{L}) \right). \qquad \square$$

An immediate corollary is the following.

Corollary 5.15. *If* DISCRETECHARACTERIZATIONTEST *fails with probability at most* $2/3$*, and* $\eta_\vartheta(\mathcal{L}) \leq (6d)^{-4}$*, then for every* $\boldsymbol{p} \in \mathrm{B}(\boldsymbol{0}, r) \cap \frac{(d+1)!}{B}\mathbb{Z}^n$ *and every* $t \in \{0, \ldots, d+1\}$*,*

$$\Pr_{\boldsymbol{q} \sim \mathcal{G}\left(\mathcal{L}, \sqrt{t^2+1}\right)} [g(\boldsymbol{p}) \neq g_{\boldsymbol{q}}(\boldsymbol{p})] < \frac{1}{4(d+2)}.$$

Proof. By Claim 5.13, ρ is at most $(4d)^{-2}$. Observe, for any $t \in \{0, \ldots, d+1\}$,

$$\begin{aligned} \Pr_{\boldsymbol{q} \sim \mathcal{G}\left(\mathcal{L}, \sqrt{t^2+1}\right)} [g(\boldsymbol{p}) \neq g_{\boldsymbol{q}}(\boldsymbol{p})] &\leq \Pr_{\boldsymbol{q}_1 \sim \mathcal{G}(\mathcal{L},1)} [g(\boldsymbol{p}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{p})] + \Pr_{\substack{\boldsymbol{q} \sim \mathcal{G}\left(\mathcal{L}, \sqrt{t^2+1}\right) \\ \boldsymbol{q}_1 \sim \mathcal{G}(\mathcal{L},1)}} [g_{\boldsymbol{q}}(\boldsymbol{p}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{p})] \\ &\leq 4(d+1) \left(\rho + 196 d^2 \eta_\vartheta(\mathcal{L}) \right), \qquad \text{(By Lemma 5.14)} \end{aligned}$$

which is at most $(4(d+2))^{-1}$, since $\rho \leq (4d)^{-2}$, and by our assumptions on $\eta_\vartheta(\mathcal{L})$. $\square$

Finally, we are ready to prove Lemma 5.12, the discrete analog of Lemma 3.3.

Proof of Lemma 5.12. By Claim 5.13, ρ is at most $(4d)^{-2}$. By the same argument as in the proof of Lemma 3.3, it is sufficient to show that for $\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{G}(\mathcal{L}, 1)$, the following two events hold simultaneously with non-zero probability, for every $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\boldsymbol{0}, r) \cap (d+1)!\mathcal{L}$ such that $\boldsymbol{p} + i\boldsymbol{q} \in \mathrm{B}(\boldsymbol{0}, r) \cap (d+1)!\mathcal{L}$ for every $i \in [d+1]$:

$$\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + i\boldsymbol{q}) = \sum_{i=0}^{d+1} \alpha_i \cdot g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + i\boldsymbol{q}) \tag{21}$$

$$\sum_{i=0}^{d+1} \alpha_i \cdot f(\boldsymbol{p} + j\boldsymbol{q}_1 + i(\boldsymbol{q} + j\boldsymbol{q}_2)) = 0, \text{ for every } j \in [d+1]. \tag{22}$$

We begin with (21):

$$\Pr_{\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{G}(\mathcal{L},1)} \left[\sum_{i=0}^{d+1} \alpha_i \cdot g(\boldsymbol{p} + i\boldsymbol{q}) = \sum_{i=0}^{d+1} \alpha_i \cdot g_{\boldsymbol{q}_1 + i\boldsymbol{q}_2}(\boldsymbol{p} + i\boldsymbol{q}) \right]$$

$$
\begin{aligned}
&\geq \Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}\left[g(\boldsymbol{p}+i\boldsymbol{q}) = g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+i\boldsymbol{q}),\ \ \forall i\in\{0,\dots,d+1\}\right]\\
&=1-\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}\left[\exists i\in\{0,\dots,d+1\} : g(\boldsymbol{p}+i\boldsymbol{q}) \neq g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+i\boldsymbol{q})\right]\\
&\geq 1-\sum_{i=0}^{d+1}\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}\left[g(\boldsymbol{p}+i\boldsymbol{q}) \neq g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+i\boldsymbol{q})\right]\\
&=1-\sum_{i=0}^{d+1}\left(\Pr_{\boldsymbol{m}\sim\mathcal{G}\left(\mathcal{L},\sqrt{i^2+1}\right)}\left[g(\boldsymbol{p}+i\boldsymbol{q}) \neq g_{\boldsymbol{m}}(\boldsymbol{p}+i\boldsymbol{q})\right] + 2\,\mathrm{d_{TV}}\left(\boldsymbol{q}_1+i\boldsymbol{q}_2, \mathcal{G}\left(\mathcal{L},\sqrt{i^2+1}\right)\right)\right). \qquad (23)
\end{aligned}
$$

Since $\boldsymbol{p}+i\boldsymbol{q}\in \mathrm{B}(\boldsymbol{0},r)\cap(d+1)!\mathcal{L}$, the probability that $g(\boldsymbol{p}+i\boldsymbol{q})\neq g_{\boldsymbol{m}}(\boldsymbol{p}+i\boldsymbol{q})$ is at most $(4(d+2))^{-1}$, by Corollary 5.15. As well, we can bound the total variation distance by 8ϑ, by applying Theorem 5.6 with parameters $k=2$, $\boldsymbol{z}=(1,i)$, $\boldsymbol{s}=(1,1)$, and noting that $s_i = 1 \geq \sqrt{2}i\eta_\vartheta(\mathcal{L})$. Thus, (23) is at least

$$
1-\sum_{i=0}^{d+1}\left(\frac{1}{4(d+2)}+16\vartheta\right) \geq 1-\left(\frac{1}{4}+16(d+2)\vartheta\right) \geq \frac{1}{2}.
$$

Next, we bound (22). Fix some $j\in[d+1]$, then

$$
\begin{aligned}
&\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}\left[\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{p}+j\boldsymbol{q}_1+i(\boldsymbol{q}+j\boldsymbol{q}_2))\neq 0\right] && \text{(Let } \boldsymbol{z}_1\triangleq\boldsymbol{p}+j\boldsymbol{q}_1, \boldsymbol{z}_2\triangleq\boldsymbol{q}+j\boldsymbol{q}_2\text{)}\\
&\leq \Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{G}(\boldsymbol{p}+j\mathcal{L},j)\\ \boldsymbol{z}_2\sim\mathcal{G}(\boldsymbol{q}+j\mathcal{L},j)}}\left[\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{z}_1+i\boldsymbol{z}_2)\neq 0\right] + 2\Big(\underbrace{\mathrm{d_{TV}}(j\boldsymbol{q}_1,\mathcal{G}(j\mathcal{L},j))}_{=0,\text{ by Fact 5.3}} + \underbrace{\mathrm{d_{TV}}(j\boldsymbol{q}_2,\mathcal{G}(j\mathcal{L},j))}_{=0,\text{ by Fact 5.3}}\Big)\\
&\leq \Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{G}(j\mathcal{L},j)\\ \boldsymbol{z}_2\sim\mathcal{G}(j\mathcal{L},j)}}\left[\sum_{i=0}^{d+1}\alpha_i f(\boldsymbol{z}_1+i\boldsymbol{z}_2)\neq 0\right] + 2(\mathrm{d_{TV}}(\mathcal{G}(j\mathcal{L},j),\mathcal{G}(\boldsymbol{p}+j\mathcal{L},j)) + \mathrm{d_{TV}}(\mathcal{G}(j\mathcal{L},j),\mathcal{G}(\boldsymbol{q}+j\mathcal{L},j)))\\
&\leq \Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{G}(j\mathcal{L},j)\\ \boldsymbol{z}_2\sim\mathcal{G}(j\mathcal{L},j)}}\left[\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{z}_1+i\boldsymbol{z}_2)\neq 0\right] + 392d^2\eta_\vartheta(\mathcal{L}) && \text{(By Corollary 5.8)}\\
&\leq \rho + 392d^2\eta_\vartheta(\mathcal{L}). && \text{(By (20))}
\end{aligned}
$$

Finally, by a union bound over all $j\in[d+1]$, the probability of event (22) can be lower bounded,

$$
\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{G}(\mathcal{L},1)}\left[\forall j\in[d+1], \sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{p}+j\boldsymbol{q}_1+i(\boldsymbol{q}+j\boldsymbol{q}_2))=0\right] \geq 1-(d+1)(\rho+392d^2\eta_\vartheta(\mathcal{L})) \geq 1-o(1),
$$

where the last inequality follows by our assumptions on ρ and $\eta_\vartheta(\mathcal{L})$.

A final union bound shows that (21) and (22) hold simultaneously with non-zero probability, and the theorem follows. □

5.3 Polynomial Within a Hypercube

Next, we obtain the discrete analog of Theorem 3.2. We argue that if the conditions of Theorem 5.11 are met, then g is consistent with a degree-d polynomial on a hypercube $[-r',r']^n\subseteq \mathrm{B}(\boldsymbol{0},r)$; we take $r'=r/\sqrt{n}$ as this is a large hypercube which can be inscribed within the cube such that no point in $[-r',r']^n$ is extremal

on the cube, however other values of r' work as well. This is done in two steps: in Theorem 5.11 we argue that g is consistent with a polynomial of degree at most dn, and in Lemma 5.17 we reduce the degree to d. As before, let $\boldsymbol{e}_i$ denote the ith standard basis vector.

Lemma 5.16. *Assume that the assumptions of Theorem 5.11 hold. Let $r' = r/\sqrt{n}$, let $d, B > 0$ satisfy $2rB/\sqrt{n} > d!$ and let $h\colon [-r', r']^n \to \mathbb{R}$. Then, the following holds: If for every $i \in [n]$ and $\boldsymbol{a} \in [-r', r']^n$ such that $\boldsymbol{a}_i = 0$, the restriction of h to the line segment $L_{\boldsymbol{a},\boldsymbol{e}_i}$, the univariate function $h_{\boldsymbol{a},\boldsymbol{e}_i}$ is consistent with a degree-d univariate polynomial on every input x for which $\boldsymbol{a} + x\boldsymbol{e}_i \in [-r', r']^n \cap \frac{(d+1)!}{B}\mathbb{Z}^n$, then h agrees with an n-variate polynomial of degree at most dn on $[-r', r']^n \cap \frac{(d+1)!}{B}\mathbb{Z}^n$.*

Proof. Let $\mathrm{H} \triangleq [-r', r']^n \cap \frac{(d+1)!}{B}\mathbb{Z}^n$. We will prove the lemma by induction on the dimension n; the case $n = 1$ is immediate. Now, assume that the statement is true for $n-1$. Let $c_1, \ldots, c_{d+1} \in [-r', r'] \cap \frac{(d+1)!}{B}\mathbb{Z}$, be $d+1$ distinct values; note that these exist since $\frac{2r'B}{(d+1)!} > d+1$. For each $i \in \{1, \ldots, d+1\}$ consider the sub-cubes $\mathrm{H}_1, \ldots, \mathrm{H}_{d+1}$ of dimension $n-1$, defined as $\mathrm{H}_i \triangleq [-r', r']^{n-1} \cap \frac{(d+1)!}{B}\mathbb{Z}^{n-1} \times c_i$. Note that all the line segments in H_i are also contained in H and therefore, by assumption, h is consistent with degree-d *univariate* polynomials on them. Thus, we can apply the induction hypothesis to argue that h is consistent with degree-$d(n-1)$ polynomials h_i on each of the sub-cubes H_i. We will combine these polynomials to form an n-variate polynomial using the following degree-d polynomials: For every $i \in [d+1]$, δ_i's are defined as,

$$\delta_i(c_j) \triangleq \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

We argue that

$$h(\boldsymbol{x}) = \sum_{i=1}^{d+1} \underbrace{\delta_i(x_n)}_{\deg=d} \cdot \underbrace{h_i(x_1, \ldots, x_{n-1})}_{\deg=d(n-1)}$$

on the lattice points inside the cube. Fix $\boldsymbol{a} \in [-r', r']^{n-1} \cap \frac{(d+1)!}{B}\mathbb{Z}^{n-1}$ and let t be a formal variable. We claim that the following two *univariate* polynomials are equal,

$$h(a_1, \ldots, a_{n-1}, t) = \sum_{i=1}^{d+1} \delta_i(t) h_i(\boldsymbol{a}).$$

The left polynomial is of degree-d by assumption, while the right polynomial is of degree-d by definition of δ_i and the fact that $\boldsymbol{a}$ is fixed. Moreover, these two polynomial agree on $d+1$ points $(a_1, \ldots, a_{n-1}, i)_{i \in [d+1]}$, and therefore they are equal.

As equality holds for every such $(\boldsymbol{a}, t) \in [-r', r']^n$, h is a polynomial of degree at most dn within $[-r', r']^n$. □

Next, we argue that the degree of h is in fact at most d.

Lemma 5.17. *Assume that the assumptions of Theorem 5.11 hold. Let $r' = r/\sqrt{n}$, let $d, B, m > 0$ satisfy $\left(\frac{2r'B}{(d+1)!(m+1)}\right)^n > n^m$, and let $h : [-r', r']^n \to \mathbb{R}$ be a polynomial of finite degree m. If for every radial line in the cube $[-r', r']^n$, the restriction of h to that line agrees with a univariate polynomial of degree at most d on points in $\frac{(d+1)!}{B}\mathbb{Z}^n$, then $m \leq d$.*

Proof. Consider the coefficient representation of the polynomial $h(x\boldsymbol{b})$ in the formal variables $x \in \mathbb{R}$ and $\boldsymbol{b} \in \mathbb{R}^n$. This representation is a polynomial of degree m in both x and $\boldsymbol{b}$, as h is degree m. Consider α_m, the coefficient of the monomial with the highest degree (in x) in h as a polynomial in the formal variables $\boldsymbol{b}$.

$$\alpha_m(\boldsymbol{b}) \triangleq \sum_{i_1+\cdots+i_n=m} \alpha_{i_1,\ldots,i_n} \prod_{j=1}^{n} b_j^{i_j},$$

and note that $\alpha_m \neq 0$, as otherwise h would have degree less than m.

Now, fix $\boldsymbol{b}$ to some value $\boldsymbol{b}^* \in [-\frac{r'}{m+1}, \frac{r'}{m+1}]^n \cap \frac{(d+1)!}{B}\mathbb{Z}^n$ such that $\alpha_m(\boldsymbol{b}^*) \neq 0$; such a point exists since there are at most n^m roots of the polynomial α_m and, by assumption, there are at least

$$\left(\frac{2r'B}{(d+1)!(m+1)}\right)^n > n^m$$

many lattice points in the cube $[-\frac{r'}{m+1}, \frac{r'}{m+1}]^n$, and at least $m+1$ lattice points on the line segment $\mathrm{L}_{\mathbf{0},\boldsymbol{b}} \cap [-r', r']^n$. The univariate polynomial $h(x\boldsymbol{b}^*)$ in the formal variable x is of degree exactly m. By assumption, it is consistent with a univariate polynomial of degree at most d on all points in $\mathrm{L}_{\mathbf{0},\boldsymbol{b}^*} \cap \frac{(d+1)!}{B}\mathbb{Z}^n$. Since there are more than $m+1$ points on the line segment, these two polynomials are identical and therefore $m \leq d$. □

5.4 Global Polynomial Representation

Finally, we argue that if the conditions of Theorem 5.11 are met, then g is a degree-d polynomial.

Theorem 5.18. *Suppose that the assumptions of Theorem 5.11 hold, and let $R > r > 0$, and let $d, B > 0$ satisfy $(rB/(d+1)!n^{3/2+d})^n > 1$. If* DISCRETECHARACTERIZATIONTEST *fails with probability at most $2/3$, then g is a degree-d, n-variate polynomial on the lattice $\mathcal{L}' \triangleq \frac{(d+1)R}{rB}\mathbb{Z}^n$ within $\mathrm{B}(\mathbf{0}, R)$.*

Proof. The proof is identical to the proof of Lemma 3.10, using Lemma 5.16 and Theorem 5.11, noting that our choice of r, B, d satisfies the hypothesis of Lemma 5.17. Choosing $\mathcal{L}'$ to be a coarser lattice than $\mathcal{L}$ guarantees that for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$, there are at least $d+1$ points on the line $\mathrm{L}_{\mathbf{0},\boldsymbol{p}} \cap B(\mathbf{0}, R)$ on which to define the value of $g(\boldsymbol{p})$, and therefore g is well defined on $\mathcal{L}'$ within $\mathrm{B}(\mathbf{0}, R)$. □

Finally, we are ready to prove the main lemma, which concludes the proof of correctness for our tester.

Proof of Lemma 5.10. Suppose that DISCRETECHARACTERIZATIONTEST fails with probability at most $2/3$. Then, by Theorem 5.18, g is a degree-d, n-variate polynomial. It remains to bound the probability that $g(\boldsymbol{p}) \neq$ DISCRETEQUERY-$g(\boldsymbol{p})$ for $\boldsymbol{p} \in \mathbb{R}^n \cap \mathcal{L} \setminus \mathrm{B}(0, r)$. If f is itself a degree-d polynomial, then DISCRETEQUERY-$g(\boldsymbol{p})$ returns $g(\boldsymbol{p})$ with probability 1, so assume otherwise. To query g on a point $\boldsymbol{p} \in \mathbb{R}^n \cap \mathcal{L}$, DISCRETEQUERY-$g(\boldsymbol{p})$ obtains $d+1$ points on the line segment $\mathrm{L}^{\mathrm{B}}_{\mathbf{0},\boldsymbol{p}}$ and then interpolate g along this line. For each of these $d+1$ points $\boldsymbol{s}$, DISCRETEQUERY-g-INBALL($\boldsymbol{s}$) samples an additional $N_6' = O(\log(1/\varepsilon))$ points $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N_6'} \sim \mathcal{G}(\mathcal{L}, 1)$, and checks whether

$$\sum_{i \in [d+1]} \alpha_i \cdot f(\boldsymbol{s} + i\boldsymbol{q}_1) = \sum_{i \in [d+1]} \alpha_i \cdot f(\boldsymbol{s} + i\boldsymbol{q}_j),$$

for all $j \in [N_6']$; it rejects if any of these checks fail. This is equivalent to checking whether $g_{\boldsymbol{q}_1}(\boldsymbol{s}) \neq g_{\boldsymbol{q}_j}(\boldsymbol{s})$; by Corollary 5.15, this occurs with probability at most $1/(4(d+2))$, since $\boldsymbol{s} \in \mathrm{B}(\mathbf{0}, r) \cap \mathcal{L}$. The probability

that this test returns an incorrect value is the probability that $g(\boldsymbol{s}) \neq g_{\boldsymbol{q}_1}(\boldsymbol{s}) = g_{\boldsymbol{q}_j}(\boldsymbol{s})$ for every $\boldsymbol{q}_j$, which is at most $(4(d+2))^{-N_6'} = 4^{-N_6'}/(d+2)^{N_6'} \leq \varepsilon/2(d+1)$, where the final inequality follows by choosing $N_6' = O(\log(1/\varepsilon))$. As DISCRETEQUERY-$g(\boldsymbol{p})$ samples $d+1$ points, the probability that these points are all recovered successfully is at least $1 - \varepsilon/2$.

A similar argument holds for points $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r) \cap \frac{1}{B}\mathbb{Z}^n$, and bounds the probability by $1 - \varepsilon/2$ as well. □

References

[ABCG93] Sigal Ar, Manuel Blum, Bruno Codenotti, and Peter Gemmell. Checking approximate computations over the reals. In *Proceedings of the 25th Annual ACM Symposium on Theory of Computing (STOC)*, pages 786–795, 1993. 2

[AGHS13] Shweta Agrawal, Craig Gentry, Shai Halevi, and Amit Sahai. Discrete gaussian leftover hash lemma over infinite domains. In *International Conference on the Theory and Application of Cryptology and Information Security (Asiacrypt)*, pages 97–116, 2013. 37, 39

[ALM03] Jose María Almira and Antonio Jesús López-Moreno. Characterizing polynomials by forward differences. *Applied Mathematics E-Notes [electronic only]*, 3, 01 2003. 50

[ALM07] Jose María Almira and Antonio Jesús López-Moreno. On solutions of the fréchet functional equation. *Journal of mathematical analysis and applications*, 332(2):1119–1133, 2007. 50

[BBBY12] Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 21–30, 2012. 2

[BCS20] Hadley Black, Deeparnab Chakrabarty, and Comandur Seshadhri. Domain reduction for monotonicity testing: A $o(d)$ tester for boolean functions in d-dimensions. In *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1975–1994, 2020. 2

[Bel19] Aleksandrs Belovs. Quantum algorithm for distribution-free junta testing. In *Proceedings of the 14th International Computer Science Symposium in Russia (CSR)*, pages 50–59, 2019. 4

[BFH+13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every locally characterized affine-invariant property is testable. In *Proceedings of the 45th Annual ACM Symposium on Theory of computing (STOC)*, pages 429–436, 2013. 10

[BFPJH21] Eric Blais, Renato Ferreira Pinto Jr, and Nathaniel Harms. VC dimension and distribution-free sample-based testing. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 504–517, 2021. 2

[BHK20] Avrim Blum, John Hopcroft, and Ravindran Kannan. *Foundations of Data Science*. Cambridge University Press, 2020. 18

[BLP+13] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. Classical hardness of learning with errors. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC)*, pages 575–584, 2013. 36

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993. 1, 4, 5

[Bsh19] Nader H. Bshouty. Almost optimal distribution-free junta testing. In *Proceedings of the 34th Computational Complexity Conference (CCC)*, pages 2:1–2:13, 2019. 4

[BY22] Arnab Bhattacharyya and Yuichi Yoshida. *Property Testing: Problems and Techniques*. Springer, Singapore, 2022. 1

[Cau21] Augustin Louis Baron Cauchy. *Cours d'analyse de l'École Royale Polytechnique: Analyse algébrique. I. re partie*. Debure frères, 1821. 3

[CFSS17] Xi Chen, Adam Freilich, Rocco A. Servedio, and Timothy Sun. Sample-based high-dimensional convexity testing. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 37:1–37:20, 2017. 2

[Cie59] Zbigniew Ciesielski. Some properties of convex functions of higher orders. *Annales Polonici Mathematici*, 7:1–7, 1959. 3, 50

[CP22] Xi Chen and Shyamal Patel. Distribution-free testing for halfspaces (almost) requires pac learning. In *Proceedings of the 33rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1715–1743, 2022. 2

[CX16] Xi Chen and Jinyu Xie. Tight bounds for the distribution-free testing of monotone conjunctions. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 54–71, 2016. 2, 4

[Dar75] Gaston Darboux. Mémoire sur les fonctions discontinues. In *Annales scientifiques de l'École normale supérieure*, volume 4, pages 57–112, 1875. 3

[DMN19] Anindya De, Elchanan Mossel, and Joe Neeman. Is your function low dimensional? In *Proceedings of the 32nd Conference on Learning Theory (COLT)*, pages 979–993, 2019. 2

[DR11] Elya Dolev and Dana Ron. Distribution-free testing for monomials with a sublinear number of queries. *Theory of Computing*, 7(1):155–176, 2011. 2, 4

[EKR01] Funda Ergün, S Ravi Kumar, and Ronitt Rubinfeld. Checking approximate computations of polynomials and functional equations. *SIAM Journal on Computing*, 31(2):550–576, 2001. 2

[Fré09] Maurice Fréchet. Une définition fonctionnelle des polynômes. *Nouvelles annales de mathématiques: journal des candidats aux écoles polytechnique et normale*, 9:145–162, 1909. 3

[FY20] Noah Fleming and Yuichi Yoshida. Distribution-free testing of linear functions on $\mathbb{R}^n$. In *11th Innovations in Theoretical Computer Science Conference (ITCS)*, volume 151 of *LIPIcs*, pages 22:1–22:19, 2020. 2, 3, 4, 11, 55

[Gaj91] Zbigniew Gajda. Local stability of the functional equation characterizing polynomial functions. *Annales Polonici Mathematici*, 52(2):119–137, 1991. 7, 27, 51

[Ger71] Roman Ger. On some properties of polynomial functions. In *Annales Polonici Mathematici*, volume 25, pages 195–203. Institute of Mathematics Polish Academy of Sciences, 1971. 50

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998. 4

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 529–540, 2013. 7

[GLR+91] Peter Gemmell, Richard Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proceedings of the 23rd Annual ACM symposium on Theory of Computing (STOC)*, pages 32–42, 1991. 2

[Gol17] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. 1

[GS09] Dana Glasner and Rocco A. Servedio. Distribution-free testing lower bound for basic boolean functions. *Theory of Computing*, 5(10):191–216, 2009. 2, 4

[Ham05] Georg Hamel. Eine basis aller zahlen und die unstetigen lösungen der funktionalgleichung: $f(x+y) = f(x) + f(y)$. *Mathematische Annalen*, 60(3):459–462, 1905. 3

[Har19] Nathaniel Harms. Testing halfspaces over rotation-invariant distributions. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 694–713, 2019. 2, 4

[HIR12] Donald H Hyers, George Isac, and Themistocles Rassias. *Stability of functional equations in several variables*, volume 34. Springer Science & Business Media, 2012. 7

[HK07] Shirley Halevy and Eyal Kushilevitz. Distribution-free property-testing. *SIAM Journal on Computing*, 37(4):1107–1138, 2007. 2, 4

[HY20] Nathaniel Harms and Yuichi Yoshida. Downsampling for testing and learning in product distributions. *arXiv preprint arXiv:2007.07449*, 2020. 2

[KMS01] Marcos Kiwi, Frédéric Magniez, and Miklos Santha. Exact and approximate testing/correcting of algebraic functions: A survey. *Electron. Colloquium Comput. Complex.*, (14), 2001. 4, 7

[KNOW14] Pravesh Kothari, Amir Nayyeri, Ryan O'Donnell, and Chenggang Wu. Testing surface area. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1204–1214, 2014. 2

[Kom89] Zygíryd Kominek. On a local stability of the Jensen functional equation. *Demonstratio Mathematica*, 22(2):499–508, 1989. 56

[KS08] Tali Kaufman and Madhu Sudan. Algebraic property testing: the role of invariance. In *Proceedings of the 40th Annual ACM Symposium on Theory of computing (STOC)*, pages 403–412, 2008. 10

[LCS+19] Zhengyang Liu, Xi Chen, Rocco A. Servedio, Ying Sheng, and Jinyu Xie. Distribution-free junta testing. *ACM Transactions on Algorithms*, 15(1):1:1–1:23, 2019. 2, 4

[Lip89] Richard J. Lipton. New directions in testing. In *Distributed Computing And Cryptography, Proceedings of a DIMACS Workshop*, volume 2 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 191–202. DIMACS/AMS, 1989. 4

[McK67] MA McKiernan. On vanishing n-th ordered differences and hamel bases. In *Annales Polonici Mathematici*, volume 19, pages 331–336. Institute of Mathematics Polish Academy of Sciences, 1967. 50

[MH02] J.C. Mason and D.C. Handscomb. *Chebyshev Polynomials*. CRC Press, 2002. 23

[MORS09] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A Servedio. Testing ± 1-weight halfspaces. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 646–657. Springer Berlin Heidelberg, 2009. 2

[MORS10a] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A. Servedio. Testing halfspaces. *SIAM Journal on Computing*, 39(5):2004–2047, 2010. 2

[MORS10b] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A. Servedio. Testing (subclasses of) halfspaces. In *Property Testing - Current Research and Surveys*, pages 334–340. Springer Berlin Heidelberg, 2010. 2

[MP13] Daniele Micciancio and Chris Peikert. Hardness of SIS and LWE with small parameters. In *Annual Cryptology Conference*, pages 21–39. Springer, 2013. 39, 54, 55

[MR07] Daniele Micciancio and Oded Regev. Worst-case to average-case reductions based on gaussian measures. *SIAM Journal on Computing*, 37(1):267–302, 2007. 9, 38, 40, 54

[Nee14] Joe Neeman. Testing surface area with arbitrary accuracy. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*, pages 393–397, 2014. 2

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006. 9

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):1–40, 2009. 9

[RS92] Ronitt Rubinfeld and Madhu Sudan. Self-testing polynomial functions efficiently and over rational domains. In *Proceedings of the 3rd Annual ACM/SIGACT-SIAM Symposium on Discrete Algorithms (SODA)*, pages 23–32, 1992. 4

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996. 4, 6

[RS97] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 475–484, 1997. 4

Appendix A A Characterization of Degree-d Polynomials

In this appendix, we show how the Local Characterization Theorem of degree-d polynomials over $\mathbb{R}$ follows from known results. We begin with several definitions.

To connect finite forward differences of f with derivatives of f, we consider the *discrete differential operator*. Let $\boldsymbol{t} = (t_1, \ldots, t_{d+1}) \in \mathbb{R}^{d+1}$, where $t_i \neq 0$ for every $i \in [d+1]$, be a vector of length $d+1$, and for any $S \subseteq [d+1]$, denote $t_S \triangleq \sum_{j \in S} t_j$. The discrete differential operator is defined as

$$D_{\boldsymbol{t}}[f](x) = D_{(t_1,\ldots,t_{d+1})}[f](x) \triangleq \sum_{S \subseteq [d+1]} (-1)^{|S|} f\left(x + t_S\right).$$

Note that the derivative of f is the limit of the corresponding differential; formally,

$$\frac{d^{d+1}}{dx^{d+1}} f(x) = \lim_{\boldsymbol{t} \to \boldsymbol{0}} \frac{D_{\boldsymbol{t}}[f](x)}{\prod_{i \in [d+1]} t_i}. \tag{24}$$

Thus, we can obtain local information about the derivative by inspecting the discrete differential. We will indirectly evaluate the discrete differential by inspecting the forward finite difference $\Delta_h[f](x)$ (defined in (4)). Indeed, for $\mathbf{1} = (1, \ldots 1) \in \mathbb{R}^{d+1}$, observe that for any $h \in \mathbb{R}$,

$$D_{h \cdot \mathbf{1}}[f](x) = \sum_{S \subseteq [d+1]} (-1)^{|S|} f\left(x + h|S|\right) = \sum_{i=0}^{d+1} (-1)^i \binom{d+1}{i} f(x + hi) = (-1)^{d+1} \Delta_h^{(d+1)}[f](x). \tag{25}$$

We will first state a useful result from [Cie59], and sketch a variant of another result from [ALM03].

Theorem A.1. *(Theorem 1 of [Cie59]) Let $f : (a,b) \to \mathbb{R}$ be* J-convex of the d-th order over (a,b)*, i.e. $\Delta_h^{(d+1)}[f](x) \geq 0$, for every x, and $h > 0$ such that $a < x < x + (d+1)h < b$. If f is bounded on a set $E \subset (a,b)$ of positive measure, then f is continuous on the interval (a,b).*

Theorem A.2. *(A variant of Theorem 2 of [ALM03][10]) Let $f : (a,b) \to \mathbb{R}$ such that $\Delta_h^{(d+1)}[f](x) = 0$, for every $x \in (a,b)$ and $h > 0$, such that $a < x < x + (d+1)h < b$. If f is continuous on a set $S \subset (a,b)$ of $d+1$ points, then f is a degree-d polynomial over (a,b).*

Proof Sketch: The proof goes through by showing that for an arbitrary $\alpha \in (a,b)$, and a large enough $M \in \mathbb{N}$, f agrees with a unique degree-d polynomial p, on a sequence of sets $\{\{\frac{i\alpha}{2^n M}\}_{i \in \mathbb{Z}} \cap (a,b)\}_{n \in \mathbb{N}}$, and then by the continuity of f on S, f and p can be proven to be arbitrarily close on S, i.e. $f(x) = p(x)$, for every $x \in S$. With $|S| = d+1$, this proves f is a degree-d polynomial. □

We are now ready to prove the Local Characterization Theorem, restated next for convenience.

Local Characterization Theorem. *Let $a, b \in \mathbb{R}$ such that $a < b$, and let $g\colon (a,b) \to \mathbb{R}$ be a univariate, bounded function. If for every $x \in (a,b)$ and sufficiently small $h > 0$, such that $a < x < x + (d+1)h < b$, $\Delta_h^{(d+1)}[g](x) = 0$, then g is a degree-d polynomial.*

Proof. Since $\Delta_h^{d+1}[g](x) = 0$ for every $x \in (a,b)$ and sufficiently small $h > 0$, such that $a < x < x + (d+1)h < b$, it follows that g is J-convex of the d-th order over (a,b). Since g is bounded as well over (a,b), by Theorem A.1 g must be continuous over (a,b). Now, invoking Theorem A.2, we thus claim g is a degree-d polynomial over (a,b). □

[10] See also [McK67, Ger71, ALM07].

Next, we prove the Discrete Local Characterization Theorem which was used in Section 5.

Discrete Local Characterization Theorem. *Fix $a > 0$, $M \geq d+1$, and let $S \triangleq \{\frac{ia}{M} : i \in \mathbb{N}\}$. If $f : [0,a] \to \mathbb{R}$ is a univariate function such that $\Delta^{(d+1)}_{\frac{a}{M}}[f](x) = 0$, for every $x \in S \cap [0,a]$ satisfying $x + (d+1)\frac{a}{M} \in [0,a]$, then f agrees with a degree-d polynomial over the points in $S \cap [0,a]$.*

Proof. For each $i \in \mathbb{N}$, let $S_i \triangleq \{\frac{ia}{M}, \ldots, \frac{(i+d)a}{M}\} \subset [0,a]$ and let $p_i(t) : [0,a] \to \mathbb{R}$ be the unique degree-d polynomial satisfying $p_i(s) = f(s)$ for all $s \in S_i$. We will argue that the polynomials p_i are identical, and thus equal to f on $S \cap [0,a]$. First, we will argue that $p_0 = p_1$. Observe that

$$0 = \Delta^{(d+1)}_{\frac{a}{M}}[f](0) = \sum_{j=0}^{d+1} \alpha_j \cdot f\left(\frac{ja}{M}\right) = \sum_{j=0}^{d} \alpha_j \cdot f\left(\frac{ia}{M}\right) + \alpha_{d+1} \cdot f\left(\frac{(d+1)a}{M}\right).$$

As p_0 was defined by interpolating the values of f on S_0, we have

$$0 = \sum_{j=0}^{d} \alpha_j \cdot p_0\left(\frac{ja}{M}\right) + \alpha_{d+1} \cdot f\left(\frac{(d+1)a}{M}\right) = \underbrace{\sum_{j=0}^{d+1} \alpha_j \cdot p_0\left(\frac{ja}{M}\right)}_{=0} + \alpha_{d+1} \cdot (f - p_0)\left(\frac{(d+1)a}{M}\right),$$

which implies that $f((d+1)a/M) = p_0((d+1)a/M)$, and hence $p_0(t) = p_1(t)$ for every $t \in [0,a]$. Repeating this argument for every $i \in \mathbb{N}$, we can conclude that $p_0(t) = p_1(t) \ldots = f(t)$ for every $t \in S \cap [0,a]$, where the equality with f follows because the p_i's are defined by interpolating the values of f on the S_i's. Thus, f agrees with a degree-d polynomial on $S \cap [0,a]$.

□

Appendix B Proofs from Section 4.3

In this appendix, we provide proofs of Lemma 4.8, and Lemma 4.10. But first we state the result from [Gaj91], that forms the basis of our Theorem 4.6:

Theorem B.1. *([Gaj91, Theorem 8]) Let X be a linear space over the rationals, $x_0 \in \mathbb{R}, d \in \mathbb{N}, \phi, a \in (0,\infty)$, and suppose that for all $x \in (x_0 - a, x_0 + a)$ and $h \in (-a, a)$, with $x + (d+1)h \in (x_0 - a, x_0 + a)$, $f\colon (x_0 - a, x_0 + a) \to X$ satisfies*

$$|\Delta^{(d+1)}_h[f](x)| \leq \phi.$$

Then, there exists a degree-d polynomial $g\colon \mathbb{R} \to X$, such that for every $x \in (x_0 - a, x_0 + a)$, $|f(x) - g(x)| \leq l''_d \phi$, where $l''_d \triangleq n_d + 2^{d+1} l'_d (2^{d+1} - 1)^{n_d}$, $n_d \triangleq \min\{k \in \mathbb{N} : (1 + k/d)^k \geq d\}$, $l'_d \triangleq \prod_{i=1}^{d} (k'_i + 1)$, $l'_0 \triangleq 1$, and $k'_i \triangleq 3 \cdot 2^{2i} + (i-1)2^{i+1} - 1$. In particular, $l'_d = \Theta(3^d \cdot 2^{d^2+d})$, $n_d \in \Omega(\log d) \cap o(d)$, and $l''_d = o(2^{8d^2})$.

Interestingly, [Gaj91] defines a function $g : (a,b) \to X$ to be a degree-d polynomial, if $\Delta^{(d+1)}_h[g](x) = 0$, for all $x \in (a,b)$ and $h > 0$, such that $a < x < x + (d+1)h < b$. Note that if $f : (a,b) \to \mathbb{R}$ is bounded on (a,b), then by Theorem B.1 $g : \mathbb{R} \to \mathbb{R}$ is also bounded, and hence a degree-d polynomial on the same interval, by the Local Characterization Theorem, thus proving Theorem 4.6. We now resume the proofs:

Lemma 4.8. *For every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, and every $t \in \{0, \ldots, d+1\}$,*

$$\Pr_{\substack{\boldsymbol{q}_1 \sim \mathcal{N}(\mathbf{0}, (t^2+1)I) \\ \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)}} \left[|g_{\boldsymbol{q}_1}(\boldsymbol{p}) - g_{\boldsymbol{q}_2}(\boldsymbol{p})| > 2^{d+2}\delta\right] \leq 4d\rho + 48d^5 r.$$

Proof. Fix some $t \in \{0, \dots, d+1\}$ and $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$. We will bound the probability that $g_{\boldsymbol{q}_1}(\boldsymbol{p})$ and $q_{\boldsymbol{q}_2}(\boldsymbol{p})$ are far from $\sum_{i=1}^{d+1}\sum_{j=1}^{d+1}\alpha_i\alpha_j \cdot f(\boldsymbol{p} + i\boldsymbol{q}_1 + j\boldsymbol{q}_2)$; the lemma will then follow by a union bound.

By definition, we have $g_{\boldsymbol{q}_2}(\boldsymbol{p}) = \sum_{i=1}^{d+1}\alpha_i \cdot f(\boldsymbol{p} + i\boldsymbol{q}_2)$. Fixing an $i \in [d+1]$, we get

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[|f(\underbrace{\boldsymbol{p}+i\boldsymbol{q}_1}_{\triangleq \boldsymbol{z}}) - g_{\boldsymbol{q}_2}(\boldsymbol{p}+i\boldsymbol{q}_1)| > \delta\right] = \Pr_{\substack{\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{p},i^2(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[\left|f(\boldsymbol{z}) - \sum_{j=1}^{d+1}\alpha_j f(\boldsymbol{z}+j\boldsymbol{q}_2)\right| > \delta\right]$$

$$\le \Pr_{\substack{\boldsymbol{z}\sim\mathcal{N}(\mathbf{0},i^2(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[\left|\sum_{j=0}^{d+1}\alpha_j \cdot f(\boldsymbol{z}+j\boldsymbol{q}_2)\right| > \delta\right] + 2\,\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0}, i^2(t^2+1)I), \mathcal{N}(\boldsymbol{p}, i^2(t^2+1)I))$$

$$\le \rho + i^2(t^2+1)r \le \rho + 20d^4 r. \qquad \text{(By (15) and Lemma 2.2)}$$

By a similar calculation, we have for every $j \in [d+1]$ that

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[|f(\underbrace{\boldsymbol{p}+j\boldsymbol{q}_2}_{\triangleq \boldsymbol{z}}) - g_{\boldsymbol{q}_1}(\boldsymbol{p}+j\boldsymbol{q}_2)| > \delta\right] = \Pr_{\substack{\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{p},j^2 I)\\ \boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)}}\left[\left|f(\boldsymbol{z}) - \sum_{i=1}^{d+1}\alpha_i f(\boldsymbol{z}+i\boldsymbol{q}_1)\right| > \delta\right]$$

$$\le \Pr_{\substack{\boldsymbol{z}\sim\mathcal{N}(\mathbf{0},j^2 I)\\ \boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)}}\left[\left|\sum_{i=0}^{d+1}\alpha_1 \cdot f(\boldsymbol{z}+i\boldsymbol{q}_1)\right| > \delta\right] + 2\,\mathrm{d_{TV}}(\mathcal{N}(\mathbf{0}, j^2 I), \mathcal{N}(\boldsymbol{p}, j^2 I))$$

$$\le \rho + j^2 r \le \rho + 4d^2 r. \qquad \text{(By (16) and Lemma 2.2)}$$

Taking a union bound over $i \in [d+1]$ and $j \in [d+1]$ respectively, it follows that

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[\left|\underbrace{\sum_{i=1}^{d+1}\alpha_i \cdot f(\boldsymbol{p}+i\boldsymbol{q}_1)}_{=g_{\boldsymbol{q}_1}(\boldsymbol{p})} - \sum_{i=1}^{d+1}\sum_{j=1}^{d+1}\alpha_i\alpha_j \cdot f((\boldsymbol{p}+i\boldsymbol{q}_1)+j\boldsymbol{q}_2)\right| > 2^{d+1}\delta\right] \le 2d\rho + 40d^5 r,$$

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}\left[\left|\underbrace{\sum_{j=1}^{d+1}\alpha_j \cdot f(\boldsymbol{p}+j\boldsymbol{q}_2)}_{=g_{\boldsymbol{q}_2}(\boldsymbol{p})} - \sum_{j=1}^{d+1}\sum_{i=1}^{d+1}\alpha_i\alpha_j \cdot f((\boldsymbol{p}+j\boldsymbol{q}_2)+i\boldsymbol{q}_1)\right| > 2^{d+1}\delta\right] \le 2d\rho + 8d^3 r.$$

Thus, by a union bound over the two previous inequalities we can conclude that

$$\Pr_{\substack{\boldsymbol{q}_1\sim\mathcal{N}(\mathbf{0},(t^2+1)I)\\ \boldsymbol{q}_2\sim\mathcal{N}(\mathbf{0},I)}}[|g_{\boldsymbol{q}_1}(\boldsymbol{p}) - g_{\boldsymbol{q}_2}(\boldsymbol{p})| > 2^{d+2}\delta] \le 4d\rho + 48d^5 r. \qquad \square$$

Lemma 4.10. *If* ApproxCharacterizationTest *fails with probability at most* $2/3$*, then for every* $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ *and sufficiently small* $h \in \mathbb{R}$*, such that* $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ *for every* $i \in [d+1]$*, we have* $|\sum_{i=0}^{d+1}\alpha_i \cdot g(\boldsymbol{p}+ih\boldsymbol{q})| \le 2^{2d+5}\delta$.

Proof. Fix $\boldsymbol{p}, \boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ and let $h \in \mathbb{R}$ be sufficiently small so that $\boldsymbol{p} + ih\boldsymbol{q} \in \mathrm{B}(\mathbf{0}, r)$ for every $i \in [d+1]$; such h's exist, as $\mathrm{B}(\mathbf{0}, r)$ is an open ball containing $\boldsymbol{p}$. We will argue that the following hold simultaneously with non-zero probability over $\boldsymbol{q}_1, \boldsymbol{q}_2 \sim \mathcal{N}(\mathbf{0}, I)$:

$$\left|\sum_{i=0}^{d+1}\alpha_i \cdot g(\boldsymbol{p}+ih\boldsymbol{q}) - \sum_{i=0}^{d+1}\alpha_i \cdot g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+ih\boldsymbol{q})\right| \le 2^{2d+4}\delta, \tag{26}$$

$$\Big|\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{p}+j\boldsymbol{q}_1+i(h\boldsymbol{q}+j\boldsymbol{q}_2))\Big|\le\delta,\text{ for every } j\in[d+1]. \tag{27}$$

Assuming that these hold, we complete the proof. Fix any $\boldsymbol{q}_1,\boldsymbol{q}_2$ satisfying both (26), and (27). Then,

$$\begin{aligned}
\Big|\sum_{i=0}^{d+1}\alpha_i\cdot g(\boldsymbol{p}+ih\boldsymbol{q})\Big| &\le \Big|\sum_{i=0}^{d+1}\alpha_i\cdot g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+ih\boldsymbol{q})\Big| + 2^{2d+4}\delta && \text{(By (26))}\\
&= \Big|\sum_{i=0}^{d+1}\alpha_i\Big(\sum_{j=1}^{d+1}\alpha_j\cdot f(\boldsymbol{p}+ih\boldsymbol{q}+j(\boldsymbol{q}_1+i\boldsymbol{q}_2))\Big)\Big| + 2^{2d+4}\delta\\
& && \text{(By definition of } g_{\boldsymbol{q}}(\boldsymbol{p}))\\
&= \Big|\sum_{j=1}^{d+1}\alpha_j\Big(\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{p}+j\boldsymbol{q}_1+i(h\boldsymbol{q}+j\boldsymbol{q}_2))\Big)\Big| + 2^{2d+4}\delta\\
&\le \Big|\sum_{j=1}^{d+1}\alpha_j\cdot\delta\Big| + 2^{2d+4}\delta && \text{(By (27))}\\
&\le 2^{d+1}\delta + 2^{2d+4}\delta = 2^{d+1}(2^{d+3}+1)\delta \le 2^{2d+5}\delta.
\end{aligned}$$

Next, we prove (26) and (27), by arguing that each holds with positive probability and then taking a union bound. For (26), we observe

$$\begin{aligned}
&\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\boldsymbol{0},I)}\Big[\Big|\sum_{i=0}^{d+1}\alpha_i\cdot g(\boldsymbol{p}+ih\boldsymbol{q})-\sum_{i=0}^{d+1}\alpha_i\cdot g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+ih\boldsymbol{q})\Big|\le 2^{2d+4}\delta\Big]\\
&\ge \Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\boldsymbol{0},I)}\Big[|g(\boldsymbol{p}+ih\boldsymbol{q})-g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+ih\boldsymbol{q})|\le 2^{d+3}\delta,\ \ \forall i\in\{0,\dots,d+1\}\Big]\\
&= 1-\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\boldsymbol{0},I)}\Big[\exists i\in\{0,\dots,d+1\}:|g(\boldsymbol{p}+ih\boldsymbol{q})-g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+ih\boldsymbol{q})|>2^{d+3}\delta\Big]\\
&\ge 1-\sum_{i=0}^{d+1}\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\boldsymbol{0},I)}\Big[|g(\boldsymbol{p}+ih\boldsymbol{q})-g_{\boldsymbol{q}_1+i\boldsymbol{q}_2}(\boldsymbol{p}+ih\boldsymbol{q})|>2^{d+3}\delta\Big] && \text{(By union bound)}\\
&= 1-\sum_{i=0}^{d+1}\Pr_{\boldsymbol{m}\sim\mathcal{N}(\boldsymbol{0},(i^2+1)I)}\Big[|g(\boldsymbol{p}+ih\boldsymbol{q})-g_{\boldsymbol{m}}(\boldsymbol{p}+ih\boldsymbol{q})|>2^{d+3}\delta\Big] && \text{(Letting } \boldsymbol{m}\triangleq\boldsymbol{q}_1+i\boldsymbol{q}_2)\\
&> 1-\sum_{i=0}^{d+1}\frac{1}{7d}=1-\frac{d+2}{7d}>\frac{1}{2}. && \text{(Applying Corollary 4.9, as } \boldsymbol{p}+ih\boldsymbol{q}\in\mathrm{B}(\boldsymbol{0},r))
\end{aligned}$$

For (9), consider some $j\in[d+1]$, then

$$\begin{aligned}
&\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\boldsymbol{0},I)}\Big[\Big|\sum_{i=0}^{d+1}\alpha_i\cdot f(\underbrace{\boldsymbol{p}+j\boldsymbol{q}_1}_{\triangleq\boldsymbol{z}_1}+i(\underbrace{h\boldsymbol{q}+j\boldsymbol{q}_2}_{\triangleq\boldsymbol{z}_2}))\Big|>\delta\Big]=\Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{N}(\boldsymbol{p},j^2I)\\ \boldsymbol{z}_2\sim\mathcal{N}(h\boldsymbol{q},j^2I)}}\Big[\Big|\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{z}_1+i\boldsymbol{z}_2)\Big|>\delta\Big]\\
&\le \Pr_{\substack{\boldsymbol{z}_1\sim\mathcal{N}(\boldsymbol{0},j^2I)\\ \boldsymbol{z}_2\sim\mathcal{N}(\boldsymbol{0},j^2I)}}\Big[\Big|\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{z}_1+i\boldsymbol{z}_2)\Big|>\delta\Big]\\
&+2\left(\mathrm{d_{TV}}\left(\mathcal{N}(\boldsymbol{0},j^2I),\mathcal{N}(\boldsymbol{p},j^2I)\right)+\mathrm{d_{TV}}\left(\mathcal{N}(\boldsymbol{0},j^2I),\mathcal{N}(h\boldsymbol{q},j^2I)\right)\right)
\end{aligned}$$

$$\leq \Pr_{\boldsymbol{z}_1,\boldsymbol{z}_2\sim\mathcal{N}(\boldsymbol{0},j^2I)}\left[\left|\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{z}_1+i\boldsymbol{z}_2)\right|>\delta\right]+j^2r+j^2hr \qquad \text{(By Lemma 2.2)}$$

$$\leq \rho+j^2r+j^2hr<\rho+8d^2r. \qquad \text{(By (17))}$$

By a union bound over all $j\in[d+1]$,

$$\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{N}(\boldsymbol{0},I)}\left[\forall j\in[d+1],\left|\sum_{i=0}^{d+1}\alpha_i\cdot f(\boldsymbol{p}+j\boldsymbol{q}_1+i(h\boldsymbol{q}+j\boldsymbol{q}_2))\right|\leq\delta\right]\geq 1-(2d\rho+16d^3r),$$

which is at least $2/3$, as $\rho\leq(30d)^{-2}$ by Claim 4.7, and we set $r=(4d)^{-6}$ in Corollary 4.9. A final union bound concludes that both (26) and (27) hold simultaneously with non-zero probability. □

Appendix C Proof of Theorem 5.6

In this appendix we prove Theorem 5.6, which is an immediate consequence of the following lemma. This lemma follows in a straightforward manner from [MP13, Theorem 3.3] . Denote by $\oplus$, the standard *direct sum* of lattices and by $\otimes$, the standard *tensor product*, and let I_n denote the $n\times n$ identity matrix.

Lemma C.1. *Let $\mathcal{L}\subseteq\mathbb{R}^n$ be a full rank lattice, and fix $k\in\mathbb{Z}_{>0}$, ϑ, $s_1,\ldots,s_k\in\mathbb{R}_{>0}$, and $\boldsymbol{z}\triangleq(z_1,\ldots,z_k)\in\mathbb{Z}^k$ such that $s_i\geq\|\boldsymbol{z}\|_\infty\sqrt{2}\eta_\vartheta(\mathcal{L})$ for every $i\in[k]$. Let $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k$ be sampled independently from $\mathcal{G}(\mathcal{L},s_i)$. Then, for $s\triangleq\sqrt{\sum_{i=1}^k(z_is_i)^2}$, the total variation distance between $\boldsymbol{y}\triangleq\sum_{i=1}^k z_i\boldsymbol{y}_i$ and $\mathcal{G}\left(\gcd\left(\boldsymbol{z}\right)\mathcal{L},s\right)$ is given by*

$$\mathrm{d_{TV}}\left(\boldsymbol{y},\mathcal{G}\left(\gcd\left(\boldsymbol{z}\right)\mathcal{L},s\right)\right)=\frac{1}{2}\sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}}\left|\frac{\tau_s(\boldsymbol{y})}{\tau_s(\gcd(\boldsymbol{z})\mathcal{L})}-\frac{\tau_s(\boldsymbol{y})\tau(L+\boldsymbol{x})}{\tau(\mathcal{L}')}\right|,$$

where the nk-dimensional lattice $\mathcal{L}'\triangleq\bigoplus_{i=1}^k s_i^{-1}\mathcal{L}=(\boldsymbol{S}\otimes I_n)^{-1}\mathcal{L}^{\oplus k}$, for $\boldsymbol{S}\triangleq\mathrm{diag}(s_1,\ldots,s_k)$, and L is the sublattice of $\mathcal{L}'$ containing the elements which fall in the kernel of $\boldsymbol{Z}\triangleq(\boldsymbol{z}^\top\boldsymbol{S})\otimes I_n$; that is, $L\triangleq\mathcal{L}'\cap\ker(\boldsymbol{Z})$. As well, $\boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}'$ onto $\ker(\boldsymbol{Z})$, where $\boldsymbol{x}'\in\mathcal{L}':\boldsymbol{Z}\boldsymbol{x}'=\boldsymbol{y}$. Furthermore,

$$\mathrm{d_{TV}}\left(\boldsymbol{y},\mathcal{G}\left(\gcd\left(\boldsymbol{z}\right)\mathcal{L},s\right)\right)\leq\frac{2k\vartheta}{1-2k\vartheta}.$$

The proof of the first part follows by recording the parameters obtained in [MP13], while the proof for the second part is provided below:

Proof. Let $\Re\colon\mathbb{C}\to\mathbb{R}$ denote the real part of a complex number. From the proof of Lemma 4.1 in [MR07], we have that for every $\boldsymbol{x}\in\mathbb{R}^n$, lattice Λ, and $\vartheta>0$, such that $\eta_\vartheta(\Lambda)\leq 1$,

$$\begin{aligned}\underbrace{\tau(\Lambda+\boldsymbol{x})}_{\geq 0}&=\underbrace{\det(\Lambda^*)}_{\geq 0}\Big(1+\sum_{\boldsymbol{w}\in\Lambda^*\setminus\{\boldsymbol{0}\}}e^{2\pi i\langle\boldsymbol{x},\boldsymbol{w}\rangle}\underbrace{\tau(\boldsymbol{w})}_{\geq 0}\Big)\\&=\det(\Lambda^*)\Big(1+\sum_{\boldsymbol{w}\in\Lambda^*\setminus\{\boldsymbol{0}\}}\underbrace{\Re(e^{2\pi i\langle\boldsymbol{x},\boldsymbol{w}\rangle})}_{\in[-1,1]}\tau(\boldsymbol{w})\Big)\in\det(\Lambda^*)\Big(1\pm\sum_{\boldsymbol{w}\in\mathcal{L}^*\setminus\{\boldsymbol{0}\}}\tau(\boldsymbol{w})\Big)\\&=\det(\Lambda^*)\Big(1\pm\underbrace{\tau(\Lambda^*\setminus\{\boldsymbol{0}\})}_{\leq\vartheta,\because\eta_\vartheta(\mathcal{L})\leq 1}\Big)=\det(\Lambda^*)(1\pm\vartheta).\end{aligned}$$

From the proof of Theorem 3.3 in [MP13], we have $\eta_{\vartheta'}(L) \le \sqrt{2}\eta_\vartheta(\mathcal{L})/\min s_i \le 1$, where $\vartheta' \triangleq (1+\vartheta)^{k-1}-1$. And for all ϑ, k such that $k\vartheta \in [0,1]$ we have $\vartheta' = (1+\vartheta)^{k-1}-1 \le 2k\vartheta$. Also, note that for every $0 \le \vartheta_1 \le \vartheta_2$, $\eta_{\vartheta_1}(\Lambda) \ge \eta_{\vartheta_2}(\Lambda)$ holds, for every lattice Λ. So, we can claim $\eta_{2k\vartheta}(L) \le \eta_{\vartheta'}(L) \le 1$, and hence we have that for every $\boldsymbol{x} \in \mathbb{R}^n, \tau(L+\boldsymbol{x}) \in \det(L^*)(1 \pm 2k\vartheta) = \det(L^*)[1-2k\vartheta, 1+2k\vartheta]$. Now observe that

$$\sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} \underbrace{\frac{\tau_s(\boldsymbol{y})}{\tau_s(\gcd(\boldsymbol{z})\mathcal{L})}}_{\triangleq p(\boldsymbol{y})} = 1 = \sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} \underbrace{\frac{\tau_s(\boldsymbol{y})\tau(L+\boldsymbol{x})}{\tau(\mathcal{L}')}}_{\triangleq q(\boldsymbol{y})} = \sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} \underbrace{p(\boldsymbol{y})c\cdot(1+\vartheta(\boldsymbol{y}))}_{=q(\boldsymbol{y})},$$

where $c \triangleq \tau_s(\gcd(\boldsymbol{z})\mathcal{L})\det(L^*)/\tau(\mathcal{L}') = \tau_s(\gcd(\boldsymbol{z})\mathcal{L})/\tau(\mathcal{L}')\det(L)$, and $\vartheta(\boldsymbol{y}) \in [-2k\vartheta, 2k\vartheta]$ for every $\boldsymbol{y} \in \gcd(\boldsymbol{z})\mathcal{L}$. Thus, we have,

$$\begin{aligned}\frac{1}{c} &= \sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} p(\boldsymbol{y})(1+\vartheta(\boldsymbol{y})) = \underbrace{\sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} p(\boldsymbol{y})}_{=1} + \underbrace{\sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} p(\boldsymbol{y})\vartheta(\boldsymbol{y})}_{p(\boldsymbol{y}) \text{ is a PMF}} \\ &= 1 + \underbrace{\mathbb{E}_{\boldsymbol{y}\sim_{p(\boldsymbol{y})}\gcd(\boldsymbol{z})\mathcal{L}}[\vartheta(\boldsymbol{y})]}_{\in[\min_{\boldsymbol{y}}\vartheta(\boldsymbol{y}),\max_{\boldsymbol{y}}\vartheta(\boldsymbol{y})]} \in [1-2k\vartheta, 1+2k\vartheta].\end{aligned}$$

Thus, $c \in [1/(1+2k\vartheta), 1/(1-2k\vartheta)]$ and $1-c \in [-2k\vartheta/(1-2k\vartheta), 2k\vartheta/(1+2k\vartheta)]$. It follows that

$$\begin{aligned}\mathrm{d_{TV}}\left(\boldsymbol{y}, \mathcal{G}\left(\gcd(\boldsymbol{z})\mathcal{L}, s\right)\right) &= \frac{1}{2}\sum_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}} \left|\frac{\tau_s(\boldsymbol{y})}{\tau_s(\gcd(\boldsymbol{z})\mathcal{L})} - \frac{\tau_s(\boldsymbol{y})\tau(L+\boldsymbol{x})}{\tau(\mathcal{L}')}\right| \le \frac{1}{2}\max_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}}\left|1-\frac{q(\boldsymbol{y})}{p(\boldsymbol{y})}\right| \\ &\le \frac{1}{2}\max_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}}\max_c |1-c(1+\vartheta(\boldsymbol{y}))| \\ &\le \frac{1}{2}\left(\max_c|1-c| + \max_c c\cdot\max_{\boldsymbol{y}\in\gcd(\boldsymbol{z})\mathcal{L}}|\vartheta(\boldsymbol{y})|\right) \\ &= \frac{1}{2}\left(\frac{2k\vartheta}{1-2k\vartheta} + \frac{1}{1-2k\vartheta}2k\vartheta\right) = \frac{2k\vartheta}{1-2k\vartheta}.\end{aligned}$$

□

Finally, setting k and ϑ such that $k\vartheta < 1/4$ proves Theorem 5.6.

Appendix D Distribution-Free Approximate Tester for Additivity

Recall that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *additive* if for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $f(\boldsymbol{x}+\boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$; a function is *linear* if it is both additive and for every $\alpha \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^n$, $f(\alpha\boldsymbol{x}) = \alpha f(\boldsymbol{x})$. In this appendix we modify the additivity tester of [FY20] to be robust against noise. This gives us a tester for additivity with better error parameters than the approximate degree-1 tester obtained from Theorem 1.2. Formally, given query access to the input function $f : \mathbb{R}^n \to \mathbb{R}$, sampling access to unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$, and constants $0 < \alpha \in \mathbb{R}$ and $0 < \varepsilon \in \mathbb{R}$, a *distribution-free approximate tester* for additivity distinguishes between the following two cases with probability at least $2/3$:

- **YES CASE:** There exists an additive function $h : \mathbb{R}^n \to \mathbb{R}$ such that for all $\boldsymbol{p} \in \mathbb{R}^n$:

$$|f(\boldsymbol{p}) - h(\boldsymbol{p})| \le \alpha;$$

- **NO CASE:** For any additive function $h : \mathbb{R}^n \to \mathbb{R}$:

$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > 21015 \cdot Rn^{1.5}\alpha] > \varepsilon.$$

We say that the tester has one-sided error if, for every f satisfying the YES CASE , the tester always accepts, with probability 1.

The main theorem of this section is the following.

Theorem D.1. *Let $\alpha, \varepsilon > 0$ and $\mathcal{D}$ be an unknown $(R, \varepsilon/4)$-concentrated distribution. There exists a one-sided error, $O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})$-query for distinguishing between the case when f is pointwise α-close to some additive function and the case when, for every additive function h, $\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > O(Rn^{1.5}\alpha)] > \varepsilon$.*

The remainder of this section is organized as follows: In Section D.1, first we describe several properties of additive functions which we will require for our tester, and give an overview of the proof of Theorem D.1. Then, we present our tester under some constraints on the unknown $\mathcal{D}$, and give informal description of the proof technique. In Section D.2 we prove the main Theorem D.1, relying on our main Lemma D.6. Section D.3 is devoted to prove the main Lemma D.6. Finally, in Section D.4, we show that our tester is actually a multiplicative error distribution-free tester, without any assumption on the unknown distribution $\mathcal{D}$.

D.1 Proof Overview and δ-Additive Functions

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is *δ-additive*, if for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ it holds that

$$|f(\boldsymbol{x} + \boldsymbol{y}) - f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq \delta.$$

Satisfying δ-additivity implies that the following inequalities hold, which will be the basis for our tester. For every $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n$, assuming f is α-close to some additive function h, we have:

$$|f(\boldsymbol{x} - \boldsymbol{y}) - f(\boldsymbol{x}) + f(\boldsymbol{y})| \leq |\underbrace{h(\boldsymbol{x} - \boldsymbol{y}) - h(\boldsymbol{x}) - h(-\boldsymbol{y})}_{=0}| + 3\alpha \leq \delta, \tag{28}$$

$$|f(\boldsymbol{x}) + f(-\boldsymbol{x})| \leq |h(\boldsymbol{x}) + h(-\boldsymbol{x})| + 2\alpha = |h(\boldsymbol{x}) - h(\boldsymbol{x})| + 2\alpha \leq \delta \tag{29}$$

$$|f(\boldsymbol{x} - \boldsymbol{y}) - f(\boldsymbol{x} - \boldsymbol{z}) - f(\boldsymbol{z} - \boldsymbol{y})| \leq |h(\boldsymbol{x} - \boldsymbol{y}) - \underbrace{(h(\boldsymbol{x} - \boldsymbol{z}) + h(\boldsymbol{z} - \boldsymbol{y}))}_{=h(\boldsymbol{x}-\boldsymbol{y})}| + 3\alpha \leq \delta \tag{30}$$

Our tester (given in Algorithm 7 and Algorithm 8) follows the general outline given in the introduction for testing linearity. First, it tests whether f satisfies δ-additivity over a set of samples drawn from the distribution $\mathcal{N}(\mathbf{0}, I)$. If this test passes with sufficiently high probability then we able to show that g — a self-corrected function of f on $\mathrm{B}(\mathbf{0}, r)$ — is $O(n^{1.5}\delta)$-close to an additive function $h : \mathbb{R}^n \to \mathbb{R}$, and furthermore, if f is δ-additive, then f and g (and therefore f and h) are close. To do so, we crucially rely on the following stability theorem for additive functions which follows from [Kom89, Theorem 2].

Theorem D.2. *Let $r > 0$ and $g : \mathrm{B}(\mathbf{0}, r) \to \mathbb{R}$. If g is δ-additive, then there exists an additive function $h : \mathbb{R}^n \to \mathbb{R}$, such that for every $\boldsymbol{x} \in \mathrm{B}(\mathbf{0}, r)$*

$$|g(\boldsymbol{x}) - h(\boldsymbol{x})| \leq 5n^{1.5}\delta.$$

Second, we show that by the way we have constructed g, we are able to approximate its value on points within $\mathrm{B}(\mathbf{0}, r)$ with high probability. Thus, for any point $\mathrm{B}(\mathbf{0}, r)$, we can estimate the distance between f and g, and therefore between f and h, the additive function which is close to g, given by Theorem D.2.

For points $\boldsymbol{p} \notin \mathrm{B}(\mathbf{0}, r)$, we map them to a point within $\mathrm{B}(\mathbf{0}, r)$ by dividing by a contraction factor $\kappa_{\boldsymbol{p}}$, defined as

$$\kappa_{\boldsymbol{p}} \triangleq \begin{cases} 1 & \text{if } \|\boldsymbol{p}\|_2 \leq r, \\ \left\lceil \frac{\|\boldsymbol{p}\|_2}{r} \right\rceil & \text{if } \|\boldsymbol{p}\|_2 > r. \end{cases}$$

Then, we approximate h on the corresponding point $\boldsymbol{p}/\kappa_{\boldsymbol{p}}$ inside the ball and map $h(\boldsymbol{p}/\kappa_{\boldsymbol{p}})$ back to $h(\boldsymbol{p})$.

We are now ready to formally define g.

The Self-Corrected Function. Let r be a sufficiently small rational; $r \triangleq 1/50$ suffices. Define the value of the self-corrected function g at a point $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ as the (weighted) median value of $g_{\boldsymbol{x}}(\boldsymbol{p}) \triangleq f(\boldsymbol{p} - \boldsymbol{x}) + f(\boldsymbol{x})$, each weighted according to its probability mass under $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, I)$. For points $\boldsymbol{p}$ outside of the ball, we project them into the ball by diving by a sufficiently large contraction factor that depends on the magnitude of $\boldsymbol{p}$.

Concretely, $g : \mathbb{R}^n b \to \mathbb{R}$ is defined as follows

$$g(\boldsymbol{p}) \triangleq \kappa_{\boldsymbol{p}} \cdot \operatorname*{med}_{\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, I)} \left[g_{\boldsymbol{x}} \left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}} \right) \right] = \kappa_{\boldsymbol{p}} \cdot \operatorname*{med}_{\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, I)} \left[f \left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}} - \boldsymbol{x} \right) + f(\boldsymbol{x}) \right].$$

The intuition for using median is that it, in the case when f is approximately additive, the median value should allow us to approximately correct the errors in f, and thus g should be close to additive. We use the median here, rather than the majority, because the majority is more affected by outliers.

D.2 Approximate Additivity Tester

Our tester is given in Algorithm 7, which uses subroutines given in Algorithm 8.

Algorithm 7: Approximate Additivity Tester

Procedure ApproxAdditivityTester($f, \mathcal{D}, \alpha, \varepsilon, R$)
- **Given :** Query access to $f \colon \mathbb{R}^n \to \mathbb{R}$, sampling access to an unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$, a noise parameter $\alpha > 0$, and a farness parameter $\varepsilon > 0$;
- $\delta \leftarrow 3\alpha$, $r \leftarrow 1/50$;
- **Reject** if TestAdditivity(f, δ) returns **Reject**;
- **for** $N_7 \leftarrow O(1/\varepsilon)$ *times* **do**
  - Sample $\boldsymbol{p} \sim \mathcal{D}$;
  - **if** $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$ **then**
    - **Reject** if $|f(\boldsymbol{p}) -$ Approximate-$g(\boldsymbol{p}, f, \delta)| > 5\delta n^{1.5} \kappa_{\boldsymbol{p}}$, or if Approximate-$g(\boldsymbol{p}, f, \delta)$ returns **Reject**.
- **Accept**.

The following lemma records the properties of g that will be guaranteed by our tester.

Algorithm 8: Additivity Subroutines

Procedure TESTADDITIVITY(f, δ)
  Given : Query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, threshold parameter $\delta > 0$;
  for $N_8 \leftarrow O(1)$ *times* **do**
    Sample $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, I)$;
    Reject if $|f(-\boldsymbol{x}) + f(\boldsymbol{x})| > \delta$;
    Reject if $|f(\boldsymbol{x} - \boldsymbol{y}) - (f(\boldsymbol{x}) - f(\boldsymbol{y}))\,| > \delta$;
    Reject if $\left|f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - \left(f\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > \delta$;
  Accept.

Procedure APPROXIMATE-$g(\boldsymbol{p}, f, \delta)$
  Given : $\boldsymbol{p} \in \mathbb{R}^n$, query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, threshold parameter $\delta > 0$;
  $N_8' \leftarrow O\left(\log \frac{1}{\varepsilon}\right)$;
  Sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_8'} \sim \mathcal{N}(\mathbf{0}, I)$;
  Reject if there exists $j \in [N_8']$ such that
   $|(f(\boldsymbol{p}/\kappa_{\boldsymbol{p}} - \boldsymbol{x}_1) + f(\boldsymbol{x}_1)) - (f(\boldsymbol{p}/\kappa_{\boldsymbol{p}} - \boldsymbol{x}_j) + f(\boldsymbol{x}_j))| > 2\delta$;
  return $\kappa_{\boldsymbol{p}}\,(f(\boldsymbol{p}/\kappa_{\boldsymbol{p}} - \boldsymbol{x}_1) + f(\boldsymbol{x}_1))$.

Lemma D.3. *With $r \triangleq 1/50$, if* TESTADDITIVITY*(f, δ) accepts with probability at least $1/3$, then g is a 14δ-additive function inside the small ball $\mathrm{B}(\mathbf{0}, r)$, and furthermore, for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$ it holds that*

$$\Pr_{\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},I)}[|g(\boldsymbol{p}) - (f(\boldsymbol{p} - \boldsymbol{x}) + f(\boldsymbol{x}))| \geq 4\delta] < 1/2.$$

We prove Theorem D.1 assuming that Lemma D.3 holds.

Proof of Theorem D.1. First, observe that if f is a noisy version of an additive function with noise bounded by α, then f is a δ-additive function for $\delta = 3\alpha$, and we claim Algorithm 7 always accepts. It is immediate that TESTADDITIVITY(f) always accepts. To see that f also passes the remaining tests, observe that, since f is α-close to an additive function h, point-wise, we can claim:

$$\begin{aligned}
|\kappa_{\boldsymbol{p}} g_{\boldsymbol{x}}(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - f(\boldsymbol{p})| &\leq \kappa_{\boldsymbol{p}}\,|g_{\boldsymbol{x}}(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - f(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| + |\kappa_{\boldsymbol{p}} f(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - f(\boldsymbol{p})| \\
&= \kappa_{\boldsymbol{p}} \underbrace{|f(\boldsymbol{p}/\kappa_{\boldsymbol{p}} - \boldsymbol{x}) + f(\boldsymbol{x}) - f(\boldsymbol{p}/\kappa_{\boldsymbol{p}})|}_{\leq\delta, \text{ by (28)}} + |\kappa_{\boldsymbol{p}} f(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - h(\boldsymbol{p}) + h(\boldsymbol{p}) - f(\boldsymbol{p})| \\
&\leq \kappa_{\boldsymbol{p}}\delta + |\kappa_{\boldsymbol{p}} f(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - \kappa_{\boldsymbol{p}} h(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) + h(\boldsymbol{p}) - f(\boldsymbol{p})| \qquad (\text{as } h(\boldsymbol{p}) = \kappa_{\boldsymbol{p}} h(\boldsymbol{p}/\kappa_{\boldsymbol{p}})) \\
&\leq \delta\kappa_{\boldsymbol{p}} + \kappa_{\boldsymbol{p}} \underbrace{|f(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - h(\boldsymbol{p}/\kappa_{\boldsymbol{p}})|}_{<\alpha} + \underbrace{|h(\boldsymbol{p}) - f(\boldsymbol{p})|}_{<\alpha} \\
&\leq \delta\kappa_{\boldsymbol{p}} + \alpha\kappa_{\boldsymbol{p}} + \alpha < 2\delta\kappa_{\boldsymbol{p}}.
\end{aligned}$$

Note that APPROXIMATE-$g(\boldsymbol{p}, f)$ never rejects, because we have $|g_{\boldsymbol{x}}(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - f(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| \leq \delta$.Then, by the triangle inequality,

$$|g_{\boldsymbol{x}_i}(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - g_{\boldsymbol{x}_j}(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| \leq 2|g_{\boldsymbol{x}}(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - f(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| \leq 2\delta.$$

We now show that if f is ε-far from all additive functions, then Algorithm 7 rejects with probability at least $2/3$. If TESTADDITIVITY(f) accepts with probability at most $1/3$, we can reject f with probability at least

$2/3$. Hence, we assume that TESTADDITIVITY(f) accepts with probability at least $1/3$. Then by Lemma D.3, the function g is 14δ-additive, inside $\mathrm{B}(\mathbf{0}, r)$. Using Theorem D.2, there is an additive function $h : \mathbb{R}^n \to \mathbb{R}$, which is $70\delta n^{1.5}$-close to g, on the small ball, i.e. for every $\boldsymbol{x} \in \mathrm{B}(\mathbf{0}, r)$, $|g(\boldsymbol{x}) - h(\boldsymbol{x})| \leq 70\delta n^{1.5}$. Since f is ε-far from any additive function, we have f is ε-far from h.

Now, we want to bound the probability that Step 2 of Algorithm 7 passes. First, we bound the probability that APPROXIMATE-$g(\boldsymbol{p}, f)$ fails to recover the value of $g(\boldsymbol{p})$ within an error of 4δ. That is, we bound the probability that $|g_{\boldsymbol{x}_1}(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - g_{\boldsymbol{x}_j}(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| \leq 2\delta$, for all $j \in [N_8']$ (so that it doesn't reject), but $|g(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - g_{\boldsymbol{x}_1}(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| > 4\delta$, by the probability that for all sampled vectors $\boldsymbol{x}_i, i \in [N_8']$, $|g(\boldsymbol{p}/\kappa_{\boldsymbol{p}}) - g_{\boldsymbol{x}_i}(\boldsymbol{p}/\kappa_{\boldsymbol{p}})| \geq 4\delta$. By Lemma D.3, the probability that we draw N_8' points which satisfy this, is less than $2^{-N_8'}$, which can be made $\leq \varepsilon/4$ by choosing the hidden constant in N_8' to be large enough.

Now that we have established that — in the case we obtained query access to approximate g inside the small ball — we get the correct approximation within 4δ with high probability, it remains to show that we can test whether f and h are close. After arguing that g is 14δ-additive in $\mathrm{B}(\mathbf{0}, r)$, it will follow using Theorem D.2, that g is close to an additive function $h : \mathbb{R}^n \to \mathbb{R}$ on all points inside $\mathrm{B}(\mathbf{0}, r)$. We argue that in the YES case, if f is close to some additive function then, for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$ (which contains the majority of the mass of the unknown distribution $\mathcal{D}$), we have $|f(\boldsymbol{p}) - h(\boldsymbol{p})| \leq O(n^{1.5}\delta)$.

While in the NO case, since f is far from any additive function, it is also far from h, and therefore

$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > 4515n^2\alpha] \geq \varepsilon.$$

And also

$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > 4515n^2\alpha : \boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)] \geq \frac{3}{4}\varepsilon.$$

If TESTADDITIVITY(f) passes with probability at least $1/3$, then by Lemma D.3, g will be 14δ-additive inside $\mathrm{B}(\mathbf{0}, r)$, and for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$, $\Pr_{\boldsymbol{p}\sim\mathcal{N}(\mathbf{0},I)}[|g(\boldsymbol{p}) - g_{\boldsymbol{x}}(\boldsymbol{p})| \geq 4\delta] < 1/2$. Consequently, by Theorem D.2, there would exist an additive function $h : \mathbb{R}^n \to \mathbb{R}$, such that for every $\boldsymbol{x} \in \mathrm{B}(\mathbf{0}, r)$, $|g(\boldsymbol{x}) - h(\boldsymbol{x})| \leq 70n^{1.5}\delta$. This gives us, for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)$,

$$|g(\boldsymbol{p}) - h(\boldsymbol{p})| = \left|\kappa_{\boldsymbol{p}} g\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right) - \kappa_{\boldsymbol{p}} h\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)\right| \leq 70n^{1.5}\delta\kappa_{\boldsymbol{p}} \leq 3500n^{1.5}\delta R \leq 7000n^2\delta.$$

Note that since f is ε-far from h we have

$$\begin{aligned}
&\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - g(\boldsymbol{p})| > 5\delta n^{1.5}\kappa_{\boldsymbol{p}} : \boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)] \\
&\geq \Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > 21015n^2\alpha : \boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)] \\
&\geq \frac{3\varepsilon}{4}
\end{aligned}$$

Indeed, the probability that Step 2 of Algorithm 7 fails to reject is at most

$$\begin{aligned}
&\left(\Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|f(\boldsymbol{p}) - g(\boldsymbol{p})| \leq 5\delta n^{1.5}\kappa_{\boldsymbol{p}} \vee \text{APPROXIMATE-}g(\boldsymbol{p}, f) \text{ fails to correctly recover } g(\boldsymbol{p}) : \boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)\right]\right)^{N_7} \\
&\leq \Big(1 - \Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - g(\boldsymbol{p})| > 5\delta n^{1.5}\kappa_{\boldsymbol{p}} : \boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)] \\
&+ \Pr_{\boldsymbol{p}\sim\mathcal{D}}[\text{APPROXIMATE-}g(\boldsymbol{p}, f) \text{ fails to correctly recover } g(\boldsymbol{p}) : \boldsymbol{p} \in \mathrm{B}(\mathbf{0}, R)]\Big)^{N_7} \\
&< \left(1 - \frac{3\varepsilon}{4} + \frac{\varepsilon}{4}\right)^{N_7} < \frac{1}{3},
\end{aligned}$$

by choosing the hidden constant in N_7 to be large enough. Therefore, Algorithm 7 rejects with probability at least $2/3$. □

It now remains to prove Lemma D.3, showing that if Algorithm 7 succeeds, then g is 14δ-additive inside $\mathrm{B}(\mathbf{0}, r)$, and can be well approximated in $\mathrm{B}(\mathbf{0}, r)$ with high probability by querying f on correlated points.

D.3 $O(\delta)$-Additivity of g Inside $\mathrm{B}(\mathbf{0}, r)$

First, we record the basic, but useful observation that if the TESTADDITIVITY subroutine passes, then each of its tests hold with high probability over $\mathcal{N}(\mathbf{0}, I)$.

Lemma D.4. *If* TESTADDITIVITY(f) *accepts with probability at least* $1/3$*, then*

$$\Pr_{\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},I)}[|f(-\boldsymbol{x}) + f(\boldsymbol{x})| \leq \delta] \geq \frac{999}{1000}, \tag{31}$$

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x}) + f(\boldsymbol{y})| \leq \delta] \geq \frac{999}{1000}, \tag{32}$$

$$\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\mathbf{0},I)}\left[\left|f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - f\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) - f\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right| \leq \delta\right] \geq \frac{999}{1000}. \tag{33}$$

Proof. Suppose for contradiction that at least one of (31), (32), and (33) does not hold. We here assume that (31) does not hold as other cases are similar.

We accept only when all the sampled points $\boldsymbol{x}$ satisfy $|f(-\boldsymbol{x}) + f(\boldsymbol{x})| \leq \delta$. By setting the hidden constant in N_8 to be large enough, this happens with probability at most

$$\left(\Pr_{\boldsymbol{x}\sim\mathcal{N}(0,I)}[|f(-\boldsymbol{x}) + f(\boldsymbol{x})| \leq \delta]\right)^{N_8} < \left(\frac{999}{1000}\right)^{N_8} < \frac{1}{3},$$

which is a contradiction. □

In order to argue that g is $O(\delta)$-additive on points within $\mathrm{B}(\mathbf{0}, r)$, we will rely on the fact that $\boldsymbol{p} + \boldsymbol{x}$ is approximately distributed as $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, I)$, if $\|\boldsymbol{p}\|_2$ is small. By Lemma 2.2, we have a bound on the total variation distance between $\boldsymbol{x}$ and $\boldsymbol{p} + \boldsymbol{x}$. Next, we will show that g is $O(\delta)$-additive within $\mathrm{B}(\mathbf{0}, r)$.

Lemma D.5. *Suppose that* (31) $-$ (33) *of Lemma D.4 hold. Then for every* $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$ *with* $\|\boldsymbol{p}\|_2, \|\boldsymbol{q}\|_2, \|\boldsymbol{p} + \boldsymbol{q}\|_2 \leq r$*, it holds that*

$$|g(\boldsymbol{p} + \boldsymbol{q}) - g(\boldsymbol{p}) - g(\boldsymbol{q})| \leq 14\delta.$$

The proof of this lemma will crucially rely on the following two lemmas, which say that the conclusions of Lemma D.4 hold with high probability, even when one of the points are fixed to some $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$. A consequence of this is that we will be able to query g within a small error, with high probability.

Lemma D.6. *Suppose that* (31) $-$ (33) *of Lemma D.4 hold. Then, for every* $\boldsymbol{p} \in \mathbb{R}^n$ *with* $\|\boldsymbol{p}\|_2 \leq r$*,*

$$\Pr_{\boldsymbol{x}\sim\mathcal{N}(\mathbf{0},I)}[|g(\boldsymbol{p}) - (f(\boldsymbol{p}-\boldsymbol{x}) + f(\boldsymbol{x}))| < 4\delta] \geq \frac{113}{125}. \tag{34}$$

The proof of this lemma will rely on an earlier stated theorem which provides a relationship between the majority and the median:

Lemma 4.1. *Let Ω be a sample space, $g : \Omega \to \mathbb{R}$ and $\mathcal{D}$ be a distribution over Ω. For any $\eta \in [0, 1/4]$, $\delta \in \mathbb{R}$, if $\Pr_{q_1,q_2\sim\mathcal{D}}[|g(q_1) - g(q_2)| < \delta] > 1 - \eta$, then $\Pr_{q_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(q_1)| < \delta] > 1 - 4\eta$, where $g_{\mathsf{med}} = \mathsf{med}_{q\sim\mathcal{D}}\{g(q)\}$.*

We provide a proof of this theorem in Appendix E. With this result in hand, we are ready to prove Lemma D.6.

Proof of Lemma D.6. Fix a point $\boldsymbol{p} \in \mathbb{R}^n$ with $\|\boldsymbol{p}\|_2 \le r$. We will bound the following probability, which can be thought as the approximate-collision probability.

$$A := \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|(f(\boldsymbol{p}-\boldsymbol{x}) + f(\boldsymbol{x})) - (f(\boldsymbol{p}-\boldsymbol{y}) + f(\boldsymbol{y}))| \le 4\delta].$$

Observe that

$$\begin{aligned}
1 - A &= \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\boldsymbol{x}) - f(\boldsymbol{y}) - f(\boldsymbol{p}-\boldsymbol{y}) + f(\boldsymbol{p}-\boldsymbol{x})| \ge 4\delta] \\
&\le \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{p}-\boldsymbol{y}) + f(\boldsymbol{p}-\boldsymbol{x})| > 3\delta] \\
&\quad + \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\boldsymbol{x}) - f(\boldsymbol{y}) - f(\boldsymbol{x}-\boldsymbol{y})| > \delta] && \text{(By Triangle Inequality)} \\
&< \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{p}-\boldsymbol{y}) + f(\boldsymbol{p}-\boldsymbol{x})| > 3\delta] + \frac{1}{1000} && \text{(By Lemma D.4 (32))}
\end{aligned}$$

To bound the first term, we observe, by the fact that $\boldsymbol{x}-\boldsymbol{p}, \boldsymbol{y}-\boldsymbol{p} \sim \mathcal{N}(-\boldsymbol{p}, I)$ and $\boldsymbol{p} \approx 0$, the random variables $\boldsymbol{x}-\boldsymbol{p}$ and $\boldsymbol{y}-\boldsymbol{p}$ should be distributed similarly to $\boldsymbol{x}$ and $\boldsymbol{y}$. Indeed,

$$\begin{aligned}
&\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\boldsymbol{x}-\boldsymbol{y}) - (f(\boldsymbol{p}-\boldsymbol{y}) - f(\boldsymbol{p}-\boldsymbol{x}))| > 3\delta] \\
&= \Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|f(\underbrace{\boldsymbol{x}-\boldsymbol{p}}_{\triangleq\tilde{\boldsymbol{x}}} - \underbrace{(\boldsymbol{y}-\boldsymbol{p})}_{\triangleq\tilde{\boldsymbol{y}}}) - (f(\boldsymbol{p}-\boldsymbol{y}) - f(\boldsymbol{p}-\boldsymbol{x}))| > 3\delta] \\
&= \Pr_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}\sim\mathcal{N}(-\boldsymbol{p},I)}[|f(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{y}}) - (f(-\tilde{\boldsymbol{y}}) - f(-\tilde{\boldsymbol{x}}))| > 3\delta] \\
&\le \Pr_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}\sim\mathcal{N}(\mathbf{0},I)}[|f(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{y}}) - (f(-\tilde{\boldsymbol{y}}) - f(-\tilde{\boldsymbol{x}}))| > 3\delta] + 2\,\mathrm{d_{TV}}\left(\mathcal{N}(\mathbf{0},I), \mathcal{N}(-\boldsymbol{p},I)\right) \\
&\le \Pr_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}}\sim\mathcal{N}(\mathbf{0},I)}[|f(\tilde{\boldsymbol{x}}-\tilde{\boldsymbol{y}}) - f(\tilde{\boldsymbol{x}}) + f(\tilde{\boldsymbol{y}})| > \delta] + \Pr_{\tilde{\boldsymbol{x}}\sim\mathcal{N}(\mathbf{0},I)}[|f(-\tilde{\boldsymbol{x}}) - f(\tilde{\boldsymbol{x}})| > \delta] \\
&\quad + \Pr_{\tilde{\boldsymbol{y}}\sim\mathcal{N}(\mathbf{0},I)}[|f(-\tilde{\boldsymbol{y}}) - f(\tilde{\boldsymbol{y}})| > \delta] + \frac{1}{50} && \text{(By Triangle Inequality, and Lemma 2.2)} \\
&\le \frac{3}{1000} + \frac{1}{50} \le \frac{23}{1000}. && \text{(By Lemma D.4 (31) – (33))}
\end{aligned}$$

Plugging this into our previous bound on A, we have

$$\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\mathbf{0},I)}[|g_{\boldsymbol{x}}(\boldsymbol{p}) - g_{\boldsymbol{y}}(\boldsymbol{p})| \le 4\delta] \ge 1 - \left(\frac{1}{1000} + \frac{23}{1000}\right) = 1 - \frac{3}{125},$$

Applying Lemma 4.1, we conclude that for every $\boldsymbol{p} \in \mathrm{B}(\mathbf{0}, r)$

$$\Pr_{\boldsymbol{q}\sim\mathcal{N}(\mathbf{0},I)}[|g(\boldsymbol{p}) - (f(\boldsymbol{p}-\boldsymbol{q}) + f(\boldsymbol{q}))| \le 4\delta] \ge 1 - 4\cdot\frac{3}{125} = \frac{113}{125}.$$

□

The following lemma is essentially condition (32) of Lemma D.4 with two fixed points.

Lemma D.7. *Suppose that* (31) $-$ (33) *of Lemma D.4 hold then, for every* $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$ *with* $\|\boldsymbol{p}\|_2, \|\boldsymbol{q}\|_2, \|\boldsymbol{p}+\boldsymbol{q}\| \le r$*, it holds that*

$$\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|g(\boldsymbol{p}+\boldsymbol{q}) - \left(f\left(\boldsymbol{p}-\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\boldsymbol{q}-\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > 5\delta\right] < \frac{177}{1000}.$$

Proof. Fix a pair of points $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$ with $\|\boldsymbol{p}\|_2, \|\boldsymbol{q}\|_2 \le r$. We can bound the probability

$$\begin{aligned}
&\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|g(\boldsymbol{p}+\boldsymbol{q}) - \left(f\left(\boldsymbol{p}-\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\boldsymbol{q}-\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > 5\delta\right] \\
\le &\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|g(\boldsymbol{p}+\boldsymbol{q}) - \left(f\left(\boldsymbol{p}+\boldsymbol{q}-\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > 4\delta\right] \\
+ &\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|f\left(\boldsymbol{p}+\boldsymbol{q}-\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - \left(f\left(\boldsymbol{p}-\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\boldsymbol{q}-\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > \delta\right]
\end{aligned}$$

To bound the first term, observe that if $\boldsymbol{x}, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, I)$, then the random variable $\boldsymbol{m} \triangleq (\boldsymbol{x}-\boldsymbol{y})/\sqrt{2} \sim \mathcal{N}(\boldsymbol{0}, I)$. Furthermore, because $\|\boldsymbol{p}+\boldsymbol{q}\|_2 \le r$, we can apply Lemma D.6 (34) and conclude that

$$\begin{aligned}
&\Pr_{\boldsymbol{x},\boldsymbol{y}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|g(\boldsymbol{p}+\boldsymbol{q}) - \left(f\left(\boldsymbol{p}+\boldsymbol{q}-\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > 4\delta\right] \\
= &\Pr_{\boldsymbol{m}\sim\mathcal{N}(\boldsymbol{0},I)}\left[|g(\boldsymbol{p}+\boldsymbol{q}) - (f((\boldsymbol{p}+\boldsymbol{q})-\boldsymbol{m})+f(\boldsymbol{m}))| > 4\delta\right] \le \frac{12}{125}.
\end{aligned}$$

To bound the second term, observe that

$$\begin{aligned}
&\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|f\left(\boldsymbol{p}+\boldsymbol{q}-\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - \left(f\left(\boldsymbol{p}-\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\boldsymbol{q}-\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > \delta\right] \\
= &\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|f\left(\frac{(\sqrt{2}\boldsymbol{q}+\boldsymbol{y})-(\boldsymbol{x}-\sqrt{2}\boldsymbol{p})}{\sqrt{2}}\right) - \left(f\left(\frac{(\sqrt{2}\boldsymbol{q}+\boldsymbol{y})-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{z}-(\boldsymbol{x}-\sqrt{2}\boldsymbol{p})}{\sqrt{2}}\right)\right)\right| > \delta\right] \\
= &\Pr_{\substack{\tilde{\boldsymbol{x}}\triangleq\boldsymbol{x}-\sqrt{2}\boldsymbol{p}\sim\mathcal{N}(-\sqrt{2}\boldsymbol{p},I) \\ \tilde{\boldsymbol{y}}\triangleq\boldsymbol{y}+\sqrt{2}\boldsymbol{q}\sim\mathcal{N}(\sqrt{2}\boldsymbol{q},I) \\ \boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}}\left[\left|f\left(\frac{\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{x}}}{\sqrt{2}}\right) - \left(f\left(\frac{\tilde{\boldsymbol{y}}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{z}-\tilde{\boldsymbol{x}}}{\sqrt{2}}\right)\right)\right| > \delta\right] \\
\le &\Pr_{\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|f\left(\frac{\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{x}}}{\sqrt{2}}\right) - \left(f\left(\frac{\tilde{\boldsymbol{y}}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{z}-\tilde{\boldsymbol{x}}}{\sqrt{2}}\right)\right)\right| > \delta\right] \\
&+ 2\left(\mathrm{d_{TV}}\left(\mathcal{N}(\boldsymbol{0},I),\mathcal{N}(-\sqrt{2}\boldsymbol{p},I)\right)+\mathrm{d_{TV}}\left(\mathcal{N}(\boldsymbol{0},I),\mathcal{N}(\sqrt{2}\boldsymbol{q},I)\right)\right) \\
\le &\frac{1}{1000}+\frac{\sqrt{2}}{25} < \frac{81}{1000}. \qquad \text{(By Lemma D.4 (33) and Lemma 2.2)}
\end{aligned}$$

Combining both of these bounds, we have

$$\Pr_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}\sim\mathcal{N}(\boldsymbol{0},I)}\left[\left|g\left(\boldsymbol{p}+\boldsymbol{q}\right) - \left(f\left(\boldsymbol{p}-\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right)+f\left(\boldsymbol{q}-\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)+f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| > 5\delta\right] < \frac{177}{1000}.$$

□

$O(\delta)$-additivity of g within $\mathrm{B}(\mathbf{0}, r)$ is an immediate consequence of these two lemmas.

Proof of Lemma D.5. Let $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^n$ be any pair of points satisfying $\|\boldsymbol{p}\|_2, \|\boldsymbol{q}\|_2, \|\boldsymbol{p}+\boldsymbol{q}\|_2 \leq r$. Our aim is to show that $|g(\boldsymbol{p}+\boldsymbol{q}) - g(\boldsymbol{p}) - g(\boldsymbol{q})| \leq 14\delta$. By a union bound we show that the probability that $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \sim \mathcal{N}(0, I)$ simultaneously satisfy:

$$\left|g(\boldsymbol{p}+\boldsymbol{q}) - \left(f\left(\boldsymbol{p} - \frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) + f\left(\boldsymbol{q} - \frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| < 5\delta, \tag{35}$$

$$\left|g(\boldsymbol{p}) - \left(f\left(\boldsymbol{p} - \frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right)\right)\right| < 4\delta, \tag{36}$$

$$\left|g(\boldsymbol{q}) - \left(f\left(\boldsymbol{q} - \frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| < 4\delta, \tag{37}$$

$$\left|f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - \left(f\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) - f\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| < \delta \tag{38}$$

is at least $1 - (177/1000 + 2 \cdot 12/125 + 1/1000) = 63/100 > 0$. Probabilities for (35), and (38) follow by Lemma D.7, and Lemma D.4 (32), respectively. For (36) and (37) we are using the fact that $(\boldsymbol{x}-\boldsymbol{z})/\sqrt{2}$, and $(\boldsymbol{z}-\boldsymbol{y})/\sqrt{2}$ are distributed as $\mathcal{N}(0, I)$ and apply Lemma D.6 (34). Fixing such a triple $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, we conclude that

$$\begin{aligned}
|g(\boldsymbol{p}+\boldsymbol{q}) - g(\boldsymbol{p}) - g(\boldsymbol{q})| \leq & \left|g(\boldsymbol{p}+\boldsymbol{q}) - \left(f\left(\boldsymbol{p} - \frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) + f\left(\boldsymbol{q} - \frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right)\right)\right| \\
& + \left|f\left(\boldsymbol{p} - \frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) + f\left(\boldsymbol{q} - \frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - g(\boldsymbol{p}) - g(\boldsymbol{q})\right| \\
\leq & 5\delta + \left|f\left(\boldsymbol{p} - \frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) - g(\boldsymbol{p})\right| \\
& + \left|f\left(\boldsymbol{q} - \frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right) + f\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right) - g(\boldsymbol{q})\right| \\
& + \left|f\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{\sqrt{2}}\right) - f\left(\frac{\boldsymbol{x}-\boldsymbol{z}}{\sqrt{2}}\right) - f\left(\frac{\boldsymbol{z}-\boldsymbol{y}}{\sqrt{2}}\right)\right| \leq 14\delta.
\end{aligned}$$

Therefore, g is 14δ-additive within $\mathrm{B}(\mathbf{0}, r)$. □

With this we are ready to prove Lemma D.3.

Proof of Lemma D.3. g is 14δ-additive by Lemma D.5. And, $g_{\boldsymbol{x}}(\boldsymbol{p}) = f(\boldsymbol{p}-\boldsymbol{x}) + f(\boldsymbol{x})$ is a good estimation (up to 4δ) for $g(\boldsymbol{p})$ with high probability $\left(\frac{113}{125} > \frac{1}{2}\right)$ for $x \sim \mathcal{N}(\mathbf{0}, I)$ by Lemma D.6(34). □

D.4 Multiplicatively-Approximate Distribution-Free Additivity Tester

In this section, we show that a small adaption of our tester give us a distribution-free tester for multiplicatively approximate additivity, without any precondition on the unknown distribution $\mathcal{D}$ (such as assuming that it is concentrated). After removing the condition of sampled points being inside $\mathrm{B}(\mathbf{0}, R)$, the adapted tester is represented in Algorithm 9.

We note that the subroutines in Algorithm 8 remain the same and still sample points from $\mathcal{N}(\mathbf{0}, I)$, in order to check that f satisfies the characterization properties, and to approximate g inside $\mathrm{B}(\mathbf{0}, r)$.

Algorithm 9: Distribution-Free Approximate Additivity Tester With Multiplicative Error

Procedure MULTAPPROXADDITIVITYTESTER($f, \mathcal{D}, \alpha, \varepsilon, R$)
  Given : Query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to an unknown $(\varepsilon/4, R)$-concentrated distribution $\mathcal{D}$, noise parameter $\alpha > 0$, farness parameter $\delta > 0$;
  $\delta \leftarrow 3\alpha, r \leftarrow 1/50$;
  Reject if TESTADDITIVITY(f, δ) returns **Reject**;
  for $N_9 \leftarrow O(1/\varepsilon)$ *times* **do**
    Sample $\boldsymbol{p} \sim \mathcal{D}$;
    Reject if $|f(\boldsymbol{p}) -$ APPROXIMATE-$g(\boldsymbol{p}, f, \delta)| > 5\delta n^{1.5}\kappa_{\boldsymbol{p}}$ or if APPROXIMATE-$g(\boldsymbol{p}, f, \delta)$ returns **Reject**.
  Accept.

Distribution-Free Multiplicatively-Approximate Tester for Additivity. Given query access to the input function $f : \mathbb{R}^n \to \mathbb{R}$, sampling access to unknown distribution $\mathcal{D}$, as well to $\mathcal{N}(\mathbf{0}, I)$, a parameter $0 < \alpha \in \mathbb{R}$ and a constant $0 < \varepsilon \in \mathbb{R}$, a distribution-free, multiplicative-approximate tester for additivity distinguishes between the following two cases with probability at least $2/3$:

- **YES CASE:** There exists an additive function $h : \mathbb{R}^n \to \mathbb{R}$ such that for all $\boldsymbol{p} \in \mathbb{R}^n$:
$$|f(\boldsymbol{p}) - h(\boldsymbol{p})| \leq \alpha;$$
- **NO CASE:** For any additive function $h : \mathbb{R}^n \to \mathbb{R}$:
$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[|f(\boldsymbol{p}) - h(\boldsymbol{p})| > 600n^{1.5}\alpha\kappa_{\boldsymbol{p}}] > \varepsilon.$$

Correctness of our tester, given in Algorithm 9, follows from this theorem.

Proof. The proof follows the same path as for Theorem D.1. We only adapt the Algorithm 9 to now test all points sampled by $\mathcal{D}$. In the **YES CASE**, the tests always accept. Indeed the TESTADDITIVITY(f) subroutine passes with probability 1, and we claim APPROXIMATE-$g(\boldsymbol{p}, f)$ never rejects and returns an approximate value $\kappa_{\boldsymbol{p}} g_{\boldsymbol{x}_1}\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)$, when queried $g(\boldsymbol{p}) = \kappa_{\boldsymbol{p}} g\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)$, where $\boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{0}, I)$. Recall that in the **YES CASE**, $\left|g_{\boldsymbol{x}_1}\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right) - f\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)\right| \leq \delta = 3\alpha$. Therefore we have, by triangle inequality, for every $\boldsymbol{p}, \boldsymbol{x}_1 \in \mathbb{R}^n$,

$$\left|f(\boldsymbol{p}) - \kappa_{\boldsymbol{p}} g_{\boldsymbol{x}_1}\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)\right| \leq \left|f(\boldsymbol{p}) - \kappa_{\boldsymbol{p}} f\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)\right| + \kappa_{\boldsymbol{p}}\left|f\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right) - g_{\boldsymbol{x}_1}\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)\right| \leq \alpha + \alpha\kappa_{\boldsymbol{p}} + \delta\kappa_p \leq 2\delta\kappa_p.$$

Last inequality by f being point-wise close to an additive function. Thus Step 4 in Algorithm 9 always passes.

In the **NO CASE**, we reject with probability at least $2/3$. Indeed, if TESTADDITIVITY(f) rejects with probability $> 2/3$ we are done. So, assume that it accepts with probability at least $1/3$, then we see that the premise of Lemma D.3 holds.

We first bound the probability of Step 4 of Algorithm 9 to pass. For this we use the fact that the probability that $\tilde{g}(\boldsymbol{p}) \triangleq$APPROXIMATE-$g(\boldsymbol{p}, f)$ fails to approximate g withing 6δ error is at most $\frac{\varepsilon}{2}$ as we proved for Theorem D.1.

$$\Pr_{\boldsymbol{p}\sim\mathcal{D}}[\text{Step 6 passes}] \leq \Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|f(\boldsymbol{p}) - \tilde{g}(\boldsymbol{p})| < 5\delta n^{1.5}\kappa_{\boldsymbol{p}}\right]$$

$$
\begin{aligned}
&\leq \Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|f(\boldsymbol{p})-g(\boldsymbol{p})|<20\delta n^{1.5}\kappa_p \vee |\tilde{g}(\boldsymbol{p})-g(\boldsymbol{p})|>6\delta\right]\\
&\leq 1-\Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|f(\boldsymbol{p})-g(\boldsymbol{p})|\geq 20\delta n^{1.5}\kappa_p\right]+\Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|\tilde{g}(\boldsymbol{p})-g(\boldsymbol{p})|>6\delta\right]\\
&\leq 1-\Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|f(\boldsymbol{p})-g(\boldsymbol{p})|\geq 20\delta n^{1.5}\kappa_p\right]+\frac{\varepsilon}{2}\\
&\leq 1-\frac{\varepsilon}{2}.
\end{aligned}
$$

For the last inequality we have to bound the probability that f and g are far, say

$$
\Pr_{\boldsymbol{p}\sim\mathcal{D}}\left[|f(\boldsymbol{p})-g(\boldsymbol{p})|\geq 20\delta n^{1.5}\kappa_p\right]\geq\varepsilon,
$$

for that we use Theorem D.2 to show that there exist an additive function $h:\mathbb{R}^n\to\mathbb{R}$, such that for every $\boldsymbol{x}\in\mathrm{B}(\mathbf{0},r)$, $|g(\boldsymbol{x})-h(\boldsymbol{x})|\leq 150n^{1.5}\delta$. This gives us, for every $\boldsymbol{p}\in\mathbb{R}^n$,

$$
|g(\boldsymbol{p})-h(\boldsymbol{p})|=\left|\kappa_{\boldsymbol{p}}g\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)-\kappa_{\boldsymbol{p}}h\left(\frac{\boldsymbol{p}}{\kappa_{\boldsymbol{p}}}\right)\right|\leq 150n^{1.5}\delta\kappa_{\boldsymbol{p}}.
$$

Note that since f is ε-far from any additive function, it is also ε-far from h and with probability ε we draw $\boldsymbol{p}\sim\mathcal{D}$ that satisfies $|f(\boldsymbol{p})-h(\boldsymbol{p})|>200\delta n^{1.5}\kappa_{\boldsymbol{p}}$. For these $\boldsymbol{p}$, it holds that

$$
200\delta n^{1.5}\kappa_{\boldsymbol{p}}<|f(\boldsymbol{p})-h(\boldsymbol{p})|<|f(\boldsymbol{p})-g(\boldsymbol{p})|+|g(\boldsymbol{p})-h(\boldsymbol{p})|\leq|f(\boldsymbol{p})-g(\boldsymbol{p})|+150\delta n^{1.5}\kappa_{\boldsymbol{p}},
$$

implying that $|f(\boldsymbol{p})-g(\boldsymbol{p})|>50\delta n^{1.5}\kappa_{\boldsymbol{p}}$. □

Appendix E Proof of Lemma 4.1

In this appendix we prove the following lemma which gives a sufficient condition for the median of any distribution to be close to a random element sampled from that distribution.

Lemma 4.1. *Let Ω be a sample space, $g:\Omega\to\mathbb{R}$ and $\mathcal{D}$ be a distribution over Ω. For any $\eta\in[0,1/4]$, $\delta\in\mathbb{R}$, if $\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{D}}[|g(\boldsymbol{q}_1)-g(\boldsymbol{q}_2)|<\delta]>1-\eta$, then $\Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}}-g(\boldsymbol{q}_1)|<\delta]>1-4\eta$, where $g_{\mathsf{med}}=\mathsf{med}_{\boldsymbol{q}\sim\mathcal{D}}\{g(\boldsymbol{q})\}$.*

Proof. Define $S_{\leq}\triangleq\{\boldsymbol{q}\in\Omega:g(\boldsymbol{q})\leq g_{\mathsf{med}}\}$, and $S_{\geq}\triangleq\{\boldsymbol{q}\in\Omega:g(\boldsymbol{q})\geq g_{\mathsf{med}}\}$. Since, g_{med} is the median of the set $\{g(\boldsymbol{q}):\boldsymbol{q}\in\Omega\}$ over $\boldsymbol{q}\sim\mathcal{D}$,

$$
\Pr_{\boldsymbol{q}\sim\mathcal{D}}[\boldsymbol{q}\in S_{\leq}]=\Pr_{\boldsymbol{q}\sim\mathcal{D}}[\boldsymbol{q}\in S_{\geq}]=\frac{1}{2}. \tag{39}
$$

Suppose for contradiction that the following hold:

$$
\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{D}}[|g(\boldsymbol{q}_1)-g(\boldsymbol{q}_2)|<\delta]>1-\eta, \tag{40}
$$

$$
\Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}}-g(\boldsymbol{q}_1)|<\delta]\leq 1-4\eta, \tag{41}
$$

By (41), for $\boldsymbol{q}_1\sim\mathcal{D}$, $g(\boldsymbol{q}_1)$ will be at least δ-far from g_{med} with probability more than 4η. We will argue that for $\boldsymbol{q}_2\sim\mathcal{D}$, $g(\boldsymbol{q}_2)$ will be at least δ-far from $g(\boldsymbol{q}_1)$, with probability more than η, contradicting (40).

Suppose that $\boldsymbol{q}_1 \in S_{\leq}$. Then, for any $\boldsymbol{q}_2 \in S_{\geq}$, we have

$$|g(\boldsymbol{q}_2) - g(\boldsymbol{q}_1)| \geq |g(\boldsymbol{q}_2) - g_{\mathsf{med}}| + |g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq 0 + \delta = \delta.$$

Similarly, if $\boldsymbol{q}_1 \in S_{\geq}$, then for any $\boldsymbol{q}_2 \in S_{\leq}$,

$$|g(\boldsymbol{q}_2) - g(\boldsymbol{q}_1)| \geq |g(\boldsymbol{q}_2) - g_{\mathsf{med}}| + |g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq 0 + \delta = \delta.$$

Therefore,

$$\begin{aligned}
&\Pr_{\boldsymbol{q}_1,\boldsymbol{q}_2\sim\mathcal{D}}[|g(\boldsymbol{q}_1) - g(\boldsymbol{q}_2)| \geq \delta] \\
&\geq \Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[\boldsymbol{q}_1 \in S_{\leq}] \cdot \Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq \delta \mid \boldsymbol{q}_1 \in S_{\leq}] \cdot \Pr_{\boldsymbol{q}_2\sim\mathcal{D}}[\boldsymbol{q}_2 \in S_{\geq}] \\
&+ \Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[\boldsymbol{q}_1 \in S_{\geq}] \cdot \Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq \delta \mid \boldsymbol{q}_1 \in S_{\geq}] \cdot \Pr_{\boldsymbol{q}_2\sim\mathcal{D}}[\boldsymbol{q}_2 \in S_{\leq}] \\
&= \frac{1}{4}\left(\Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq \delta \mid \boldsymbol{q}_1 \in S_{\leq}] + \Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq \delta \mid \boldsymbol{q}_1 \in S_{\geq}]\right) && \text{(By (39))} \\
&= \frac{1}{4}\left(\Pr_{\boldsymbol{q}_1\sim\mathcal{D}}[|g_{\mathsf{med}} - g(\boldsymbol{q}_1)| \geq \delta]\right) \\
&> \frac{4\eta}{4} = \eta. && \text{(By (41))}
\end{aligned}$$

□