Passive and Active Testing of Linear Functions Over the Boolean Hypercube[⋆]

Abbas Mehrabian

Department of Combinatorics and Optimization, University of Waterloo,
Waterloo, Ontario, Canada
amehrabi@uwaterloo.ca

Abstract. One of the motivations for property testing is the idea that testing can provide a fast preprocessing step before learning. In the *standard* property testing, the algorithm is able to query the objective function on any point of its choice, whereas in most machine learning applications, the data points are not chosen by the algorithm, but appear randomly from some distribution. In the *passive* property testing model, the data points appear randomly from some distribution, and are automatically labelled. Recently Balcan et al. [1] introduced the more realistic *active property testing model*, in which the algorithm receives a sample of reasonable size of points, and then asks for the label of sample points of its choice. We study the problems of active and passive property testing of linear functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$ under the uniform distribution. We show that the query complexity for passive testing is $\Omega(n)$ and $O(n + \varepsilon^{-1})$, and the query complexity for active testing is $\Omega(n/\log n)$ and $O(n/(\varepsilon \log n))$, where $\varepsilon > 0$ is the error parameter.

Keywords: active property testing, linear functions, boolean hypercube

1 Introduction

One of the motivations for property testing is the idea that testing can provide a fast preprocessing step before learning. For instance, if one wants to learn a linear threshold function that approximates some unknown function f, one can quickly test whether f is 'close' or 'far' from being a linear threshold function, and then either proceed to the learning step (with confidence) or extend its set of hypotheses. Refer to Ron [6] for a comprehensive survey on property testing results that could be useful in machine learning applications.

In the *standard* property testing, the algorithm is able to query the unknown function on any point of its choice. However, in most machine learning applications, data points are not chosen by the algorithm, but appear randomly from some distribution. Hence the standard property testing model is not very realistic.

In the *passive* property testing model, the data points appear randomly from some distribution, which makes the model more realistic. However, in this model all

[⋆] This is a 'Type 2. Ongoing Research Work' submission to ICALP 2013 Satellite Workshop on Learning Theory and Complexity data points are automatically labelled, i.e., it is assumed that seeing an unlabelled data point is as costly as seeing a labelled data point. In many applications, sampling from the domain is not very expensive, but labelling a sample point is expensive. For instance, in the problem of machine learning for medical diagnosis, it is cheap to sample a patient from the set of all patients, but running a medical test on a patient is much more expensive. The *active* property testing model, introduced very recently by Balcan, Blais, Blum, and Yang [1], was developed to capture such a setting. In this model, which is motivated by the notion of *active learning*, the algorithm receives a sample of reasonable size of points, and then asks for the label of sample points of its choice.

One of the first problems studied in (standard) property testing was testing linear functions. In this manuscript, we study the problems of active and passive property testing of linear functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$. To the best of our knowledge, no results regarding these problems have appeared in the literature. Our main results are the following two theorems.

Theorem 1. *The query complexity for passive testing of linear functions from* $\mathbb{Z}_2^n$ *to* $\mathbb{Z}_2$ *under the uniform distribution is* $\Omega(n)$ *and* $O(n + \varepsilon^{-1})$*, where* $\varepsilon > 0$ *is the error parameter.*

Theorem 2. *The query complexity for active testing of linear functions from* $\mathbb{Z}_2^n$ *to* $\mathbb{Z}_2$ *under the uniform distribution is* $\Omega(n/\log n)$ *and* $O(n/(\varepsilon \log n))$*, where* $\varepsilon > 0$ *is the error parameter.*

In Section 2 we formally define the testing models. In Section 3 we review known results on standard property testing of linear functions. In Sections 4 and 5 we prove Theorems 1 and 2, respectively. Finally, in Section 6 we mention possible extensions of results and potentials for future work.

2 The Models

Let A be a domain. For a probability distribution D over A, we write $X \overset{d}{\sim} D$ if X is a random variable distributed as D. For two functions $f, g : A \to B$ and a distribution D over A, let

$$\mathrm{dist}_D(f, g) := \mathbb{P}_{X \overset{d}{\sim} D}\left[f(X) \neq g(X)\right] .$$

Note that the distance is always in $[0, 1]$. For the rest of this manuscript, we assume that D is the uniform distribution, and so we drop the subscript D and simply write $\mathrm{dist}(f, g)$. Let P be a *property* of all functions from A to B; that is, P defines a subset of all such functions. For a function $f : A \to B$, let

$$\mathrm{dist}(f, P) := \inf\{\mathrm{dist}(f, g) : g \in P\} .$$

One can think of $\mathrm{dist}(f, P)$ as the minimum number of points in which the value of f must be changed so that a function in P is obtained.

In the following definitions, A is a domain, B is a range, and P is a subset of all B-valued functions over A.

Definition 1 (Standard Property Testing [7, 4]). *A* standard property tester *with query complexity q for property P is a randomized decision algorithm that given a distance parameter $\varepsilon > 0$ and query access to an unknown function $f : A \to B$, asks the value of f on q points, and satisfies the following.*

- *If $f \in P$, then accepts with probability at least $2/3$.*
- *If $\operatorname{dist}(f, P) > \varepsilon$, then rejects with probability at least $2/3$.*

Definition 2 (Passive Property Testing [4, 5]). *A* passive property tester *with query complexity s for property P is a randomized decision algorithm that is given a distance parameter $\varepsilon > 0$ and the values of f on s sample points chosen independently and uniformly at random from A, and satisfies the following, where the probabilities are taken over the sample points and the internal coin flips of the algorithm.*

- *If $f \in P$, then accepts with probability at least $2/3$.*
- *If $\operatorname{dist}(f, P) > \varepsilon$, then rejects with probability at least $2/3$.*

Definition 3 (Active Property Testing [1]). *An* active property tester *with sample complexity s and query complexity q for property P is a randomized decision algorithm that given a distance parameter $\varepsilon > 0$ and s sample points chosen independently and uniformly at random from A, asks the value of f on q of the sample points, and satisfies the following, where the probabilities are taken over the sample points and the internal coin flips of the algorithm.*

- *If $f \in P$, then accepts with probability at least $2/3$.*
- *If $\operatorname{dist}(f, P) > \varepsilon$, then rejects with probability at least $2/3$.*

Remark 1. In general we have

$$\begin{aligned}\text{query complexity in standard model} &\le \text{query complexity in active model}\\ &\le \text{query complexity in passive model}\,.\end{aligned}$$

Remark 2. An active tester whose sample complexity equals the domain size, has almost the same query complexity as a standard tester. However, we are interested in active testers with sample complexity significantly smaller than (usually polylogarithmic in) the domain size.

Let $\mathcal{L}_n$ denote the set of all linear functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$. We denote the domain, $\mathbb{Z}_2^n$, by $\mathfrak{D}$ and denote the set $\{1, 2, \ldots, n\}$ by $[n]$. All logarithms are in natural base.

3 Standard Testing

Definition 4 (Linear function). *Let G and H be two groups. A function $f : G \to H$ is linear if for every $x, y \in G$ we have $f(x + y) = f(x) + f(y)$.*

Linearity testing was the first property testing problem studied, though the term ‘property testing’ was not used at that time. Blum, Luby, and Rubinfield [3] proved the following theorem in 1993.

Theorem 3. *The query complexity of testing linear functions is $\Theta(1/\varepsilon)$.*

Here we prove a special case of this theorem: when the domain is $\mathbb{Z}_2^n$ and the range is $\mathbb{Z}_2$. Note that in this case the VC-dimension of the set of linear functions is n, so a tester requires significantly less queries than a learner. The tester runs Algorithm 1, which is very simple. The difficulty lies in the analysis.

Algorithm 1 Standard property tester for linear functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$

```
Require: $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and $\varepsilon > 0$
1: for $i = 1$ to $2/\varepsilon$ do
2:     Uniformly and independently select $X, Y \in \mathbb{Z}_2^n$.
3:     If $f(X) + f(Y) \neq f(X + Y)$ then reject.
4: end for
5: If no iteration caused rejection, then accept.
```

Clearly the query complexity of Algorithm 1 is $\Theta(1/\varepsilon)$. To show correctness, we need to show two things:

(a) If $f \in \mathcal{L}_n$, then Algorithm 1 accepts with probability at least 2/3.
(b) If $\mathrm{dist}(f, \mathcal{L}_n) > \varepsilon$, then Algorithm 1 rejects with probability at least 2/3.

Item (a) is obvious, since any linear function passes all tests in line 3 successfully. Hence if $f \in \mathcal{L}_n$, then Algorithm 1 accepts with probability 1. Item (b) follows immediately from the following lemma.

Lemma 1. *There exists an absolute constant $\xi > 0$ such that the following is true. Let $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ be arbitrary. If we choose $X, Y \in \mathbb{Z}_2^n$ independently and uniformly at random, then we have*

$$\mathbb{P}\left[f(X) + f(Y) \neq f(X + Y)\right] \geq \xi \, \mathrm{dist}(f, \mathcal{L}_n) \, .$$

Blum et al. [3] proved Lemma 1 with $\xi = 2/9$. Three years later, Bellare et al. [2] proved using Discrete Fourier analysis that $\xi = 1$ works as well. We prove that Lemma 1 is true with $\xi = 1/7$. This follows from Lemma 4 in Section 5.

Now, let f be a function with $\mathrm{dist}(f, \mathcal{L}_n) > \varepsilon$. Then by Lemma 1 (say with $\xi = 1$), in every iteration of Algorithm 1, the algorithm rejects in line 3 with probability at least ε. Moreover, the iterations are independent. Hence, the probability that the algorithm accepts is at most $(1 - \varepsilon)^{2/\varepsilon} \leq \exp(-\varepsilon)^{2/\varepsilon} < 1/3$, and this concludes the proof of (b) and of Theorem 3.

4 Passive Testing

In this section we prove Theorem 1, starting with the upper bound. The passive tester is illustrated in Algorithm 2. Note that the query complexity of Algorithm 2 is clearly $O(n + 1/\varepsilon)$. To show correctness, we need to show two things:

Algorithm 2 Passive property tester for linear functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$

```
Require: $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and $\varepsilon > 0$
 1: $S \leftarrow 12n$ elements from $\mathbb{Z}_2^n$ chosen independently and uniformly at random
 2: if $S$ contains $n$ linearly independent elements $b_1, b_2, \ldots, b_n$ then
 3:     Define $g : \mathbb{Z}_2^n \to \mathbb{Z}_2$ as $g(x) := \sum_{i=1}^n \langle x, b_i \rangle f(b_i)$.
 4: else
 5:     accept
 6: end if
 7: for $i = 1$ to $2/\varepsilon$ do
 8:     Uniformly and independently select $Y \in \mathbb{Z}_2^n$.
 9:     If $f(Y) \neq g(Y)$ then reject.
10: end for
11: If no iteration caused rejection, then accept.
```

(a) If $f \in \mathcal{L}_n$, then Algorithm 2 accepts with probability at least 2/3.
(b) If $\mathrm{dist}(f, \mathcal{L}_n) > \varepsilon$, then Algorithm 2 rejects with probability at least 2/3.

We first prove (a). Notice that if f is a linear function, then it is uniquely determined by its effect on a linear basis. Namely, if $\{b_1, b_2, \ldots, b_n\}$ is a set of n linearly independent vectors in $\mathfrak{D}$, then

$$\forall x \in \mathfrak{D} : \qquad f(x) = \sum_{i=1}^{n} \langle x, b_i \rangle f(b_i) \; .$$

Hence if f is linear, then either Algorithm 2 accepts in line 5, or else the function g would be equal to f on all points, and thus Algorithm 2 passes all tests in line 9, and this proves (a).

Now, we turn to proving (b). Assume that $\mathrm{dist}(f, \mathcal{L}_n) > \varepsilon$. First, it is not hard to show that the probability that line 5 is reached is at most 1/6. In fact, the expected number of random vectors to choose from $\mathfrak{D}$ uniformly at random in order to span the space is at most $2n$, and the claim follows by applying Markov's inequality.

Now, since g is linear and $\mathrm{dist}(f, \mathcal{L}_n) > \varepsilon$, we have $\mathrm{dist}(f, g) > \varepsilon$. Thus every Y chosen in line 8 causes rejection with probability at least ε. As the Y's are chosen independently, the probability that the algorithm passes all tests in line 9 is at most $(1 - \varepsilon)^{2/\varepsilon} \leq \exp(-\varepsilon)^{2/\varepsilon} < 1/6$. Applying the union bound completes the proof of (b) and the upper bound in Theorem 1.

For proving the lower bound in Theorem 1 we will need a lemma.

Lemma 2. *Let $\mathcal{F}$ be the set of all functions $f : \mathfrak{D} \to \mathbb{Z}_2$. Suppose that an adversary chooses a function f either uniformly at random from $\mathcal{L}_n$, or uniformly at random from $\mathcal{F}$. For every linearly independent set $\{x_1, \ldots, x_k\} \subseteq \mathfrak{D}$, there is no algorithm that given the labelled data set $((x_1, f(x_1)), (x_2, f(x_2)), \ldots, (x_k, f(x_k)))$ can determine, with probability at least 3/5, whether f is chosen from $\mathcal{L}_n$ or $\mathcal{F}$.*

Proof. If f is chosen randomly from $\mathcal{F}$, then all the 2^k possible label vectors $(f(x_1), f(x_2), \ldots, f(x_k))$ are equally likely to show up. Hence, the probability that any particular label vector $(f(x_1), f(x_2), \ldots, f(x_k))$ comes up for the tester is 2^{-k}.

On the other hand, if f is chosen randomly from $\mathcal{L}_n$, then again all the 2^k possible label vectors $(f(x_1), f(x_2), \ldots, f(x_k))$ are equally likely to show up. The reason is that the set $\{x_1, x_2, \ldots, x_k\}$ can be extended to a linear basis $(x_1, x_2, \ldots, x_k, x_{k+1}, x_{k+2}, \ldots, x_n)$. Each prescribed $\{0,1\}$-vector $(f_1, f_2, \ldots, f_n)$ determines exactly one linear function, namely $f(z) = \sum \langle z, x_i \rangle f_i$, which takes the prescribed values on the basis. This means that there are exactly 2^{n-k} linear functions that take the value $f(x_i)$ on all x_i, for all $1 \leq i \leq k$. Since the total number of linear functions is 2^n, the probability that the particular label vector $(f(x_1), f(x_2), \ldots, f(x_k))$ comes up for the tester is $2^{n-k}/2^n = 2^{-k}$. □

Now, we show that the query complexity for passive testing is at least $n/2$, completing the proof of Theorem 1. Let $k = n/2$ and let $X_1, X_2, \ldots, X_k$ be vectors chosen independently and uniformly at random from $\mathfrak{D}$. For a given nonempty $A \subseteq \{1, 2, \ldots, k\}$, we have $\mathbb{P}\left[\sum_{i \in A} X_i = 0\right] = 2^{-n}$. Hence, by the union bound, the probability that there exists *some* nonempty $A \subseteq \{1, 2, \ldots, k\}$ satisfying $\sum_{i \in A} X_i = 0$ is at most $2^{k-n} = 2^{-n/2}$. This means that, in particular, with probability larger than 5/6, the sample points given to the tester are linearly independent. So by Lemma 2, if the tester wants to distinguish between a randomly chosen linear function and a randomly chosen function, then its error would be at least $\frac{5}{6} \times \frac{3}{5} = \frac{1}{2}$. This completes the proof, as a uniformly chosen random function is $\left(\frac{1}{4}\right)$-far from $\mathcal{L}_n$ with probability $1 - o(1)$.

5 Active Testing

In this section we prove Theorem 2. We start by giving an active tester with query complexity $O(n/\varepsilon \log n)$. The idea of the algorithm is similar to the idea for the standard tester (Algorithm 1): the tester tries to find a set $(x_1, x_2, \ldots, x_m)$ with $\sum x_i = 0$, queries the values of f on these points, and checks that $\sum f(x_i) = 0$. However, as the tester cannot choose the x_i's, it should try to find a linearly dependent set (of as small cardinality as possible) in the sample. The following lemma essentially determines how small a linearly dependent set can one expect in a sample with polynomial size. For a positive integer s, $\mathcal{U}^s$ denotes the uniform distribution over $\mathfrak{D}^s$.

Lemma 3. *Let $w, \delta > 0$ be constants, $s = n^{1+\delta}$ and $q = wn/\log n$. Also let $(X_1, X_2, \ldots, X_s) \overset{d}{\sim} \mathcal{U}^s$.*

(a) If $w > \log(2)/\delta$, then with probability approaching 1 as n goes to infinity, there is a subsequence $(Y_1, Y_2, \ldots, Y_q)$ of $(X_1, X_2, \ldots, X_s)$ with $\sum_{i=1}^{q} Y_i = 0$.

(b) If $w < \log(2)/\delta$, then with probability approaching 1 as n goes to infinity, for every $1 \leq q' \leq q$ and every subsequence $(Y_1, Y_2, \ldots, Y_{q'})$ of $(X_1, X_2, \ldots, X_s)$ we have $\sum_{i=1}^{q'} Y_i \neq 0$.

Proof. (a) Assume that $w > \log(2)/\delta$. Let $m = \binom{s}{q}$ and let $A_1, A_2, \ldots, A_m$ be all the subsets of $[s]$ of size q. Define $\{0,1\}$-valued random variables $Z_1, Z_2, \ldots, Z_m$ as follows. For $1 \leq i \leq m$, let $Z_i = 1$ if and only if $\sum_{j \in A_i} X_j = 0$. Notice that

$\sum_{i=1}^{m} Z_i > 0$ if and only if there is a subsequence $(Y_1, Y_2, \ldots, Y_q)$ of $(X_1, X_2, \ldots, X_s)$ with $\sum_{i=1}^{q} Y_i = 0$. Let $Z = \sum_{i=1}^{m} Z_i$. Hence we only need to show that $\mathbb{P}[Z = 0]$ approaches zero as n goes to infinity.

It is easy to see that $\mathbf{E}[Z_i] = \mathbb{P}[Z_i = 1] = 2^{-n}$ and thus by linearity of expectation,

$$\mathbf{E}[Z] = m2^{-n} = \binom{s}{q} 2^{-n} \geq \left(\frac{s}{q}\right)^q 2^{-n} = \exp\left(n(w\delta - \log 2 + o(1))\right) ,$$

which approaches infinity as n goes to infinity, as $w\delta > \log 2$.

Next we compute the variance of Z. It is not hard to show that for every $i \neq j$ we have $\mathbb{P}[Z_i = Z_j = 1] = 2^{-2n}$, which means that the covariance of Z_i and Z_j is zero. Consequently,

$$\mathbf{Var}[Z] = \sum_{i=1}^{m} \mathbf{Var}[Z_i] = m2^{-n}\left(1 - 2^{-n}\right) = \mathbf{E}[Z]\left(1 - 2^{-n}\right) < \mathbf{E}[Z] .$$

Finally, Chebyshev's inequality gives

$$\mathbb{P}[Z = 0] \leq \mathbb{P}[|Z - \mathbf{E}[Z]| \geq \mathbf{E}[Z]] \leq \frac{\mathbf{Var}[Z]}{\mathbf{E}[Z]^2} < \frac{1}{\mathbf{E}[Z]} ,$$

which approaches zero as n goes to infinity. Thus, with probability approaching 1 as n goes to infinity, we have $Z > 0$, and the proof is complete.

(b) We omit the proof, which uses linearity of expectation and some straightforward calculations as in part (a). □

Before proceeding to describing the tester, let us say why part (b) of Lemma 3 together with Lemma 2 imply the lower bound in Theorem 2. Consider an active tester that is given an unlabelled sample S of size n^c for some fixed c, and assume that, the query complexity is $o(n/\log n)$. Note that by Lemma 3(b), with probability larger than $5/6$ *every* subset of S of size $o(n/\log n)$ is linearly independent. Assume that the tester asks the value of f on a set $A \subseteq S$ of size $o(n/\log n)$, In particular, with probability larger than $5/6$, the set A is linearly independent. So by Lemma 2, if the tester wants to distinguish between a randomly chosen linear function and a randomly chosen function, then its error would be at least $\frac{5}{6} \times \frac{3}{5} = \frac{1}{2}$. This completes the proof, as a uniformly chosen random function is $\left(\frac{1}{4}\right)$-far from $\mathcal{L}_n$ with probability $1 - o(1)$.

Now, we describe the active tester. The tester is illustrated in Algorithm 3. The sample complexity of Algorithm 3 is $O(n^2/\varepsilon)$, i.e., polylogarithmic in the domain size, and the query complexity of Algorithm 3 is $O(n/\varepsilon \log n)$. To show correctness, we need to show two things:

(a) If $f \in \mathcal{L}_n$, then Algorithm 3 accepts with probability at least $2/3$.
(b) If $\operatorname{dist}(f, \mathcal{L}_n) > \varepsilon$, then Algorithm 3 rejects with probability at least $2/3$.

Item (a) is obvious, as any linear function passes all the tests in lines 7 and 8.

Algorithm 3 Active property tester for linear functions from $\mathbb{Z}_2^n$ to $\mathbb{Z}_2$

```
Require: $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ and $\varepsilon > 0$
 1: $m \leftarrow n/\log n$.
 2: for $i = 1$ to $16/\varepsilon$ do
 3:     Sample a set $S$ of size $n^2$ from $\mathbb{Z}_2^n$.
 4:     if there exist $x_1, x_2, \ldots, x_m \in S$ with $\sum_{i=1}^{m} x_i = 0$
        AND there exist $y_1, y_2, \ldots, y_{m-1} \in S$ with $\sum_{i=1}^{m-1} y_i = 0$ then
 5:         choose an $m$-tuple $(X_1, X_2, \ldots, X_m) \in S^m$ with $\sum_{i=1}^{m} X_i = 0$ uniformly at random.
 6:         choose an $(m-1)$-tuple $(Y_1, Y_2, \ldots, Y_{m-1}) \in S^{m-1}$ with $\sum_{i=1}^{m-1} Y_i = 0$ uniformly at random.
 7:         if $\sum_{i=1}^{m} f(X_i) \neq 0$ then reject.
 8:         if $\sum_{i=1}^{m-1} f(Y_i) \neq 0$ then reject.
 9:     end if
10: end for
11: If no iteration caused rejection, then accept.
```

To prove (b), we need a definition and a lemma. Let k be a positive integer, and let $f : \mathfrak{D} \to \mathbb{Z}_2$ be arbitrary. Define $\varepsilon_0(f) := 0$ and

$$\varepsilon_k(f) := \mathbb{P}_{(X_1, X_2, \ldots, X_{k+1}) \overset{d}{\sim} \mathcal{U}^{k+1}} \left[\sum_{i=1}^{k+1} f(X_i) \neq f\left(\sum_{i=1}^{k+1} X_i \right) \right] .$$

The following lemma is the key for proving (b).

Lemma 4. *For every $f : \mathfrak{D} \to \mathbb{Z}_2$ and every $k > 0$ we have*

$$\mathrm{dist}(f, \mathcal{L}_n) \leq 7 \max\{\varepsilon_k(f), \varepsilon_{k-1}(f)\} . \tag{1}$$

Notice that letting $k = 1$ gives $\mathrm{dist}(f, \mathcal{L}_n) \leq 7 \max\{\varepsilon_1(f), \varepsilon_0(f)\} = 7\varepsilon_1(f)$, which implies Lemma 1 with $\xi = 1/7$. Thus Lemma 4 is, in a certain sense, a generalization of Lemma 1. We remark that one cannot hope to give a generalization of Lemma 1 as is, namely one cannot prove that for every positive integer k there exists an η_k such that $\mathrm{dist}(f, \mathcal{L}_n) \leq \eta_k \varepsilon_k(f)$. To see this, let f be the constant 1 function. Then $\mathrm{dist}(f, \mathcal{L}_n) = 1/2$, however, for every even k we have $\varepsilon_k(f) = 0$.

Before proving Lemma 4, let us show that it implies item (b) above. Assume that $\mathrm{dist}(f, \mathcal{L}_n) > \varepsilon$. Then in each iteration of Algorithm 3, the condition in line 4 is true with probability at least $7/8$ by Lemma 3(a). Conditional on this, at least one of the tests in lines 7 and 8 fail with probability at least $\max\{\varepsilon_{m-2}, \varepsilon_{m-3}\} \geq \varepsilon/7$ by Lemma 4. Thus each iteration fails with probability at least $\varepsilon/7 \times 7/8 = \varepsilon/8$. Since the iterations are independent, the probability that the tester accepts is at most $(1 - \varepsilon/8)^{16/\varepsilon} \leq \exp(-2) < 1/3$.

The rest of this section is devoted to the proof of Lemma 4. Ron [6] presented a proof of Lemma 1 with $\xi = 1/6$. The proof of Lemma 4 here follows the line of that proof, except that the proof of one of the sub-lemmas, namely Lemma 7, requires an extra step. Fix a function $f : \mathfrak{D} \to \mathbb{Z}_2$ and a positive integer k, and let $\varepsilon_k = \varepsilon_k(f)$

and $\varepsilon_{k-1} = \varepsilon_{k-1}(f)$. For a point $x \in \mathfrak{D}$ and a sequence $A \in \mathfrak{D}^k$, define

$$V_A(x) := \sum_{a \in A} f(a) + f\left(x + \sum_{a \in A} a\right) .$$

Note that if f was linear then we would have $V_A(x) = f(x)$ for all A and $x \in \mathfrak{D}$. One can think of $V_A(x)$ as the 'vote that A casts' on the value of $f(x)$. Note that, indeed,

$$\varepsilon_k = \varepsilon_k(f) = \mathbb{P}_{(X,A) \overset{d}{\sim} \mathcal{U}^1 \times \mathcal{U}^k} \left[f(X) \neq V_A(X) \right] .$$

Also, define the function $g : \mathfrak{D} \to \mathbb{Z}_2$ as follows. For $x \in \mathfrak{D}$, let $g(x) = 0$ if

$$\mathbb{P}_{A \overset{d}{\sim} \mathcal{U}^k} \left[V_A(x) = 0 \right] \geq 1/2 ,$$

and let $g(x) = 1$ otherwise. Note that g is the majority vote taken over all $A \in \mathfrak{D}^k$. That is, we have $g(x) = 0$ if and only if most of the $A \in \mathfrak{D}^k$ 'believe' that $f(x)$ must be zero. Therefore, for every $x \in \mathfrak{D}$ we have $\mathbb{P}_{A \overset{d}{\sim} \mathcal{U}^k} \left[V_A(x) = g(x) \right] \geq 1/2$.

The proofs of Lemmas 5 and 6 below are very similar to the proofs of Lemma 3.3 and Claim 3.5 in Ron [6], respectively, thus we omit them.

Lemma 5. *We have* $\operatorname{dist}(f, g) \leq 2\varepsilon_k$.

Lemma 6. *For every $x \in \mathfrak{D}$ we have* $\mathbb{P}_{A \overset{d}{\sim} \mathcal{U}^k} \left[V_A(x) \neq g(x) \right] \leq 2\varepsilon_k$.

Lemma 7. *If $6\varepsilon_k + \varepsilon_{k-1} < 1$, then g is a linear function.*

Proof. Assume that $6\varepsilon_k + \varepsilon_{k-1} < 1$ and let $x, y \in \mathfrak{D}$ be arbitrary, and we show $g(x + y) = g(x) + g(y)$. For this, let $A = (T_1, T_2, \ldots, T_{k-1}, S) \overset{d}{\sim} \mathcal{U}^k$, and let $B = (T_1, T_2, \ldots, T_{k-1}, S + x)$. Note that $B \overset{d}{\sim} \mathcal{U}^k$. We show that with positive probability (over the k-tuple A) all of the following four events happen:

$$g(x) = \sum_{i=1}^{k-1} f(T_i) + f(S) + f\left(\sum_{i=1}^{k-1} T_i + S + x\right) \tag{2}$$

$$g(y) = \sum_{i=1}^{k-1} f(T_i) + f(S + x) + f\left(\sum_{i=1}^{k-1} T_i + S + x + y\right) \tag{3}$$

$$g(x + y) = \sum_{i=1}^{k-1} f(T_i) + f(S) + f\left(\sum_{i=1}^{k-1} T_i + S + x + y\right) \tag{4}$$

$$0 = \sum_{i=1}^{k-1} f(T_i) + f(S + x) + f\left(\sum_{i=1}^{k-1} T_i + S + x\right) \tag{5}$$

Equation (2) just means that $g(x) = V_A(x)$, which is wrong with probability at most $2\varepsilon_k$ by Lemma 6. Equation (3) just means that $g(y) = V_B(y)$, which is wrong with probability at most $2\varepsilon_k$ by Lemma 6. Equation (4) just means that $g(x + y) = V_A(x + y)$, which is wrong with probability at most $2\varepsilon_k$ by Lemma 6. Equation (5) is wrong with probability exactly ε_{k-1} by the definition of ε_{k-1}. As $6\varepsilon_k+\varepsilon_{k-1} < 1$, with positive probability (where the probability is taken over the k-tuple A) all equations (2)–(5) are true, hence there exist certain $(t_1, t_2, \ldots, t_{k-1}, s)$ for which (2)–(5) are true, which implies (by adding them up) that $g(x) + g(y) + g(x + y) = 0$. □

We now have all the ingredients to prove Lemma 4.

Proof (of Lemma 4). Let $f : \mathfrak{D} \to \mathbb{Z}_2$ and $k > 0$. We need to show that (1) holds. If $\max\{\varepsilon_k(f), \varepsilon_{k-1}(f)\} \geq 1/7$, then (1) obviously holds as $\mathrm{dist}(f, P)$ is bounded by one for any function f and any property P. Therefore we may assume that $\max\{\varepsilon_k(f), \varepsilon_{k-1}(f)\} < 1/7$. Hence the function g is linear by Lemma 7. But then by Lemma 5,

$$\mathrm{dist}(f, \mathcal{L}_n) \leq \mathrm{dist}(f, g) \leq 2\varepsilon_k \leq 7 \max\{\varepsilon_k(f), \varepsilon_{k-1}(f)\} ,$$

and this completes the proof. □

6 Future Work

First of all, there is a additive (respectively, multiplicative) gap of $O(1/\varepsilon)$ in Theorem 1 (respectively, Theorem 2) and closing these gaps is left as an open problem. It would be interesting to extend the upper bounds in Theorems 1 and 2 to arbitrary underlying distributions, instead of just the uniform distribution. My conjecture is that both theorems are true for any distribution. Notice that the lower bounds extend automatically.

Another possible direction is to extend the domain and range. For any two groups G and H, one can define the property of 'being linear' for functions from G to H, and indeed the standard property testing result, Theorem 3, works in this general setting. It would be interesting to extend Theorems 1 and 2 to arbitrary groups, instead of just $\mathbb{Z}_2^n$ and $\mathbb{Z}_2$. Although in property testing settings the query complexity seems to be the important parameter, it would also be interesting to investigate time complexities of Algorithms 2 and 3.

References

1. M.-F. Balcan, E. Blais, A. Blum, L. Yung. Active property testing. In Proceedings of *FOCS 2012* (available on arXiv:1111.0897).
2. M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.
3. M. Blum, M. Luby, and R. Rubinfield. Self-testing/correcting with applications to numerical problems. *Journal of the ACM*, 47:549–595, 1993.
4. O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.
5. M. Kearns and D. Ron. Testing problems with sub-learning complexity. *Journal of Computer and System Sciences*, 61(3):428–456, 2000.
6. D. Ron. Property testing: a learning theory perspective. *Foundations and trends in machine learning*, Vol. 1, No. 3 (2008), pp. 307–402.
7. R. Rubinfield and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.