Distribution-Free Testing of Linear Functions on $\mathbb{R}^n$

Noah Fleming[1]
University of Toronto, Toronto, Canada
noahfleming@cs.toronto.edu

Yuichi Yoshida
National Institute of Informatics, Tokyo, Japan
yyoshida@nii.ac.jp

Abstract

We study the problem of testing whether a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is linear (i.e., both additive and homogeneous) in the *distribution-free* property testing model, where the distance between functions is measured with respect to an unknown probability distribution over $\mathbb{R}^n$. We show that, given query access to f, sampling access to the unknown distribution as well as the standard Gaussian, and $\varepsilon > 0$, we can distinguish additive functions from functions that are ε-far from additive functions with $O\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$ queries, independent of n. Furthermore, under the assumption that f is a continuous function, the additivity tester can be extended to a distribution-free tester for linearity using the same number of queries. On the other hand, we show that if we are only allowed to get values of f on sampled points, then any distribution-free tester requires $\Omega(n)$ samples, even if the underlying distribution is the standard Gaussian.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms

Keywords and phrases Property Testing, Distribution-Free Testing, Linearity Testing

Digital Object Identifier 10.4230/LIPIcs.ITCS.2020.22

Funding *Noah Fleming*: Supported by MITACS JSPS and NSERC.
Yuichi Yoshida: Supported by JSPS KAKENHI Grant Number JP17H04676 and JP18H05291.

1 Introduction

Property testing of Boolean functions studies the problem where, given query access to a function $f\colon \{0,1\}^n \to \{0,1\}$ and a parameter $\varepsilon > 0$, the goal is to distinguish with high probability the case that f satisfies some predetermined property P from the case that f is ε-far from satisfying P. That is, whether we need to change the values of $f(x)$ for at least an ε-fraction of $x \in \{0,1\}^n$ before f satisfies P. Since the seminal work by Blum, Luby and Rubinfeld [11], property testing has become a thriving field, and many properties of Boolean functions have been shown to be testable with a number of queries independent of n, including linear functions [11], low-degree polynomials [8, 26] and k-juntas [9, 10, 18]. For an introductory survey, we recommend [21].

In contrast to Boolean functions, only a few properties of functions on a Euclidean space, that is, $\mathbb{R}^n$, have been studied. For a measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$, $\varepsilon > 0$, and a property P, we say that f is *ε-far* from P if

$$\Pr_{x\sim\mathcal{N}(0,I)}[f(x) \neq g(x)] > \varepsilon,$$

for any measurable function $g\colon \mathbb{R}^n \to \mathbb{R}$ satisfying P, where $\mathcal{N}(0,I)$ is the standard Gaussian. We say that an algorithm is a *tester* for a property P if, given query access to a measurable function

[1] Work done while visiting the National Institute of Informatics.

$f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to the standard Gaussian, and $\varepsilon > 0$, it accepts with probability at least $2/3$ when f satisfies P, and rejects with probability at least $2/3$ when f is ε-far from P. Testability of a variety of properties has been considered, including surface area of a set [29, 34], half spaces [31–33], linear separators [3], high-dimensional convexity [13], and linear k-junta [15].

Although the standard Gaussian is natural, it rarely appears in practice. In fact, we typically have little, if any, information about the underlying distribution. This raises the question of whether we can test when the underlying distribution of the data is unknown. For a measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$, $\varepsilon > 0$, a distribution $\mathcal{D}$ over $\mathbb{R}^n$, and a property P, we say that f is *ε-far* from P with respect to $\mathcal{D}$ if

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

for any measurable function $g\colon \mathbb{R}^n \to \mathbb{R}$ satisfying P. We say that an algorithm is a *distribution-free tester* for a property P if, given query access to a measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to an *unknown* distribution $\mathcal{D}$ over $\mathbb{R}^n$ as well as the standard Gaussian, and $\varepsilon > 0$, it accepts with probability at least $2/3$ when f satisfies P, and rejects with probability at least $2/3$ when f is ε-far from P with respect to $\mathcal{D}$. Distribution-free property testing is an attractive model because it makes minimal assumptions on the environment, and models the scenario most often occurring in practice.

We say that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *additive* if $f(x) + f(y) = f(x+y)$ for any $x, y \in \mathbb{R}^n$. In this work, we consider distribution-free testing of additivity of functions $f\colon \mathbb{R}^n \to \mathbb{R}$ and show the following.

► **Theorem 1.** *There exists a one-sided error distribution-free tester for additivity of $f\colon \mathbb{R}^n \to \mathbb{R}$ with $O\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$ queries.*

Previously no algorithm was known even when the underlying distribution $\mathcal{D}$ is the standard Gaussian. As there is a trivial lower bound of $\Omega\left(\frac{1}{\varepsilon}\right)$, the query complexity of our tester is almost tight.

We say that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *homogeneous* if $cf(x) = f(cx)$ for any $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$. A function that is both additive and homogeneous is said to be *linear*. Although additivity and linearity are equivalent for functions over finite groups, there are (pathological) functions $f\colon \mathbb{R}^n \to \mathbb{R}$ that are additive but not homogeneous. Hence, the testability of additivity does not immediately imply the testability of linearity. However, when the input function is guaranteed to be continuous, we can also test linearity.

► **Theorem 2.** *Suppose that the input function is guaranteed to be continuous. Then, there exists a one-sided error distribution-free tester for linearity with $O\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$ queries.*

It is also natural to assume that we can get values of the input function only on sampled points. Specifically, we say that a (distribution-free) tester is *sample-based* if it accesses the input function $f\colon \mathbb{R}^n \to \mathbb{R}$ through points sampled from the distributions $\mathcal{D}$ and $\mathcal{N}(0, I)$. We show a strong lower bound for sample-based testers.

► **Theorem 3.** *Any sample-based tester for the linearity of functions $f\colon \mathbb{R}^n \to \mathbb{R}$ requires $\Omega(n)$ samples, even when $\mathcal{D} = \mathcal{N}(0, I)$.*

This lower bound is tight; it is not difficult to see that $O(n)$ samples suffices to test linearity. Indeed $O(n)$ samples will, with high probability, contain n linearly independent vectors. The evaluations of f on these vectors uniquely determines the linear function. This theorem shows a sharp contrast between query-based and sample-based testers for properties of functions on a Euclidean space. We note that we can show the same lower bound for testing additivity with an almost identical proof.

1.1 Related Work

The question of property testing first appeared (implicitly) in the work of Blum, Luby and Rubinfeld [11]. Among the problems that they studied was linearity testing. Their algorithm, now famously known as the BLR test, has played a key role in the design of probabilistically checkable proofs [2, 5, 25] and this connection was some of the early motivation for the field of property testing. Since the original paper, the parameters of the BLR test have been extensively refined. Much of this work focused on reducing the amount of randomness, due to this being a key parameter in probabilistically checkable proofs, as well as analyzing the rejection probability (see [36] for a survey). Another line of works considered the testing linearity over more general domains. The works of [7, 11, 35] showed that the BLR test can be used to test the linearity of any function with $f\colon G \to H$ for finite groups G and H with $O(1/\varepsilon)$ queries. Following this, a body of work [1, 17, 19, 27] constructed testers for linearity of functions $f\colon S \to \mathbb{R}$, where S is a finite subset of rational numbers, and the distance is measured with respect to the uniform distribution over S. See [28] for a survey. These results were phrased in terms of approximate self-testing and correcting programs. In this setting the queries to f return a finite approximation of $f(x)$. Although these results are arguably the most related to our work, our proof differs significantly from theirs and instead takes inspiration from the original BLR test.

Distribution-free testing (for graph properties) was first defined by Goldreich et al. [22], though the first distribution-free testers for non-trivial properties appeared much later in the work of Halevy and Kushilevitz [23]. Subsequently, distribution-free testers have been considered for a variety of Boolean functions including low-degree polynomials, dictators, and monotone functions [23], k-juntas [6, 12, 23, 30], conjunctions, decision lists, and linear threshold functions [20], monotone and non-monotone monomials [16], and monotone conjunctions [14, 20]. However, to our knowledge the only (partial) distribution-free tester for a class of functions on the Euclidean space is due to Harms [24] who gave an efficient tester for half spaces, that is, functions $f\colon \mathbb{R}^n \to \{0, 1\}$ of the form $f(x) = \mathrm{sgn}(w^\top x - \theta)$ for some $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$, over any rotationally invariant distribution.

1.2 Proof Technique

The construction of our tester for additivity will be done in two steps. First, we construct a constant-query tester for additivity over the standard Gaussian distribution $\mathcal{N}(0, I)$. Our tester will accept linear functions with probability 1, and so the majority of the work is in showing that if the test accepts the given function $f\colon \mathbb{R}^n \to \mathbb{R}$ with high probability, then f is close to an additive function. To do so, we show that if f passes a series of tests then there exists a related function $g\colon \mathbb{R}^n \to \mathbb{R}$, defined from f, which is additive. Furthermore, if f is linear then $f = g$. The definition of g will allow us to obtain query access to it with high probability, and so we can simply estimate the distance between f and g. At a high-level, this is somewhat similar to the BLR test, however operating over $\mathcal{N}(0, I)$ rather than the uniform distribution presents its own set of non-trivial challenges. We discuss these, as well as the definition of g at the start of Section 3.1.

It is fairly straightforward to generalize this tester for additivity to a distribution-free tester. To do so, we run the additivity tester for the standard Gaussian, except that testing the distance between f and g will now be done using samples from the unknown $\mathcal{D}$. This crucially relies on our ability to draw samples from the standard Gaussian.

Any additive function $f\colon \mathbb{R}^n \to \mathbb{R}$ is linear over the rationals, meaning that $f(qx) = qf(x)$ for every $q \in \mathbb{Q}$. Therefore, in order to test linearity it remains to test whether this holds also for irrationals. Assuming that f is continuous we are able to modify our tester to show that this implies that the additive function g is continuous as well. We then leverage the fact that any continuous additive function is linear in order to obtain our linearity tester.

To prove Theorem 3, the lower bound on sample-based testers for linearity, we construct two distributions, one supported on linear functions, and the other supported on functions which are far from linear. Consider drawing a function f from one of these two distributions with equal probability. By Yao's minimax principle it suffices to show that any deterministic algorithm which receives n samples from $\mathcal{N}(0, I)$, together with their evaluations on f, is unable to distinguish, with high probability, which of the two distribution f came from. To construct the distribution on linear functions, we sample $w \sim \mathcal{N}(0, I)$ and return $f(x) := w^\top x$. Our distribution on functions which are far from linear is designed so that any function f from this distribution satisfies $f(x+y) \neq f(x)+f(y)$ with probability 1 over $x, y \sim \mathcal{N}(0, I)$. To do so, for every $x \in \mathbb{R}^n$ we sample ε_x from a one-dimensional Gaussian and return $f(x) := w^\top x + \varepsilon_x$. It is not difficult to show that such functions are far from linear.

1.3 Organization

The remainder of the paper is organized as follows. In Section 2 we review several useful facts about probability distributions. In Section 3 we develop our distribution-free tester for additivity by first constructing a tester for additivity over the standard Gaussian in Section 3.1. We generalize this tester to the distribution-free setting in Section 3.2 and to a tester for linearity in Section 4. Finally, we end with our lower bound on the sampling model in Section 5.

2 Preliminaries

Let $\mathcal{D}$ and $\mathcal{D}'$ be probability distributions on the same domain Ω. Then, the *total variation distance* between them, denoted by $\mathrm{d_{TV}}(\mathcal{D}, \mathcal{D}')$, is defined as

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

The *Kullback-Leibler divergence* (or *KL-divergence*) of $\mathcal{D}'$ from $\mathcal{D}$, denoted $\mathrm{d_{KL}}(\mathcal{D} \| \mathcal{D}')$, is defined as

$$\mathrm{d_{KL}}(\mathcal{D} \| \mathcal{D}') = \int_\Omega \mathcal{D}(x) \log\left(\frac{\mathcal{D}(x)}{\mathcal{D}'(x)}\right) dx.$$

We will use the KL-divergence to upper bound the total variation distance, using the following inequality.

▶ **Theorem 4** (Pinsker's Inequality). *Let $\mathcal{D}$ and $\mathcal{D}'$ be probability distributions on the same domain Ω. Then,*

$$\mathrm{d_{TV}}(\mathcal{D}, \mathcal{D}') \leq \sqrt{\frac{1}{2} \mathrm{d_{KL}}(\mathcal{D} \| \mathcal{D}')}.$$

The following allows us to bound the KL-divergence between two Gaussian distributions.

▶ **Lemma 5.** *Let $\mathcal{D} = \mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{D}' = \mathcal{N}(\mu_2, \Sigma_2)$ be multivariate Gaussian distributions with $\mu_1, \mu_2 \in \mathbb{R}^n$ and invertible $\Sigma_1, \Sigma_2 \in \mathbb{R}^{n \times n}$. Then,*

$$\mathrm{d_{KL}}(\mathcal{D} \| \mathcal{D}') = \frac{1}{2}\left(\log\left(\frac{\det \Sigma_2}{\det \Sigma_1}\right) + \mathrm{tr}\left((\Sigma_2)^{-1}\Sigma_1\right) - n + (\mu_2 - \mu_1)^\top \Sigma_2^{-1}(\mu_2 - \mu_1)\right).$$

We record a useful lemma about total variation distance of Gaussians with shared covariance matrices.

▶ **Lemma 6.** *Consider two Gaussian distributions $\mathcal{N}(\mu_1, \Sigma), \mathcal{N}(\mu_2, \Sigma)$ with shared invertible covariance matrices $\Sigma \in \mathbb{R}^{n\times n}$. Then $\mathrm{d_{TV}}(\mathcal{N}(\mu_1, \Sigma), \mathcal{N}(\mu_2, \Sigma)) \leq \phi$ holds if $\|\mu_1 - \mu_2\|_2 \leq 2\phi/\sqrt{\|\Sigma^{-1}\|_2}$.*

Proof. Denote $\mu := \mu_1 - \mu_2$. By Lemma 5, $\mathrm{d_{TV}}(\mathcal{N}(\mu_1, \Sigma), \mathcal{N}(\mu_2, \Sigma)) = \sqrt{\frac{1}{4}\mu^\top \Sigma^{-1}\mu}$. Now, because Σ is PSD, $\mu^\top\Sigma^{-1}\mu \leq \|\mu\|_2^2\|\Sigma^{-1}\|_2$, where $\|\cdot\|_2$ is the spectral matrix norm. Therefore, we have $\mathrm{d_{TV}}(\mathcal{N}(\mu_1, \Sigma), \mathcal{N}(\mu_2, \Sigma)) \leq \frac{1}{2}\|\mu\|_2\sqrt{\|\Sigma^{-1}\|_2} \leq \phi$.

◀

3 Testing Additivity

In this section, we develop our distribution-free tester for additivity. For convenience, we first describe a simpler tester for additivity over the standard Gaussian distribution $\mathcal{N}(0, I)$ in Section 3.1. Then, in Section 3.2, we describe how to generalize this algorithm to test additivity over an unknown distribution.

3.1 Tester for the Standard Gaussian

Our goal in this section is to design a constant-query tester for the additivity of a measurable function $f: \mathbb{R}^n \to \mathbb{R}$ over the standard Gaussian.

▶ **Theorem 7.** *There exists a one-sided error $O\left(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon}\right)$-query tester for additivity over the standard Gaussian.*

At a high-level, our tester consists of two steps. First, we test whether f satisfies additivity over a set of samples drawn from the distribution. If f passes this test, then we conclude that there must be an additive function $g: \mathbb{R}^n \to \mathbb{R}$, which is a self-corrected version of f. Second, by testing the value of f on a correlated set of points, we are able to get query access to g with high probability, and therefore we can simply estimate the distance between f and g. Our tester relies on the fact that it has one-sided error: if f is additive then our test passes with probability 1. Otherwise, if f is non-additive and the second step passes with high probability, then f and g must be close.

The first step is inspired by the BLR test. Indeed, the evaluation of the function g at a point p is defined as the (weighted) majority value of $f(p - x) + f(x)$ over all $x \sim \mathcal{N}(0, I)$ (where, $f(p - x) + f(x)$ is weighted according to the probability of drawing $x \sim \mathcal{N}(0, I)$). However, there are some significant challenges in generalizing the BLR test to the standard Gaussian, the most obvious of which is that unlike the uniform distribution, every point in the support of the distribution does not have equal probability. In particular, $p - x$ is not distributed as $x \sim \mathcal{N}(0, I)$ for fixed $p \neq 0$. In order to overcome this, we exploit the fact that for additive functions f, we have $f(x) = qf(x/q)$ for every rational q. This allows us to restrict attention to a small ball $\mathrm{B}(0, 1/r)$ of radius $1/r$ centered at the origin. Then, for $p \in \mathrm{B}(0, 1/r)$, $p - x$ is approximately distributed as x for small enough $1/r$. Thus, we get around the issue of unevenly weighted points by defining g within $\mathrm{B}(0, 1/r)$, and then extrapolating to define g over $\mathbb{R}^n$.

Concretely, we will define g as follows. First, let r be a sufficiently large integer ($r = 50$ suffices). For each point $p \in \mathbb{R}^n$ define

$$k_p := \begin{cases} 1 & \text{if } \|p\|_2 \leq 1/r, \\ \lceil r \cdot \|p\|_2 \rceil & \text{if } \|p\|_2 > 1/r. \end{cases}$$

Now, define $g: \mathbb{R}^n \to \mathbb{R}$ as

$$g(p) := k_p \cdot \operatorname*{maj}_{\mathcal{N}(0,I)} \left[f\left(\frac{p}{k_p} - x\right) + f\left(x\right) \right],$$

Algorithm 1: Standard Gaussian Additivity Tester

Given : Query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to the distribution $\mathcal{N}(0, I)$;

Reject if TESTADDITIVITY(f) returns **Reject**;

for $N_1 := O(1/\varepsilon)$ *times* **do**

- Sample $p \sim \mathcal{N}(0, I)$;
- **Reject** if $f(p) \neq$ QUERY-$g(p, f)$ or if QUERY-$g(p, f)$ returns **Reject**.

Accept.

Algorithm 2: Subroutines

Procedure TESTADDITIVITY(f)

- **Given :** Query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to the distribution $\mathcal{N}(0, I)$;
- **for** $N_2 := O(1)$ *times* **do**
  - Sample $x, y, z \sim \mathcal{N}(0, I)$;
  - **Reject** if $f(-x) \neq -f(x)$;
  - **Reject** if $f(x-y) \neq f(x) - f(y)$;
  - **Reject** if $f\left(\frac{x-y}{2}\right) \neq f\left(\frac{x-z}{2}\right) + f\left(\frac{z-y}{2}\right)$;
- **Accept**.

Procedure QUERY-$g(p, f)$

- **Given :** $p \in \mathbb{R}^n$, query access to $f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to $\mathcal{N}(0, I)$;
- $N_2' := O(\log \frac{1}{\varepsilon})$;
- Sample $x_1, \ldots, x_{N_2'} \sim \mathcal{N}(0, I)$;
- **Reject** if there exists $i, j \in [N_2']$ such that $f(p/k_p - x_i) + f(x_i) \neq f(p/k_p - x_j) + f(x_j)$;
- **return** $k_p\left(f(p/k_p - x_1) + f(x_1)\right)$.

where $\mathrm{maj}_{\mathcal{N}(0,I)}$ is the *weighted majority function* where a value $f(p/k_p - x) + f(x)$ is weighted according to its probability mass under $x \sim \mathcal{N}(0, I)$. Observe that either $p \in \mathrm{B}(0, 1/r)$, or $g(p)$ first maps p to a point p/k_p in $\mathrm{B}(0, 1/r)$. The value of g is the most likely value (according to $\mathcal{N}(0, I)$) of $f(p/k_p - x) + f(x)$. If f is close to additive, then taking this majority should allow us to correct for the errors in f.

An equivalent definition of g which will be useful is the following. For $p \in \mathbb{R}^n$ let P_p be the Lebesgue measurable function such that $\int_A P_p(x)dx$ gives the probability (over $\mathcal{N}(0, I)$) that $f(p/k_p - x) + f(x)$ takes value in A. Then g is defined as $g(p) := \mathrm{argmax}_x\, P_p(x)$ if $P_p(x) \geq 1/2$.

Our algorithm is given in Algorithm 1, which uses subroutines given in Algorithm 2. The QUERY-g subroutine allows us to obtain query access to g with high probability, while the TESTADDITIVITY subroutine tests the conditions that we require in order to prove that g is additive.

▶ **Lemma 8.** *If* TESTADDITIVITY*(f) accepts with probability at least* $1/10$*, then g is a well-defined, additive function, and furthermore,* $\Pr_{x \sim \mathcal{N}(0,I)}[g(p) \neq k_p(f(p/k_p - x) + f(x))] < 1/2$.

We first prove Theorem 7 assuming that Lemma 8 holds.

Proof of Theorem 7. First, observe that if f is an additive function then Algorithm 1 always accepts. Indeed, it is immediate that TESTADDITIVITY(f) always accepts. To see that f also passes the remaining tests, observe that by additivity, $k_p\left(f(p/k_p - x) + f(x)\right) = k_p f(p/k_p) = f(p)$, where the final inequality holds because $k_p \in \mathbb{Z}$ and by homogeneity over the rationals $f(qx) = qf(x)$ for every $q \in \mathbb{Q}$.

We now show that if f is ε-far from all additive functions then Algorithm 1 rejects with probability at least $2/3$. If TESTADDITIVITY(f) accepts with probability at most $1/10$, we can reject f with probability at least $1 - 1/10 > 2/3$. Hence, we assume that TESTADDITIVITY(f) accepts with probability at least $1/10$. Then by Lemma 8, the function g is additive and hence f is ε-far from g. Now, we want to bound the probability that Step 2 of Algorithm 1 passes.

First, we bound the probability that QUERY-$g(p, f)$ fails to recover the value of $g(p)$. That is, we bound the probability that $f(p/k_p - x_i) + f(x_i) = f(p/k_p - x_j) + f(x_j)$ for all $i, j \in [N_2']$, but $g(p) \neq k_p\left(f(p/k_p - x_i) + f(x_i)\right)$. By Lemma 8, the probability that we draw N_2' points which satisfy this is at most $2^{-N_2'} \leq \varepsilon/2$ by choosing the hidden constant in N_2' to be large enough. Therefore, the probability that we correctly recover $g(p)$ is at least $1 - \varepsilon/2$.

Now that we have established that we can obtain query access to g with high probability, it remains to show that we can test whether f and g are close. Indeed, the probability that Step 2 of Algorithm 1 fails to reject is at most

$$
\begin{aligned}
&\left(\Pr_{p \sim \mathcal{N}(0,I)}[f(p) = g(p) \vee \text{QUERY-}g(p,f) \text{ fails to correctly recover } g(p)]\right)^{N_1} \\
&\leq \left(1 - \Pr_{p \sim \mathcal{N}(0,I)}[f(p) \neq g(p)] + \Pr_{p \sim \mathcal{N}(0,I)}[\text{QUERY-}g(p,f) \text{ fails to correctly recover } g(p)]\right)^{N_1} \\
&< \left(1 - \frac{\varepsilon}{2}\right)^{N_1} < \frac{1}{10},
\end{aligned}
$$

by choosing the hidden constant in N_1 to be large enough. Therefore, Algorithm 1 rejects with probability at least $1 - 1/10 > 2/3$. ◀

It remains to prove Lemma 8 showing that if Algorithm 1 succeeds, then g is an additive function with high probability.

3.1.1 Additivity of the Function g

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

▶ **Lemma 9.** *If* TESTADDITIVITY*(f) accepts with probability at least* $1/10$*, then*

$$\Pr_{x,y \sim \mathcal{N}(0,I)}[f(x-y) = f(x) - f(y)] \geq \frac{99}{100}, \tag{1}$$

$$\Pr_{x \sim \mathcal{N}(0,I)}[f(-x) = -f(x)] \geq \frac{99}{100}, \tag{2}$$

$$\Pr_{x,y,z \sim \mathcal{N}(0,I)}\left[f\left(\frac{x-y}{2}\right) = f\left(\frac{x-z}{2}\right) + f\left(\frac{z-y}{2}\right)\right] \geq \frac{99}{100}. \tag{3}$$

Proof. Suppose for contradiction that at least one of (1), (2), and (3) does not hold. We here assume that (1) does not hold as other cases are similar.

We accept only when all the sampled pairs (x, y) satisfy $f(x + y) = f(x) + f(y)$. By setting the hidden constant in N_2 to be large enough, this happens with probability at most

$$\left(1 - \Pr_{x,y \sim \mathcal{N}(0,I)}[f(x+y) \neq f(x) + f(y)]\right)^{N_2} < \left(\frac{99}{100}\right)^{N_2} < \frac{1}{10},$$

which is a contradiction. ◀

In order to argue that g is additive, we will first argue that g is additive on points within the tiny ball $\mathrm{B}(0,1/r)$. To do so, we will crucially use the fact that $p-x$ is distributed approximately as $x \sim \mathcal{N}(0,I)$ if $\|p\|_2$ is small. By Lemma 6 we have a bound on the total variation distance between x and $x+p$.

▷ Claim 10. Let $p \in \mathbb{R}^n$ satisfying $\|p\|_2 \leq k/r$ for some $k \in \mathbb{Z}^{>0}$. Then $\mathrm{d_{TV}}(\mathcal{N}(0,I), \mathcal{N}(p,I)) \leq k/100$.

Proof. By Lemma 6, for $\mathrm{d_{TV}}(\mathcal{N}(0,I), \mathcal{N}(p,I)) \leq k/100$ it is enough to show that p satisfies $\|0-p\|_2 \leq 2k/(100\sqrt{\|I\|_2})$. Because $\|p\|_2 \leq k/r \leq 2k/100 = 2k/(100\sqrt{\|I\|_2})$. ◁

After arguing that g is additive in $\mathrm{B}(0,1/r)$, it will follow that g is additive elsewhere because g is defined by extrapolating the value of g within this ball. Therefore, we will focus on proving the additivity of g within $\mathrm{B}(0,1/r)$.

▶ **Lemma 11.** *Suppose that* (1) – (3) *of Lemma 9 hold. For every $p, q \in \mathbb{R}^n$ with $\|p\|_2, \|q\|_2, \|p+q\|_2 \leq 1/r$ it holds that $g(p+q) = g(p) + g(q)$.*

The proof of this lemma will crucially rely on the following two lemmas, which say that the conclusions of Lemma 9 hold with high probability even when one of the points are fixed to a point $\mathrm{B}(0,1/r)$. A consequence of this is that g is well-defined.

▶ **Lemma 12.** *Suppose that* (1) – (3) *of Lemma 9 hold, then g is well-defined, and for every $p \in \mathbb{R}^n$ with $\|p\|_2 \leq 1/r$,*

$$\Pr_{x\sim\mathcal{N}(0,I)}[g(p) = f(p-x) + f(x)] \geq \frac{9}{10}.$$

Proof. Fix a point $p \in \mathbb{R}^n$ with $\|p\|_2 \leq 1/r$. We will bound the following probability.

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

Observe that

$$\begin{aligned}
A &= \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x) - f(y) \neq f(p-y) - f(p-x)] \\
&\leq \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x) - f(y) \neq f(x-y)] + \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x-y) \neq f(p-y) - f(p-x)] \\
&< \frac{1}{100} + \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x-y) \neq f(p-y) - f(p-x)] && \text{(By Lemma 9)}
\end{aligned}$$

It remains to bound the second term. Intuitively, because $x-p, y-p \sim \mathcal{N}(-p,I)$ and $p \approx 0$, the random variables $p-x$ and $p-y$ should be distributed similarly to x and y. Indeed,

$$\begin{aligned}
&\Pr_{x,y\sim\mathcal{N}(0,I)}[f(x-y) \neq f(p-y) - f(p-x)] \\
&= \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x-p+p-y) \neq f(p-y) - f(p-x)] \\
&= \Pr_{x,y\sim\mathcal{N}(-p,I)}[f(x-y) \neq f(-y) - f(-x)] \\
&\leq \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x-y) \neq f(-y) - f(-x)] + 2\,\mathrm{d_{TV}}\Big(\mathcal{N}(0,I), \mathcal{N}(-p,I)\Big) \\
&\leq \Pr_{x,y\sim\mathcal{N}(0,I)}[f(x-y) \neq f(x) - f(y)] + \frac{2}{100} + 2\Pr_{x\sim\mathcal{N}(0,I)}[f(-x) \neq f(x)] && \text{(Claim 10)} \\
&\leq \frac{3}{100} + \frac{2}{100} = \frac{5}{100}. && \text{(By (1) and (2) in Lemma 9)}
\end{aligned}$$

Plugging this into our previous bound on A, we can conclude that

$$A \geq 1 - \left(\frac{1}{100} + \frac{5}{100}\right) = 1 - \frac{6}{100} > \frac{9}{10}.$$

Next, we bound A above in terms of the probability that $g(p) \neq f(p-x) + f(x)$. Define $P_p \colon \mathbb{R}^n \to \mathbb{R}^+$ to be the bounded Lebesgue-measurable function such that $\int_B P_p(x)dx$ is the probability that $f(p-x)+f(x)$ takes value in the (measurable) set B. By Hölder's inequality with $p = 1, q = \infty$ we have

$$A = \int_{\mathbb{R}} P_p^2(x)dx \leq \|P_p\|_\infty \int_{\mathbb{R}} P_p(x)dx = \|P_p\|_\infty,$$

where the last equality follows because P_p is a density and $\int_{\mathbb{R}} P_p(x)dx = 1$ holds. Therefore,

$$\frac{9}{10} \leq A \leq \|P_p\|_\infty.$$

Because $\operatorname{argmax}_x P_p(x) \geq 9/10 > 1/2$, we have $g(p) = \operatorname{argmax}_x P_p(x)$ and hence $\Pr_{x\sim\mathcal{N}(0,I)}[g(p) = f(p-x) + f(x)] \geq 9/10$. ◀

The following lemma is essentially condition (3) of Lemma 9 with two fixed points.

▶ **Lemma 13.** *Suppose that* (1) – (3) *of Lemma 9 hold then, for every* $p, q \in \mathbb{R}^n$ *with* $\|p\|_2, \|q\|_2, \|p+q\| \leq 1/r$*,*

$$\Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[g(p+q) \neq f\Big(p - \frac{x-z}{2}\Big) + f\Big(q - \frac{z-y}{2}\Big) + f\Big(\frac{x-y}{2}\Big)\right] \leq \frac{2}{10}.$$

Proof. Fix a pair of points $p, q \in \mathbb{R}^n$ with $\|p\|_2, \|q\|_2 \leq 1/r$. We can bound the probability

$$\begin{aligned}
&\Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[g(p+q) \neq f\Big(p - \frac{x-z}{2}\Big) + f\Big(q - \frac{z-y}{2}\Big) + f\Big(\frac{x-y}{2}\Big)\right] \\
&\leq \Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[g(p+q) \neq f\Big(p + q - \frac{x-y}{2}\Big) + f\Big(\frac{x-y}{2}\Big)\right] \\
&\quad + \Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[f\Big(p+q - \frac{x-y}{2}\Big) \neq f\Big(p - \frac{x-z}{2}\Big) + f\Big(q - \frac{z-y}{2}\Big)\right]
\end{aligned}$$

To bound the first term, observe that if $x, y \sim \mathcal{N}(0, I)$, then the random variable $(x-y)/2$ is also distributed according to $\mathcal{N}(0, I)$. Furthermore, because $\|p+q\|_2 \leq 1/r$, we can apply Lemma 12 and conclude that

$$\Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[g(p+q) \neq f\Big(p + q - \frac{x-y}{2}\Big) + f\Big(\frac{x-y}{2}\Big)\right] \leq \frac{1}{10}.$$

To bound the second term, observe that

$$\begin{aligned}
&\Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[f\Big(p+q - \frac{x-y}{2}\Big) \neq f\Big(p - \frac{x-z}{2}\Big) + f\Big(q - \frac{z-y}{2}\Big)\right] \\
&= \Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[f\left(\frac{(2q+y)-(x-2p)}{2}\right) \neq f\left(\frac{(2q+y)-z}{2}\right) + f\left(\frac{z-(x-2p)}{2}\right)\right] \\
&= \Pr_{\substack{x\sim\mathcal{N}(-2p,I)\\ y\sim\mathcal{N}(2q,I)\\ z\sim\mathcal{N}(0,1)}}\left[f\Big(\frac{y-x}{2}\Big) \neq f\Big(\frac{y-z}{2}\Big) + f\Big(\frac{z-x}{2}\Big)\right] \\
&\leq \Pr_{x,y,z\sim\mathcal{N}(0,I)}\left[f\Big(\frac{x-y}{2}\Big) \neq f\Big(\frac{x-z}{2}\Big) + f\Big(\frac{z-y}{2}\Big)\right] + \mathrm{d_{TV}}\Big(\mathcal{N}(0,I), \mathcal{N}(-2p,I)\Big) \\
&\quad + \mathrm{d_{TV}}\Big(\mathcal{N}(0,I), \mathcal{N}(2q,I)\Big) \\
&\leq \frac{1}{100} + \frac{2}{100} + \frac{2}{100} = \frac{5}{100}. \qquad \text{(By Lemma 9 and Claim 10)}
\end{aligned}$$

Combining both of these bounds, we have $\Pr_{x,y,z\sim\mathcal{D}}[g(p+q) \neq f(p - \frac{x-z}{2}) + f(q - \frac{z-y}{2}) + f(\frac{x-y}{2})] \leq 1/10 + 5/100 \leq 2/10$. ◀

The additivity of g within $\mathrm{B}(0, 1/r)$ is an immediate consequence of these two lemmas.

Proof of Lemma 11. Let $p, q \in \mathbb{R}^n$ be any pair of points satisfying $\|p\|_2, \|q\|_2, \|p+q\|_2 \leq 1/r$. Our aim is to show that $g(p+q) = g(p) + g(q)$. By a union bound over Lemmas 12 and 13, the probability that $x, y, z \sim \mathcal{N}(0, I)$ simultaneously satisfy

1. $g(p+q) = f(p - \frac{x-z}{2}) + f(q - \frac{z-y}{2}) + f(\frac{x-y}{2})$,
2. $g(p) = f(p - \frac{x-z}{2}) + f(\frac{x-z}{2})$,
3. $g(q) = f(q - \frac{z-y}{2}) + f(\frac{z-y}{2})$,
4. $f(\frac{x-y}{2}) = f(\frac{x-z}{2}) - f(\frac{z-y}{2})$

is at least $1 - (2/10 + 2 \cdot 1/10 + 1/10) > 0$. Here we are using the fact that $((x-y)/2)$ is distributed as $\mathcal{N}(0, I)$. Fixing such a triple (x, y, z), we conclude that

$$\begin{aligned} g(p+q) &= f\Big(p - \frac{x-z}{2}\Big) + f\Big(q - \frac{z-y}{2}\Big) + f\Big(\frac{x-y}{2}\Big) \\ &= g(p) + g(q) + f\Big(\frac{x-y}{2}\Big) - f\Big(\frac{x-z}{2}\Big) - f\Big(\frac{z-y}{2}\Big) \\ &= g(p) + g(q). \end{aligned}$$

Therefore g is additive within $\mathrm{B}(0, 1/r)$. ◀

Finally, we argue that g is additive everywhere. Intuitively this should be true because the values of g on points outside of $\mathrm{B}(0, 1/r)$ are defined by extrapolating the values of g on points within $\mathrm{B}(0, 1/r)$, where we know g is additive. For the proof, it will be useful to record the following fact.

▶ Fact 14. Provided that (1) – (3) of Lemma 9 hold then, for every $p \in \mathbb{R}^n$ with $\|p\|_2 \leq 1/r$ and $c \in \mathbb{Z}^{>0}$, we have $g(p) = cg(p/c)$.

Proof. Observe that $g(p) = g((c/c)p) = g\left(\sum_{i=1}^{c} p/c\right) = \sum_{i=1}^{c} g(p/c) = c \cdot g(p/c)$, where the third equality follows by Lemma 11, noting that $\|kp/c\|_2 \leq 1/r$ for every $k \in [c-1]$ ◀

Proof of Lemma 8. Fix a pair of points $p, q \in \mathbb{R}^n$, we will argue that $g(p+q) = g(p) + g(q)$. Recall that $g(p) := k_p g(p/k_p)$, $g(q) := k_q g(q/k_q)$, and $g(p+q) := k_{p+q} g((p+q)/k_{p+q})$. Then,

$$g(p) + g(q) = k_p \cdot g\left(\frac{p}{k_p}\right) + k_q \cdot g\left(\frac{p}{k_q}\right) = k_p k_q k_{p+q} \cdot g\left(\frac{p}{k_p k_q k_{p+q}}\right) + k_p k_q k_{p+q} \cdot g\left(\frac{p}{k_p k_q k_{p+q}}\right),$$

where the second equality follows by Fact 14, noting that $k_p, k_q, k_{p+q} \in \mathbb{Z}^{>0}$ and so $p/k_p, q/k_q \in \mathrm{B}(0, 1/r)$. Furthermore, because $p/(k_p k_q k_{p+q}), q/(k_p k_q k_{p+q}), (p+q)/(k_p k_q k_{p+q}) \in \mathrm{B}(0, 1/r)$, we can apply Lemma 11 to obtain

$$\begin{aligned} k_p k_q k_{p+q} \left(g\left(\frac{p}{k_p k_q k_{p+q}}\right) + g\left(\frac{p}{k_p k_q k_{p+q}}\right)\right) &= k_p k_q k_{p+q} \cdot g\left(\frac{p+q}{k_p k_q k_{p+q}}\right) \\ &= k_{p+q} \cdot g\left(\frac{p+q}{k_{p+q}}\right) \\ &= g(p+q), \end{aligned}$$

where the second equality follows by Fact 14, noting that $k_p k_q \in \mathbb{Z}^{>0}$ and $(p+q)/k_{p+q} \in \mathrm{B}(0, 1/r)$.

Finally, by Lemma 12, g is well-defined within $\mathrm{B}(0, 1/r)$. Because g is defined by extrapolating from its value within this ball, it is well-defined everywhere. ◀

▶ Remark 15. This tester (and the same proof) will in fact work over any Gaussian $\mathcal{N}(0, \Sigma)$ for arbitrary covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ by setting the value of r to be $50\sqrt{\|\Sigma^{-1}\|_2}$.

Algorithm 3: Distribution-Free Additivity Tester

Given : query access to $f : \mathbb{R}^n \to \mathbb{R}$, sampling access to an unknown distribution $\mathcal{D}$, and sampling access to $\mathcal{N}(0, I)$;

1 **Reject** if TESTADDITIVITY(f) returns **Reject**;
2 **for** $N_3 := O(1/\varepsilon)$ *times* **do**
3   Sample $p \sim \mathcal{D}$;
4   **Reject** if $f(p) \neq$ QUERY-$g(p, f)$ or if QUERY-$g(p, f)$ returns **Reject**.
5 **Accept**.

3.2 Distribution-Free Tester

In this section, we prove Theorem 1 by adapting our tester for additivity over the standard Gaussian (Algorithm 1) to a distribution-free tester.

Assuming that we are able to draw samples from the standard Gaussian (or in fact any Gaussian), the modification to Algorithm 1 is straight forward. Indeed, we will only have to modify Algorithm 1, the two subroutines will remain the same. Let $\mathcal{D}$ be our unknown distribution by which we will measure the distance of f to an additive function. The high-level idea is to first run the TESTADDITIVITY subroutine over the standard Gaussian. If it passes, then we know that with high probability g is additive. We can obtain query access to $g(p)$ (with high probability) as before by sampling points $x \sim \mathcal{N}(0, I)$ and checking that the values of $k_p(f(p/k_p - x) + f(x))$ agree for all of the x that we sample. To test whether f and g are ε-far according to $\mathcal{D}$ it suffices to sample points $p \sim \mathcal{D}$ and check whether $f(p)$ and $g(p)$ agree.

Our algorithm is given in Algorithm 3. We stress that both subroutines TESTADDITIVITY and QUERY-$g(p_i)$ are being performed over $\mathcal{N}(0, I)$, i.e., they do not use $\mathcal{D}$.

Proof of Theorem 1. The proof is nearly identical to the proof of Theorem 7. Again, observe that if f is an additive function then Algorithm 3 always accepts.

It remain to show that if f is ε-far from additive functions, then Algorithm 3 rejects with probability at least $2/3$. If TESTADDITIVITY(f) accepts with probability at most $1/10$, we can reject f with probability at least $1 - 1/10 > 2/3$. Hence, we assume that TESTADDITIVITY(f) accepts with probability at least $1/10$. By Lemma 8, the function g is additive and hence f is ε-far from g. Note that the probability that QUERY-$g(p, f)$ fails to correctly recover $g(p)$ is at most $\varepsilon/2$ by the same argument as before. It remains to bound the probability that Step 3 fails to reject, which is

$$\left(\Pr_{p \sim \mathcal{N}(0,I)} [f(p) = g(p) \vee \text{QUERY-}g(p) \text{ fails to correctly recover } g(p)] \right)^{N_3} < \left(1 - \frac{\varepsilon}{2}\right)^{N_3} < \frac{1}{10}$$

by choosing the hidden constant in N_3 to be large enough, by the same argument as before. Therefore, Algorithm 3 rejects with probability at least $1 - 1/10 > 2/3$. ◀

4 Testing Linearity of Continuous Functions

In this section, we prove Theorem 2 by adapting the tester from the previous section (Algorithm 3) to test whether f is linear, given that f is a continuous function.

We would like to argue that if f is continuous and Algorithm 3 passes then g is in fact a linear function with high probability. However, in order to exploit continuity, we need f to satisfy $f(-x) = -f(x)$ for every $x \in \mathbb{R}^n$. First, we will show how to argue that g is linear assuming that $f(-x) = -f(x)$. After that, we will handle the case when this property does not hold.

▶ **Lemma 16.** *If $f\colon \mathbb{R}^n \to \mathbb{R}$ is a continuous function satisfying $f(-x) = -f(x)$ and the assumptions of Lemma 8 hold, then the function g is linear.*

The proof will rely on the following claim which was originally proved by Darboux in 1875.

▷ Claim 17. Any additive function $f\colon \mathbb{R}^n \to \mathbb{R}$ which is continuous at a point $x_0 \in \mathbb{R}^n$ is a linear function.

Proof. First, it is well-known that any additive function which is continuous at a point is continuous everywhere (see e.g., [4]). Next, we argue that the continuity of f implies that $f(rx) = rf(x)$ for every $r \in \mathbb{R}$ and $x \in \mathbb{R}^n$. Because f is additive, this homogeneity holds for every $r \in \mathbb{Q}$, so it suffices to assume that r is irrational.

Fix $x \in \mathbb{R}^n$ and irrational r. Then for any $\zeta > 0$, we can always find $\tilde{r} \in \mathbb{Q}$ such that $|\tilde{r} - r| < \zeta$ and $\|\tilde{r}x - rx\|_2 < \zeta$. Now, by the continuity of f, for any $\xi > 0$ there exists $\zeta > 0$ such that whenever $\|\tilde{r}x - rx\|_2 < \zeta$, we have $|f(\tilde{r}x) - f(rx)| < \xi$. Now, take a sequence $\{\xi_i\}_i$ with $\xi_i \to 0$ and consider the corresponding sequence $\{\zeta_i\}_i$ with $\zeta_i \to 0$. Let $\{\tilde{r}_i\}_i$ with $r_i \in \mathbb{Q}$ be the sequence of approximations such that $|\tilde{r}_i - r| \le \zeta_i$ and $\|\tilde{r}_i x - rx\|_2 \le \zeta_i$. Then,

$$|f(rx) - rf(x)| \le |f(rx) - f(\tilde{r}_i x)| + |f(\tilde{r}_i x) - rf(x)| \le \xi_i + |\tilde{r}_i f(x) - rf(x)| \le \xi_i + \zeta_i |f(x)|.$$

Because $\zeta_i, \xi_i \to 0$, $|f(rx) - rf(x)| \to 0$ and so $f(rx) = rf(x)$. ◁

With this claim in hand, we are ready to prove Lemma 16.

Proof of Lemma 16. Let f be a continuous function satisfying $f(-x) = -f(x)$. By Lemma 8, the function g is additive. Conditioned on this event, we will show that the continuity of f implies that g is linear as well. To do so, we will argue that g is continuous at the origin and then appeal to Claim 17 to conclude that g is linear.

Let B be a ball of mass $1/2$ (with respect to $\mathcal{N}(0, I)$) centred at the origin. Let $\{p_i\}_i$ be any sequence of points with $p_i \in B$, $\|p_i\|_2 \le 1/r$ and $p_i \to 0$. Now, let $\{x_i\}_i$ be a sequence of points such that $g(p_i) = f(p_i - x_i) + f(x_i)$ and $x_i \in B$. Such a sequence exists because, by Lemma 8 $\Pr_{x \sim \mathcal{N}(0,I)}[g(x) = f(p_i - x) + f(x)] \ge 1/2$ and so for every p_i there must exist such an x_i in B.

Let S be the ball centred at the origin with twice the radius of B. As S is compact and f is continuous, f is uniformly continuous on S. Thus for every $\xi > 0$, there exists $\zeta > 0$ such that $|f(p_i - x_i) - f(-x_i)| = |f(p_i - x_i) + f(x_i)| < \xi$ whenever $\|(p_i - x_i) + x_i\|_2 < \zeta$. Now, take a sequence $\{\xi_i\}_i$ with $\xi_i \to 0$ and consider the corresponding sequence $\{\zeta_i\}_i$. As $p_i \to 0$, for every i, there exists j such that $\|(p_j - x_j) + x_j\|_2 < \zeta_i$ which in particular implies that $|g(p_j)| = |f(p_j - x_j) + f(x_j)| < \xi_i$. Thus, $g(p_i) \to 0$, and g is continuous at the origin. By Claim 17, we can conclude that g is a linear function. ◀

Now we consider the case when $f(-x) \ne -f(x)$ for some x. Luckily, in this case we can *force* f to satisfy $f(-x) = -f(x)$. To do so, we test whether f is $\varepsilon/2$-far from satisfying this property. If it is, then we reject f, otherwise, we can replace f with a function f' guaranteed to satisfy this property, by defining

$$f'(x) := \frac{f(x) - f(-x)}{2}.$$

We then continue to work over f' rather than f. Our modified algorithm is given in Algorithm 4, which uses Algorithm 5 as a subroutine.

▷ Claim 18. If FORCENEGATIVITY$(f, \mathcal{D})$ accepts with probability at least $1/10$, then $\Pr_{x \sim \mathcal{D}}[f(x) = f'(x)] \ge 1 - \varepsilon$.

Algorithm 4: Distribution-Free Linearity Tester

Given : query access to a continuous $f\colon \mathbb{R}^n \to \mathbb{R}$, sampling access to an unknown distribution $\mathcal{D}$, and sampling access to $\mathcal{N}(0, I)$;

Reject if ForceNegativity($f, \mathcal{D}$) returns **Reject**;
Let f' be the returned function;
Reject if TestAdditivity(f') returns **Reject**;
for $N_4 := O(1/\epsilon)$ *times* **do**
  Sample $p \sim \mathcal{D}$;
  Reject if $f'(p) \neq$ Query-$g(f', p)$ or if Query-$g(f', p)$ returns **Reject**.
Accept.

Algorithm 5: Force Negativity Subroutine

Procedure ForceNegativity($f, \mathcal{D}$)
  Given : query-Access to $f\colon \mathbb{R}^n \to \mathbb{R}$ and sampling access to an unknown distribution $\mathcal{D}$;
  for $N_5 := O(1/\epsilon)$ *times* **do**
    Sample $x \sim \mathcal{D}$;
    Reject if $f(-x) \neq -f(x)$;
  Return a function $f'\colon \mathbb{R}^n \to \mathbb{R}$ where $f'(x) := \frac{f(x) - f(-x)}{2}$;

Proof. Suppose for contradiction that $\Pr_{x \sim \mathcal{D}}[f(x) = f'(x)] \leq 1 - \varepsilon$. Observe that for a point $x \in \mathbb{R}$, $f'(x) \neq f(x)$ iff $f(-x) \neq -f(x)$. Therefore, by choosing the hidden constant in N_5 to be large enough, the probability that all the sampled points x satisfy $f(x) = -f(x)$ is at most

$$\left(\Pr_{x \sim \mathcal{D}}[f(-x) = f(x)] \right)^{N_5} < (1 - \varepsilon)^{N_5} \leq \frac{1}{10},$$

which is a contradiction. ◁

Therefore if ForceNegativity($f, \mathcal{D}$) accepts with probability at least $1/10$, f and f' are $\varepsilon/2$-close. Furthermore, because f is continuous and f' is the sum of continuous functions, f' is continuous as well, and so we can proceed with f' in place of f.

Proof of Theorem 2. First, observe that if f is linear then $f = f'$ and Algorithm 4 always accepts.

Now, we show that if f is ε-far from linear functions, then Algorithm 4 rejects with probability at least $2/3$. If either the TestAdditivity subroutine or the ForceNegativity subroutine passes with probability at most $1/10$, we can reject f with probability at least $1 - 1/10 > 2/3$. Hence, we assume both the subroutines pass with probability at least $1/10$. Then by Lemma 18, f is $\varepsilon/2$-close to f', which means that f' is $\varepsilon/2$-far from linear. Also by Lemma 16, because f' is continuous and satisfies $f'(-x) = -f'(x)$, the function g is linear, and so f' is $\varepsilon/2$-far from g. Therefore, Algorithm 4 rejects f with probability at least $1 - 1/10 > 2/3$. ◀

5 Lower Bounds on Testing Linearity in the Sampling Model

In this section, we prove Theorem 3, that is, we show without query access, any tester requires a linear number of samples in order to test linearity and additivity over the standard Gaussian. We note that we can obtain the same lower bound for testing additivity just by replacing linearity with additivity in the proof.

By Yao's minimax principle it suffices to construct two distributions, $\mathcal{D}_{\text{yes}}$ over linear functions and $\mathcal{D}_{\text{no}}$ over functions which are (with probability 1) $1/3$-far from linear such that any deterministic n-sample algorithm cannot distinguish between them with probability at least $2/3$. Let $\delta \in \mathbb{R}^{\geq 0}$ be some parameter to be set later; we will think of δ as tiny. Instances from these two distributions are generated as follows:

- $\mathcal{D}_{\text{yes}}$: Sample $w \sim \mathcal{N}(0, I)$ and return $f(x) := \langle w, x \rangle$.
- $\mathcal{D}_{\text{no}}$: Sample $w \sim \mathcal{N}(0, I)$ and for every $x \in \mathbb{R}^n$ sample $\varepsilon_x \sim \mathcal{N}(0, \delta)$. Return $f(x) := \langle w, x \rangle + \varepsilon_x$.

The functions in the support of $\mathcal{D}_{\text{yes}}$ are linear by definition. It remains to show that the instances in the support of $\mathcal{D}_{\text{no}}$ are far from linear.

▶ **Lemma 19.** *With probability* 1 *any* $f \sim \mathcal{D}_{\text{no}}$ *is* $1/3$*-far from linear.*

The proof of this lemma will hinge on the following claim.

▷ Claim 20. Let $f \sim \mathcal{D}_{\text{no}}$, for $x, y, z \sim \mathcal{N}(0, 1)$, $\Pr[f(\frac{x-y}{2}) \neq f(\frac{x-z}{2}) + f(\frac{z-y}{2})] = 1$.

Proof. Observe that $\Pr\left[f\left(\frac{x-y}{2}\right) = f\left(\frac{x-z}{2}\right) + f\left(\frac{z-y}{2}\right)\right] = \Pr\left[\varepsilon_{(x-y)/2} = \varepsilon_{(x-z)/2} + \varepsilon_{(z-y)/2}\right]$, where the probability is over $\varepsilon_{(x-y)/2}, \varepsilon_{(x-z)/2}, \varepsilon_{(z-y)/2} \sim \mathcal{N}(0, \delta)$. Define the random variable $z := \varepsilon_{(x-y)/2} - \varepsilon_{(x-z)/2} - \varepsilon_{(z-y)/2}$, and note that z is distributed according to $\mathcal{N}(0, 3\delta)$. Then

$$\Pr_{z \sim \mathcal{N}(0,3)}\left[\varepsilon_{(x-y)/2} = \varepsilon_{(x-z)/2} + \varepsilon_{(z-y)/2}\right] = \Pr_{z \sim \mathcal{N}(0,3)}[z = 0].$$

By standard arguments, we have $\Pr_{z \sim \mathcal{N}(0,3\delta)}[z = 0] = 0$. ◀

Proof of Lemma 19. Let f^* be the closest linear function to f. For a point $x \in \mathbb{R}^n$, say that $f(x)$ is *bad* if $f(x) \neq f(x^*)$. Construct the following matrix: the rows are labelled by every triple $(\frac{x-y}{2}, \frac{x-z}{2}, \frac{z-y}{2})$ and there are three columns. The entries at row $(\frac{x-y}{2}, \frac{x-z}{2}, \frac{z-y}{2})$ are $f(\frac{x-y}{2})$, $f(\frac{x-z}{2})$, and $f(\frac{z-y}{2})$. Note that because $x, y, z \sim \mathcal{N}(0, 1)$, the points $\frac{x-y}{2}$, $\frac{x-z}{2}$, $\frac{z-y}{2}$ are distributed according to $\mathcal{N}(0, 1)$.

Henceforth, we will measure mass in terms of probability mass over $\mathcal{N}(0, 1)$. By Claim 20, the probability that each row contains a bad entry is 1. Therefore, there must be some column for which the probability mass of the bad entries is at least $1/3$. This implies that a mass of at least $1/3$ of f must be changed to obtain f^*. Because f^* is the closest linear function to f, this implies that f is $1/3$-far from linear. ◀

Having defined our distributions over linear and far-from-linear functions, it remains to argue that no algorithm receiving n samples can distinguish between them with high probability.

Proof of Theorem 3. Let $\mathcal{D}$ be the distribution that with probability $1/2$ draws $f \sim \mathcal{D}_{\text{yes}}$ and otherwise draws $f \sim \mathcal{D}_{\text{no}}$. Let A be any deterministic algorithm which receives n samples $x_1, \ldots, x_n \sim \mathcal{N}(0, I)$. By Yao's minimax principle, it suffices to show that A cannot correctly distinguish which distribution of the distributions $\mathcal{D}_{\text{yes}}$ or $\mathcal{D}_{\text{no}}$ a given sample $f \sim \mathcal{D}$ comes from with probability at least $2/3$. That is, we would like to show that

$$\left| \Pr_{\substack{f \sim \mathcal{D}_{\text{yes}} \\ x_1, \ldots, x_n \sim \mathcal{N}(0,I)}}[A(f(x_1), \ldots, f(x_n)) = 1] - \Pr_{\substack{f \sim \mathcal{D}_{\text{no}} \\ x_1, \ldots, x_n \sim \mathcal{N}(0,I)}}[A(f(x_1), \ldots, f(x_n)) = 1] \right| \quad (4)$$

is $o(1)$. Suppose for contradiction that an algorithm A exists that with probability at least $2/3$ distinguishes these distributions.

Observe that the (4) can be bounded from above by the total variation distance between the distributions $(f^y(x_1), \ldots, f^y(x_n))$ for $f^y \sim \mathcal{D}_{\text{yes}}$, and $(f^n(x_1), \ldots, f^n(x_n))$ for $f^n \sim \mathcal{D}_{\text{no}}$, for

$x_1, \ldots, x_n \sim \mathcal{N}(0, I)$, as applying the algorithm A can only make the total variation distance smaller. By the definition of $\mathcal{D}_{\text{yes}}$ and $\mathcal{D}_{\text{no}}$, this means bounding the total variation distance between $(w_y^\top x_1, \ldots, w_y^\top x_n)$ and $(w_n^\top x_1 + \varepsilon_{x_1}, \ldots, w_n^\top x_n + \varepsilon_{x_n})$, where $w_y \sim \mathcal{D}_{\text{yes}}$ and $w_n \sim \mathcal{D}_{\text{no}}$

Now, let $X \in \mathbb{R}^n$ be the matrix whose rows are $x_1, \ldots, x_n$. Because $w_y, w_n \sim \mathcal{N}(0, I)$ and $\varepsilon_{x_i} \sim \mathcal{N}(0, \delta)$, it follows that

$$(w^\top x_1, \ldots, w^\top x_n) \sim \mathcal{N}(0, XX^\top),$$
$$(w_n^\top x_1, \ldots, w_n^\top x_n) + (\varepsilon_{x_1}, \ldots, \varepsilon_{x_n}) \sim \mathcal{N}(0, XX^\top + \delta I).$$

Therefore,

$$(4) \le \mathrm{d}_{\mathrm{TV}}(\mathcal{N}(0, XX^\top), \mathcal{N}(0, XX^\top + \delta I)).$$

To bound this distance we will appeal to Pinkser's inequality and Lemma 5. Thus, it will be useful to first record some facts about the covariance matrices of these distribution. First, we show that the rows of the matrix X are linearly independent with high probability.

▶ **Fact 21.** $\Pr_{x_1, \ldots, x_n \sim \mathcal{N}(0,1)}[\mathrm{span}(x_1, \ldots, x_n) = \mathbb{R}^n] = 1.$

It follows that the covariance matrices of these two distributions are positive definite with high probability.

▷ Claim 22. With probability 1 the matrices $XX^\top$ and $XX^\top + \delta I$ are positive definite.

Proof. That $XX^\top \succ 0$ is immediate from Fact 21, which implies that rows of X are linearly independent with probability 1. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $XX^\top$. To prove that $XX^\top + \delta I \succ 0$ note that adding δI simply adds δ to each of the eigenvalues. Thus, the eigenvalues of $XX^\top + \delta I$ are all positive. ◀

With these facts in hand we turn to bounding the total variation distance between $\mathcal{N}(0, XX^\top)$ and $\mathcal{N}(0, XX^\top + \delta I)$. Denote by $\Sigma_{\text{yes}} := XX^\top$ and $\Sigma_{\text{no}} := XX^\top + \delta I$. By Pinkser's inequality (Theorem 4) and Lemma 5,

$$\mathrm{d}_{\mathrm{TV}}\Big(\mathcal{N}(0, XX^\top), \mathcal{N}(0, XX^\top + \delta I)\Big) \le \sqrt{\frac{1}{4}\left(\log\left(\frac{\det \Sigma_{\text{yes}}}{\det \Sigma_{\text{no}}}\right) + \mathrm{tr}\left(\Sigma_{\text{yes}}^{-1}\Sigma_{\text{no}}\right) - n\right)}.$$

We will bound each of these terms separately.

Bounding the Determinant.

For simplicity of notation, we will bound the inverse of $\det(\Sigma_{\text{yes}})/\det(\Sigma_{\text{NO}})$ below. We have

$$\begin{aligned}\frac{\det \Sigma_{\text{no}}}{\det \Sigma_{\text{yes}}} &= \frac{\det(XX^\top + \delta I)}{\det(XX^\top)} \\ &= \det\left(XX^\top (XX^\top)^{-1} + \delta (XX^\top)^{-1}\right) \\ &= \det\left(I + \delta (XX^\top)^{-1}\right).\end{aligned}$$

▷ Claim 23. If A is a diagonalizable matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ then $\det(A + I) = \prod_{i=1}^n (\lambda_i + 1)$.

Applying this claim, we have $\det(I + \delta(XX^\top)^{-1}) = (\delta\lambda_1^{-1} + 1)\ldots(\delta\lambda_n^{-1} + 1)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $XX^\top$. By Claim 22 the matrix $XX^\top$ is positive definite and so $\lambda_i > 0$ for all i. Therefore, $(\delta\lambda_i^{-1} + 1) > 1$ for all i, and we can conclude that $\det \Sigma_{\text{n}}/\det \Sigma_{\text{yes}} > 1$. Thus we can upper bound $\det \Sigma_{\text{yes}}/\det \Sigma_{\text{no}}$ by 1.

Bounding the Trace.

Next, we bound

$$\begin{aligned}
\operatorname{tr}\left(\Sigma_{\text{yes}}^{-1}\Sigma_{\text{no}}\right) &= \operatorname{tr}\left(\left(XX^\top\right)^{-1}(XX^\top + \delta I)\right) \\
&= \operatorname{tr}\left(I + \delta\left(X^\top\right)^{-1}X^{-1}\right) \\
&\le \operatorname{tr}(I) + \delta \operatorname{tr}\left(\left(X^\top\right)^{-1}X^{-1}\right) \\
&= n + \delta \sum_{i,j} (X_{i,j}^{-1})^2 \\
&\le n + \delta n^2 \cdot \lambda_{\max}(X^{-1})^2,
\end{aligned}$$

where $\lambda_{\max}$ is the largest eigenvalue of X^{-1}. Noting that the eigenvalues of X^{-1} are the inverse of the eigenvalues of X, we have $\operatorname{tr}\left(\Sigma_{\text{yes}}^{-1}\Sigma_{\text{no}}\right) \le n + \delta n^2/\lambda_{\min}(X)^2$. Setting $\delta := C\lambda_{\min}(X)^2/n^2$ for some tiny $C > 0$ to be set later, we can conclude that $\operatorname{tr}\left(\Sigma_{\text{yes}}^{-1}\Sigma_{\text{no}}\right) \le n + C$.

Completing the proof.

Putting our previous bounds together we conclude that

$$\mathrm{d}_{\mathrm{TV}}\left(\mathcal{N}(0, XX^\top), \mathcal{N}(0, XX^\top + \delta I)\right) \le \sqrt{\frac{1}{4}\left(\log(1) + n + C - n\right)} = \frac{1}{2}C^{1/2}.$$

By our previous argument we have

$$(4) \le \mathrm{d}_{\mathrm{TV}}\left(\mathcal{N}(0, XX^\top), \mathcal{N}(0, XX^\top + \delta I)\right) \le \frac{1}{2}C^{1/2}.$$

Setting $C < (2/3)^2$ contradicts our assumption of the existence of an algorithm A which distinguishes a sample drawn from $\mathcal{D}_{\text{yes}}$ from one drawn from $\mathcal{D}_{\text{no}}$ with probability at least $2/3$, completing the proof. ◀

Finally, observe that the same proof goes through for testing additivity as well. Indeed, $\mathcal{D}_{\text{yes}}$ is supported on additive functions, while $\mathcal{D}_{\text{no}}$ is supported on functions which are far from additive with probability 1.

▶ **Corollary 24.** *Any sampler for additivity of functions $f\colon \mathbb{R}^n \to \mathbb{R}$ requires $\Omega(n)$ samples when $\mathcal{D} = \mathcal{N}(0, I)$.*

References

1 Sigal Ar, Manuel Blum, Bruno Codenotti, and Peter Gemmell. Checking approximate computations over the reals. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC)*, pages 786–795, 1993. URL: https://doi.org/10.1145/167088.167288, doi:10.1145/167088.167288.

2 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and hardness of approximation problems. In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 14–23, 1992. URL: https://doi.org/10.1109/SFCS.1992.267823, doi:10.1109/SFCS.1992.267823.

3 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. URL: https://doi.org/10.1109/FOCS.2012.64, doi:10.1109/FOCS.2012.64.

4 Robert Gardner Bartle and Donald R Sherbert. *Introduction to real analysis*. Hoboken, NJ: Wiley, 2011.

5 Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, pcps and non-approximability - towards tight results. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 422–431, 1995. URL: `https://doi.org/10.1109/SFCS.1995.492573`, `doi:10.1109/SFCS.1995.492573`.

6 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. URL: `https://doi.org/10.1007/978-3-030-19955-5_5`, `doi:10.1007/978-3-030-19955-5\_5`.

7 Michael Ben Or, Don Coppersmith, Mike Luby, and Ronitt Rubinfeld. Non-abelian homomorphism testing, and distributions close to their self-convolutions. *Random Structures & Algorithms*, 32(1):49–70, January 2008.

8 Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal Testing of Reed-Muller Codes. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 488–497, 2010.

9 Eric Blais. Testing juntas nearly optimally. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 151–158, 2009.

10 Eric Blais, Amit Weinstein, and Yuichi Yoshida. Partially symmetric functions are efficiently isomorphism testable. *SIAM Journal on Computing*, 44(2):411–432, 2015.

11 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. URL: `https://doi.org/10.1016/0022-0000(93)90044-W`, `doi:10.1016/0022-0000(93)90044-W`.

12 Nader H. Bshouty. Almost optimal distribution-free junta testing. In *Proceedings of the 34th Computational Complexity Conference (CCC)*, pages 2:1–2:13, 2019. URL: `https://doi.org/10.4230/LIPIcs.CCC.2019.2`, `doi:10.4230/LIPIcs.CCC.2019.2`.

13 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. URL: `https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2017.37`, `doi:10.4230/LIPIcs.APPROX-RANDOM.2017.37`.

14 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. URL: `https://doi.org/10.1137/1.9781611974331.ch5`, `doi:10.1137/1.9781611974331.ch5`.

15 Anindya De, Elchanan Mossel, and Joe Neeman. Is your function low-dimensional? *arXiv e-prints*, page arXiv:1806.10057, 2018. `arXiv:1806.10057`.

16 Elya Dolev and Dana Ron. Distribution-free testing for monomials with a sublinear number of queries. *Theory of Computing*, 7(1):155–176, 2011. URL: `https://doi.org/10.4086/toc.2011.v007a011`, `doi:10.4086/toc.2011.v007a011`.

17 Funda Ergün, Ravi Kumar, and Ronitt Rubinfeld. Checking approximate computations of polynomials and functional equations. *SIAM Journal on Computing*, 31(2):550–576, 2001. URL: `https://doi.org/10.1137/S0097539798337613`, `doi:10.1137/S0097539798337613`.

18 Eldar Fischer, Guy Kindler, Dana Ron, Shmuel Safra, and Alex Samorodnitsky. Testing juntas. *Journal of Computer and System Sciences*, 68(4):753–787, 2004.

19 Peter Gemmell, Richard J. 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. URL: `https://doi.org/10.1145/103418.103429`, `doi:10.1145/103418.103429`.

20 Dana Glasner and Rocco A. Servedio. Distribution-free testing lower bounds for basic boolean functions. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 494–508, 2007. URL: `https://doi.org/10.1007/978-3-540-74208-1_36`, `doi:10.1007/978-3-540-74208-1\_36`.

21 Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. URL: http://www.cambridge.org/us/catalogue/catalogue.asp?isbn=9781107194052, doi:10.1017/9781108135252.

22 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. URL: https://doi.org/10.1145/285055.285060, doi:10.1145/285055.285060.

23 Shirley Halevy and Eyal Kushilevitz. Distribution-free property-testing. *SIAM Journal on Computing*, 37(4):1107–1138, 2007. URL: https://doi.org/10.1137/050645804, doi:10.1137/050645804.

24 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. URL: https://doi.org/10.1137/1.9781611975482.44, doi:10.1137/1.9781611975482.44.

25 Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 627–636, 1996. URL: https://doi.org/10.1109/SFCS.1996.548522, doi:10.1109/SFCS.1996.548522.

26 Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM Journal on Computing*, 36(3):779–802, 2006. URL: https://doi.org/10.1137/S0097539704445615, doi:10.1137/S0097539704445615.

27 Marcos A. Kiwi, Frédéric Magniez, and Miklos Santha. Approximate testing with relative error. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC)*, pages 51–60, 1999. URL: https://doi.org/10.1145/301250.301269, doi:10.1145/301250.301269.

28 Marcos A. Kiwi, Frédéric Magniez, and Miklos Santha. Exact and approximate testing/correcting of algebraic functions: A survey. In *Theoretical Aspects of Computer Science, Advanced Lectures (First Summer School on Theoretical Aspects of Computer Science, Tehran, Iran, July 2000)*, pages 30–83, 2000. URL: https://doi.org/10.1007/3-540-45878-6_2, doi:10.1007/3-540-45878-6_2.

29 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. URL: https://doi.org/10.1137/1.9781611973402.89, doi:10.1137/1.9781611973402.89.

30 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. URL: https://doi.org/10.1145/3264434, doi:10.1145/3264434.

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

32 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. 2010. URL: https://doi.org/10.1007/978-3-642-16367-8_27, doi:10.1007/978-3-642-16367-8_27.

33 Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A Servedio. Testing$\pm$1-weight halfspace. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM)*, pages 646–657. 2009.

34 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. URL: https://doi.org/10.1145/2591796.2591807, doi:10.1145/2591796.2591807.

35 Kenta Oono and Yuichi Yoshida. Testing properties of functions on finite groups. *Random Structures & Algorithms*, 49(3):579–598, 2016.

36 Sofya Raskhodnikova and Ronitt Rubinfeld. Linearity and group homomorphism testing/testing hadamard codes. *Encyclopedia of Algorithms*, pages 1–6, 2014.