Property Testing Bounds for Linear and Quadratic Functions via Parity Decision Trees

Abhishek Bhrushundi[1], Sourav Chakraborty[1], and Raghav Kulkarni[2]

[1] {abhishek_bhr,sourav}@cmi.ac.in
Chennai Mathematical Institute
India
[2] kulraghav@gmail.com
Center for Quantum Technologies
Singapore

Abstract. In this paper, we study linear and quadratic Boolean functions in the context of property testing. We do this by observing that the query complexity of testing properties of linear and quadratic functions can be characterized in terms of complexity in another model of computation called *parity decision trees*.

The observation allows us to characterize testable properties of linear functions in terms of the approximate l_1 norm of the Fourier spectrum of an associated function. It also allows us to reprove the $\Omega(k)$ lower bound for testing k-linearity due to Blais et al [8]. More interestingly, it rekindles the hope of closing the gap of $\Omega(k)$ vs $O(k \log k)$ for testing k-linearity by analyzing the randomized parity decision tree complexity of a fairly simple function called E_k that evaluates to 1 if and only if the number of 1s in the input is exactly k. The approach of Blais et al. using communication complexity is unlikely to give anything better than $\Omega(k)$ as a lower bound.

In the case of quadratic functions, we prove an adaptive two-sided $\Omega(n^2)$ lower bound for testing affine isomorphism to the inner product function. We remark that this bound is tight and furnishes an example of a function for which the trivial algorithm for testing affine isomorphism is the best possible. As a corollary, we obtain an $\Omega(n^2)$ lower bound for testing the class of *Bent* functions.

We believe that our techniques might be of independent interest and may be useful in proving other testing bounds.

1 Introduction

The field of property testing broadly deals with determining whether a given object satisfies a property $\mathcal{P}$ or is very different from all the objects that satisfy $\mathcal{P}$. In this paper, the objects of interest are Boolean functions on n variables, i.e. functions of the form

$$f : \{0,1\}^n \to \{0,1\}.$$

A Boolean function property $\mathcal{P}$ is a collection of Boolean functions. Given a function g and a parameter ϵ, the goal of a tester is to distinguish between the following two cases:

– $g \in \mathcal{P}$
– g differs from every function in $\mathcal{P}$ in at least ϵ fraction of points in $\{0,1\}^n$.

The query complexity for testing $\mathcal{P}$ is the number of queries (of the form "what is the value of g at $x \in \{0,1\}^n$?") made by the best tester that distinguishes between the above two cases. If the queries made by the tester depend on the answers to the previous queries, the tester is called *adaptive*. Also, if the tester accepts whenever $g \in \mathcal{P}$, it is called *one-sided*.

Testing of Boolean function properties has been extensively studied over the last couple of decades (See [16,28]). Examples of problems that have been studied are linearity testing [11], k-junta testing [17,7], monotonicity testing [18,13], k-linearity testing [19,8,12] etc. An important problem in the area is to characterize Boolean function properties whose query complexity is *constant* (i.e., independent of n, though it can depend on ϵ). For example, such a characterization is known in the case of graph properties [1]. Though a general characterization for function properties is not yet known, there has been progress for some special classes of properties. In this paper, we attempt characterizing one such class: properties which only consist of linear functions. More specifically, we try to characterize all properties $\mathcal{P}$ of linear Boolean functions which can be tested using constant number of queries.

An example of a property of linear functions is one that contains all parities on k variables. The problem of testing this property is known as k-linearity testing. While this problem had been studied earlier [19], recently Blais et al. [8] used communication complexity to obtain a lower bound of $\Omega(k)$ on the query complexity of adaptive testers for k-linearity. The best known upper bound in the case of adaptive testers is $O(k \log k)$. Whereas a tight bound of $\Theta(k \log k)$ is known for the non-adaptive case [12], a gap still exists for adaptive testing: $\Omega(k)$ vs $O(k \log k)$. In this paper we give another approach to obtain the $\Omega(k)$ lower bound for the adaptive query complexity. While the lower bound technique of Blais et al.[8] is unlikely to give a bound beyond $\Omega(k)$, our technique has the potential of proving a better lower bound. We remark that other proof techniques for the lower bound have also been studied [9].

A rich class of properties for which characterizing constant query testability has been studied are properties that are invariant under natural transformations of the domain. For example, [22,4,3] study invariance under affine/linear transformations in this context. Properties that consist of functions isomorphic to a given function also form an important subclass. The testing of such properties is commonly referred to as *isomorphism testing*, and has seen two directions of study: testing if a function is equivalent to a given function up to permutation of coordinates [14,10], and testing affine/linear isomorphism.

Our second result[1] concerns testing affine isomorphism. A function f is affine isomorphic to g if there is an invertible affine transformation T such that $f \circ T = g$. Recently, Wimmer and Yoshida [30] characterized the query complexity of testing affine/linear isomorphism to a function in terms of its *Fourier norm*. We

[1] This result appears in the preprint [5]

complement their work by providing the first example of a function for which the query complexity of testing affine isomorphism is the largest possible. As a corollary, we also prove an adaptive two-sided $\Omega(n^2)$ lower bound for testing the class of *Bent* functions which are an important and well-studied class of Boolean functions in cryptography (See [24,25]).
Grigorescu et al. concurrently and independently obtained these results in [20] using a different proof technique. In fact, they prove an $2^{\Omega(n)}$ lower bound for testing Bent functions. We believe that our proof is arguably simpler and more modular, and is also amenable to generalizations (for example, to the quantum setting), even though the bound we obtain for Bent functions is weaker.

The main technique used in proving all our results is a connection between testing properties of linear and quadratic functions, and parity decision trees. Connections between linear functions and parity decision trees have been both implicitly [9] and explicitly [12] observed in earlier papers. Another connection that we exploit for proving some of our results is the one between parity decision tree depth and communication complexity. Similar connections were known earlier, see for example [31]. We remark that, to the best of our knowledge, our result is the first that combines the two connections, giving yet another way of relating property testing lower bounds to communication complexity (Blais et al. [8] observe such a connection in much more generality). Thus, we believe that our techniques might be of independent interest.

1.1 Our results and techniques

Property testing and parity decision trees We give a connection between testing properties of linear functions and parity decision trees. The following is an informal statement of the connection:

Theorem 1. *For every property $\mathcal{P}$ of linear functions on n variables, one can associate a Boolean function $E_{\mathcal{P}}$ on n variables such that there is an adaptive q-query tester for determining if a given f is in $\mathcal{P}$ or $1/2$-far from $\mathcal{P}$ only if there is a randomized parity decision tree that makes q queries to compute $E_{\mathcal{P}}$.*

A similar connection holds in the case of quadratic functions.

Theorem 2. *For every property $\mathcal{P}$ of quadratic functions on n variables, one can associate a Boolean function $E_{\mathcal{P}}$ on n^2 variables such that there is an adaptive q query tester for determining if a given f is in $\mathcal{P}$ or $1/4$-far from $\mathcal{P}$ only if there is a randomized parity decision tree that makes q queries to compute $E_{\mathcal{P}}$.*

All the results that follow use the above connections crucially. Another important ingredient for some of the results is a connection between parity decision trees and the communication complexity of XOR functions. We discuss this in detail in Section 3.

Characterization of testable properties of linear functions Theorem 1 allows us to characterize the constant query testability of a property $\mathcal{P}$ of linear functions in terms of the approximate L_1 norm of $E_{\mathcal{P}}$.

Theorem 3. *A property $\mathcal{P}$ of linear functions is constant query testable if and only if $||\widehat{E_{\mathcal{P}}}||_1^{1/4}$ is constant.*

This is the first such characterization of linear function properties and we hope our result is a small step towards our understanding of function properties testable in constant number of queries.

Testing k-linearity We also obtain an alternate proof of the lower bound for testing k-linearity due to Blais et al. [8].

Theorem 4. *Any adaptive two-sided tester for testing k-linearity requires $\Omega(k)$ queries.*

The idea behind the proof is as follows. Applying Theorem 1 in the case of k-linearity, $E_{\mathcal{P}}$ turns out to be equal to the function E_k that outputs 1 if and only if there are exactly k 1s in the input string. Thus, to prove Theorem 4 we lower bound the randomized parity decision tree complexity of E_k by $\Omega(k)$.
Note that this leaves open the possibility of proving a tight $\Omega(k \log k)$ bound for testing k-linearity by improving our lower bound on the randomized parity decision tree complexity of E_k.

Lower bound for testing affine isomorphism Let $\mathbf{IP}_n(x)$ denote the inner product function $\sum_{i=1}^{n/2} x_i x_{n/2+i}$. We consider the problem of testing affine isomorphism to $\mathbf{IP}_n(x)$ and prove a tight lower bound.

Theorem 5. *Any adaptive two-sided tester for testing affine isomorphism to $\mathbf{IP}_n(x)$ requires $\Omega(n^2)$ queries.*

The proof of Theorem 5 is similar to that of Theorem 4, though in this case, $E_{\mathcal{P}}$ turns out to be E_n, a function that maps graphs on n vertices to $\{0,1\}$, and outputs 1 if and only if the input graph's adjacency matrix is nonsingular over $\mathbb{F}_2$. As mentioned before, this is the first example of a function for which testing affine isomorphism requires $\Omega(n^2)$ queries ($O(n^2)$ is a trivial upper bound for any function and follows from a folklore result).

It can be show that testing the set of quadratic Bent functions reduces to testing affine isomorphism to $\mathbf{IP}_n(x)$. Thus, Theorem 5 gives a lower bound for testing the set of quadratic Bent functions. Furthermore, using a result from [15], the following corollary can be obtained.

Corollary 1. *Any adaptive two-sided tester for testing the set of Bent functions requires $\Omega(n^2)$ queries.*

2 Preliminaries

2.1 Boolean functions

Recall that functions mapping $\{0,1\}^n$ to $\{0,1\}$ are called Boolean functions. A Boolean function is linear if it is expressible as $\sum_{i \in S} x_i$ for $S \subseteq [n]$ over $\mathbb{F}_2$. The set of linear functions will be denoted by $\mathcal{L}$.

A Boolean function is quadratic if it can be expressed as a polynomial of degree at most two over $\mathbb{F}_2$. We shall denote the set of quadratic functions by $\mathcal{Q}$, and the set of homogeneous quadratic functions (no linear terms) by $\mathcal{Q}_0$. By a property of linear or quadratic functions, we shall always mean a subset of $\mathcal{L}$ or $\mathcal{Q}$.

For Boolean functions f and g, $dist(f,g) = \Pr_x[f(x) \neq g(x)]$. The notion can be extended to sets of Boolean functions S and T in a natural way: $dist(S,T) = \min_{f\in S, g\in T} dist(f,g)$. We state a simple but useful observation:

Observation 1 *If f and g are linear (quadratic) functions then either $f = g$ or $dist(f,g) \geq 1/2$ ($dist(f,g) \geq 1/4$).*

2.2 Property testing

Let $\mathcal{P}$ be a property of Boolean functions on n variables. We say a randomized algorithm $\mathcal{A}$ ϵ-tests $\mathcal{P}$, if given oracle access to the truth table of an input function f, $\mathcal{A}$ determines with probability at least $2/3$ whether $f \in \mathcal{P}$, or $dist(f,\mathcal{P}) \geq \epsilon$. The number of queries made by the best tester for ϵ-testing $\mathcal{P}$ is known as the query complexity of $\mathcal{P}$. It is denoted by $Q^\epsilon(\mathcal{P})$ and may be a function of n.

Remark When testing properties of linear functions, it is common to assume that the input function is promised to be a linear function. For a property $\mathcal{P}$ of linear functions, we denote the query complexity of testing $\mathcal{P}$ under such a promise by $Q_1(\mathcal{P})$.

For technical reasons, it will be useful to consider such a notion for quadratic functions. For a property $\mathcal{P} \subseteq \mathcal{Q}$ of quadratic functions, we shall denote by $Q_2(\mathcal{P})$ the query complexity of testing $\mathcal{P}$ under the promise that the input is always a function in $\mathcal{Q}_0$. Observation 1 implies the following statement.

Observation 2 *Let $\mathcal{P}$ be a property of linear functions. Then, $Q^{1/2}(\mathcal{P}) \geq Q_1(\mathcal{P})$. Similarly, in the case of quadratic functions, $Q^{1/4}(\mathcal{P}) \geq Q_2(\mathcal{P})$*

It can also be shown that:

Observation 3 *If $Q_1(\mathcal{P}) = Q$ then $\forall \epsilon \in (0, 1/4)$, $Q^\epsilon(\mathcal{P}) \leq O_\epsilon(Q \log Q)$*

A proof appears in the appendix of the full version [6].

Let G be a group that acts on $\{0,1\}^n$. A function f is G-isomorphic to another function g if there is a $\phi \in G$ such that $f \circ \phi = g$. For a fixed function g, the problem of testing G-isomorphism to g is to test if an input function f is G-isomorphic to g, or ϵ-far from all functions that are G-isomorphic to g. A folklore result gives a trivial upper bound for the problem:

Lemma 1. *Testing G-isomorphism to a function g can be done in $O(\log |G|)$ queries.*

When G is the group of invertible affine transformations, the problem is known as affine isomorphism testing. The above lemma gives us the following corollary:

Corollary 2. *$O(n^2)$ queries suffice to test affine isomorphism.*

2.3 Parity decision trees

Parity decision trees extends the model of ordinary decision trees such that one may query the parity of a subset of input bits, i.e. the queries are of form "is $\sum_{i \in S} x_i \equiv 1 \pmod 2$? " for an arbitrary subset $S \subseteq [n]$. We call such queries parity queries.

Let f be a Boolean function. For a parity decision tree P_f for f, let $C(P_f, x)$ denote the number of parity queries made by P_f on input x. The parity decision tree complexity of f is $D_{\oplus}(f) = \min_{P_f} \max_x C(P_f, x)$.
Note that $D_{\oplus}(f) \leq D(f)$, where $D(f)$ is the deterministic decision tree complexity of f, as the queries made by a usual decision tree, "is $x_i = 1$?", are also valid parity queries.

A bounded error randomized parity decision tree $R^f_{\oplus}$ is a probability distribution over all deterministic decision trees such that for every input the expected error of the algorithm is bounded by $1/3$. The cost $C(R^f_{\oplus}, x)$ is the highest possible number of queries made by $R^f_{\oplus}$ on x, and the bounded error randomized decision tree complexity of f is $R_{\oplus}(f) = \min_{R^f_{\oplus}} \max_x C(R^f_{\oplus}, x)$

For a Boolean function f, it turns out that $R_{\oplus}(f)$ can be lower bounded by the randomized communication complexity of the so-called XOR function $f(x \oplus y)$ (See [23] for the definition of randomized communication complexity and XOR functions). So we have the following lemma.

Lemma 2. $R_{\oplus}(f) \geq \frac{1}{2} RCC(f(x \oplus y))$

Proof. Given a Boolean function $f : \{0,1\}^n \to \{0,1\}$ on n consider the communication game where x is with Alice and y is with Bob and they want to compute $f(x \oplus y)$ with error bounded by $1/3$. Let $RCC(f(x \oplus y))$ denote the randomized communication complexity of this communication game.
Given a randomized parity decision tree $R^f_{\oplus}$, Alice and Bob can convert it into a protocol by simulating the parity queries made by $R^f_{\oplus}$ by two bits of communication, and thus the inequality follows.

3 Property testing and parity trees

In this section we describe a relation between the testing complexity of a property of linear/quadratic functions and the parity decision tree complexity of an associated function. We remark that such connections have been observed before in the case of linear functions, though, to the best of our knowledge, such an observation had not been made for quadratic functions before our work.

3.1 Parity trees and linear functions

Let $e_i \in \{0,1\}^n$ denote the Boolean string whose i^{th} bit is 1 and all other bits are 0. For any linear function f let us define a string $B(f) \in \{0,1\}^n$ such that the i^{th} bit of $B(f)$ is 1 iff $f(e_i) = 1$. The following lemma is easy to prove:

Lemma 3. *The map $B : \mathcal{L} \to \{0,1\}^n$ gives a bijection between the set $\mathcal{L}$ and strings of length n.*

Now let $\mathcal{P} \subseteq \mathcal{L}$ be a set of linear functions. Given a linear function f we want a tester $\mathcal{T}$ that makes queries to the truth table of f and determines whether f is in $\mathcal{P}$ or is ϵ-far from $\mathcal{P}$. Let us define a set $S_{\mathcal{P}} \subseteq \{0,1\}^n$ as $S_{\mathcal{P}} := \{B(f) \mid f \in \mathcal{P}\}$.

Lemma 4. *For any $\mathcal{P} \subseteq \mathcal{L}$ and any $f \in \mathcal{L}$ we have:*

- *$f \in \mathcal{P}$ if and only if $B(f) \in S_{\mathcal{P}}$ and*
- *f is $1/2$-far from $\mathcal{P}$ if and only if $B(f) \notin S_{\mathcal{P}}$*

We omit the proof of Lemma 4 as it follows directly from Lemma 3 and Observation 1.

Now, by Lemma 4, testing where f is in $\mathcal{P}$ or is $1/2$-far from $\mathcal{P}$ is exactly same as deciding if $B(f) \in S_{\mathcal{P}}$. Furthermore, we can translate the queries made by the tester $\mathcal{T}$ to the truth table of f into parity queries to the string $B(f)$, and vice-versa. Since f is linear, we have $f(x) = \bigoplus_i x_i \cdot f(e_i)$. Now, if $S_x := \{i \mid x_i = 1\}$ then, whenever $\mathcal{T}$ queries f at x, it can be equivalently viewed as the query $\bigoplus_{i \in S_x} (B(f))_i$ made to $B(f)$.

Consider the Boolean function $E_{\mathcal{P}} : \{0,1\}^n \to \{0,1\}$, where $E_{\mathcal{P}}(x) = 1$ iff $B^{-1}(x) \in \mathcal{P}$. From the above discussion, deciding "is $x \in S_{\mathcal{P}}$?" is same as deciding "is $E_{\mathcal{P}}(x) = 1$?" Thus we have:

Theorem 6. *There is a tester that makes q queries for determining if a linear function f satisfies the property $\mathcal{P}$ or is $1/2$-far from satisfying $\mathcal{P}$ if and only if there is a randomized parity decision that makes q queries for computing $E_{\mathcal{P}}$. Equivalently, $Q_1(\mathcal{P}) = R_{\oplus}(E_{\mathcal{P}})$.*

Notice that Theorem 1 follows from this by using $Q^{1/2}(\mathcal{P}) \geq Q_1(\mathcal{P})$ from Observation 2.

3.2 Parity trees and quadratic functions

Let $\mathcal{G}_n \subseteq \{0,1\}^{n^2}$ denote the set of graphs on n vertices. For any homogeneous quadratic function $f \in \mathcal{Q}_0$ let us define a graph $G(f)$ with vertex set $[n]$ such that the edge $\{i,j\}$ is present in $G(f)$ iff $x_i x_j$ occurs as a monomial when f is expressed as a polynomial over $\mathbb{F}_2$. The following observation follows from the way we constructed $G(f)$:

Observation 4 *The map $G : \mathcal{Q}_0 \to \mathcal{G}_n$ is a bijection.*

Let $\mathcal{P} \subseteq \mathcal{Q}$ be a property of quadratic functions, and let $S_{\mathcal{P}} = \{G(f) \mid f \in \mathcal{P} \cap \mathcal{Q}_0\}$. It now easily follows from Observation 4 and 1 that:

Lemma 5. *For any $\mathcal{P} \subseteq \mathcal{Q}$ and any $f \in \mathcal{Q}_0$ we have:*

- *$f \in \mathcal{P}$ if and only if $G(f) \in S_{\mathcal{P}}$ and*
- *f is $1/4$-far from $\mathcal{P}$ if and only if $G(f) \notin S_{\mathcal{P}}$*

Thus, the above lemma says that testing whether a given $f \in \mathcal{Q}_0$ is in $\mathcal{P}$ or $1/4$-far from $\mathcal{P}$ is exactly the same as deciding if $G(f)$ is in $S_{\mathcal{P}}$.

Let $\mathcal{A}$ be an algorithm that tests if a $f \in \mathcal{Q}_0$ is in $\mathcal{P}$ or $1/4$-far from it. We now describe how to translate queries made by $\mathcal{A}$ to the truth table of f to parity queries to the adjacency matrix of the graph $G(f)$. Given $y \in \{0,1\}^n$ and a graph G on the vertex set $[n]$, we denote by $G[y]$ the induced graph on the vertex set $\{i|\ y_i = 1\}$. It is not hard to see that the value $f(y)$ is exactly the parity of the number of edges in $G(f)[y]$. Thus, any query to the truth table of f can be translated to a parity query to the adjacency matrix of $G(f)$, but unlike in the case of linear functions, the translation works only in one direction. To be more precise, an arbitrary parity query to the adjacency matrix of $G(f)$ cannot be translated into a query to the truth table of f.

Consider the Boolean function $E_{\mathcal{P}} : \mathcal{G}_n \to \{0,1\}$, where $E_{\mathcal{P}}(H) = 1$ iff $G^{-1}(H) \in \mathcal{P}$, and observe that deciding "is $H \in S_{\mathcal{P}}$?" is same as deciding "is $E_{\mathcal{P}}(H) = 1$?". Combining the observations made above, we get:

Lemma 6. *There is an adaptive tester that makes q queries for determining if a given $f \in \mathcal{Q}_0$ satisfies the property $\mathcal{P}$ or is $1/4$-far from satisfying $\mathcal{P}$ only if there is a randomized parity decision that makes q queries for computing $E_{\mathcal{P}}$. Equivalently, $Q_2(\mathcal{P}) \geq R_{\oplus}(E_{\mathcal{P}})$.*

Combining Lemma 6 and Observation 2, we get a more general result:

Theorem 7 (Formal statement of Theorem 2). *There is an adaptive tester that makes q queries for determining if a given f satisfies the property $\mathcal{P}$ or is $1/4$-far from satisfying $\mathcal{P}$ only if there is a randomized parity decision tree that makes q queries for computing $E_{\mathcal{P}}$. Equivalently, $Q^{1/4}(\mathcal{P}) \geq R_{\oplus}(E_{\mathcal{P}})$.*

4 Characterizing testable properties of linear functions

In this section we give a characterization of properties of linear functions that are testable using constant number of queries. We will use some standard concepts from Fourier analysis of Boolean functions and the reader is referred to [26] for an introduction to the same.

Recall that for a Boolean[2] function f, $||\widehat{f}||_1^{\epsilon}$ denotes the minimum possible $||\widehat{g}||_1$ over all g such that $|f(x) - g(x)| \leq \epsilon$ for all x. We begin by proving the following lemma:

Lemma 7. *There are constants $c_1, c_2 > 0$ such that for sufficiently large n, if $f : \{0,1\}^n \to \{-1,+1\}$ is a Boolean function, then $c_1 \cdot \log ||\widehat{f}||_1^{1/4} \leq R_{\oplus}(f) \leq c_2 \cdot (||\widehat{f}||_1^{1/4})^2$*

[2] For the purpose of this section, it will be convenient to assume that the range of a Boolean function is $\{-1,+1\}$.

Proof. For the first inequality, we obtain from Lemma 2 that $RCC(f(x \oplus y)) \leq 2R_{\oplus}(f)$. Now, it is well known that $RCC(f(x \oplus y)) = \Omega(\log ||\widehat{f}||_1^{1/4})$ (see for instance [23]) and thus we have

$$R_{\oplus}(f) \geq 1/2 \cdot RCC(f(x \oplus y)) = \Omega(\log ||\widehat{f}||_1^{1/4}) \tag{1}$$

To see the second inequality, we will construct a randomized parity decision tree[3] $\mathcal{T}$ with query complexity $O((||\widehat{f}||_1^{1/4})^2)$ that computes f. Let $g : \{0,1\}^n \to \mathbb{R}$ be a function that point-wise 1/4-approximates f (i.e. for all x, $|f(x) - g(x)| \leq 1/4$) such that $||\widehat{g}||_1$ is the minimum among all functions that 1/4-approximate f. Let $\mathcal{D}_g$ denote a distribution on subsets of $[n]$ such that a set S has probability $|\hat{g}(S)|/||\widehat{g}||_1$.

We define the randomized parity decision tree $\mathcal{T}$ as follows. $\mathcal{T}$ makes d (the parameter will be fixed later) random parity queries $S_1, S_2 \ldots S_d$, such that each S_i is distributed according to $\mathcal{D}_g$. Let $X_1, X_2, \ldots X_d$ be random variables such that

$$X_i = \frac{\text{sign}(\hat{g}(S_i))(-1)^{\sum_{j \in S_i} x_j}}{||\widehat{g}||_1}$$

Here the sign function sign(x) outputs -1 is $x < 0$, and 1 otherwise. Finally, the tree outputs sign$(\sum_{i=1}^{d} X_i)$.
The first thing to note is that

$$\mathbb{E}[X_i] = \sum_{S \subseteq [n]} \frac{\text{sign}(\hat{g}(S_i))(-1)^{\sum_{j \in S_i} x_j}}{||\widehat{g}||_1} \frac{|\hat{g}(S)|}{||\widehat{g}||_1} = \frac{g(x)}{(||\widehat{g}||_1)^2}$$

Let $X = \sum_{i=1}^{d} X_i$. Then, $\mathbb{E}[X] = d \cdot g(x)/(||\widehat{g}||_1)^2$. Setting $d = 100 \cdot (||\widehat{g}||_1)^2$, we get $\mathbb{E}[X] = 100 \cdot g(x)$.

Now each X_i is bounded and lies in $[-1/||\widehat{g}||_1, +1/||\widehat{g}||_1]$. Thus by Hoeffding's inequality we have

$$\Pr[|X - \mathbb{E}[X]| \geq 50] \leq \exp\left(\frac{-2 \cdot (50)^2}{400}\right) = \exp\left(\frac{-25}{2}\right). \tag{2}$$

Since g point-wise 1/4-approximates f, sign$(g(x))$ = sign$(f(x))$ = $f(x)$. Also, it is easy to see that, if $|X - \mathbb{E}[X]| \leq 50$, sign$(X)$ = sign$(\mathbb{E}[X])$ = sign$(g(x))$. Thus, by Equation 2, sign$(X) = f(x)$ with very high probability.

The above argument shows that $\mathcal{T}$ is a randomized decision tree that computes f with high probability and makes $O((||\widehat{g}||_1)^2) = O((||\widehat{f}||_1^{1/4})^2)$ queries. This proves that

$$R_{\oplus}(f) = O((||\widehat{f}||_1^{1/4})^2) \tag{3}$$

Combining Equations 1 and 3 we can obtain the statement of the Lemma.

[3] We shall assume that $\mathcal{T}$'s range is $\{-1, +1\}$

Let $\mathcal{P}$ be a property of linear functions, and $Q_1(\mathcal{P})$ denote the query complexity of testing $\mathcal{P}$ when the input function is promised to be linear. Then, from the above lemma and Theorem 6, there exist constants $c_1, c_2 > 0$ such that for large enough n,

$$c_1 \cdot \log ||\widehat{E_{\mathcal{P}}}||_1^{1/4} \leq Q_1(\mathcal{P}) \leq c_2 \cdot (||\widehat{E_{\mathcal{P}}}||_1^{1/4})^2 \tag{4}$$

Using Observation 2 and 3 and Equation 4, we get, for $\epsilon \in (0, 1/4)$, there exists a constant c_2' that depends on ϵ such that for large enough n:

$$c_1 \cdot \log ||\widehat{E_{\mathcal{P}}}||_1^{1/4} \leq Q^{1/4}(\mathcal{P}) \leq Q^{\epsilon}(\mathcal{P}) \leq c_2'(||\widehat{E_{\mathcal{P}}}||_1^{1/4})^2 \log \left(||\widehat{E_{\mathcal{P}}}||_1^{1/4} \right)$$

Thus, we can conclude the Theorem 3 from the discussion: a property $\mathcal{P}$ of linear functions is testable using constant number of queries if and only if $||\widehat{E_{\mathcal{P}}}||_1^{1/4}$ is constant.

5 Testing k-linearity

In this section we apply the result from Section 3 to prove a lower bound for testing k-linearity. We shall use $\mathcal{P}$ to denote the set of k-linear functions on n variables.

Let $E_k : \{0,1\}^n \to \{0,1\}$ denote the Boolean function that outputs 1 if and only if the number of 1s is *exactly* k. Recall a notation from Section 3: for any linear function f we can define a string $B(f) \in \{0,1\}^n$ such that $B(f)_i = 1$ iff $f(e_i) = 1$. We observe the following:

Observation 5 *A Boolean function f is k-linear if and only if $B(f)$ has exactly k 1s.*

Thus, $E_{\mathcal{P}}$ is exactly the function E_k. Using Theorem 6 we have the following:

$$Q_1(\mathcal{P}) = R_{\oplus}(E_k) \tag{5}$$

Thus, if we can obtain a lower bound of $\Omega(k \log k)$ on the randomized parity decision tree complexity of E_k then we would obtain a tight bound for adaptive k-linearity testing (This would follow from Observation 2: $Q^{1/2}(\mathcal{P}) \geq Q_1(\mathcal{P})$). Unfortunately we are unable to obtain such a lower bound yet. Instead we can obtain a lower bound of $\Omega(k)$ that matches the previous known lower bound for k-linearity testing [8].

Using Lemma 2, we have that $R_{\oplus}(E_k) \geq \frac{1}{2} RCC(E_k(x \oplus y))$. Furthermore, Huang et al. [21] show that[4]:

Lemma 8. $RCC(E_k(x \oplus y)) = \Omega(k)$

[4] Actually, Huang et al. show that $RCC(E_{>k}(x \oplus y)) = \Omega(k)$, but their proof can be used to obtain the same lower bound for $RCC(E_k(x \oplus y))$. Alternatively, the lower bound may be obtained via a reduction to k-DISJ, a problem considered in [8].

Using Equation 5 and Lemma 8, we have $Q_1(\mathcal{P}) = \Omega(k)$. Finally, Observation 2 gives us $Q^{1/2}(\mathcal{P}) = \Omega(k)$:

Theorem 8 (Formal statement of Theorem 4). *Any adaptive two-sided tester for $1/2$-testing k-linearity must make $\Omega(k)$ queries.*

Thus we obtain a lower bound of $\Omega(k)$ using the lower bound for the randomized communication complexity of the XOR function $E_k(x \oplus y)$. Note that using this method we cannot expect to obtain a better lower bound as there is an upper bound of $O(k)$ on the communication complexity. But there is hope that one may be able to obtain a better lower bound for the parity decision tree complexity of E_k directly.

On the other hand, if one is able to construct a randomized parity decision tree of depth $O(k)$ for deciding E_k, Lemma 5 immediately implies a tester for k-linearity that makes $O(k)$ queries under the promise that the input function is linear. Notice that the exact complexity for even the promise problem is not known and the best upper bound is $O(k \log k)$. (while, naturally, the lower bound is $\Omega(k)$.)

6 Testing affine isomorphism to the inner product function

The main result of this section is that $1/4$-testing affine isomorphism to the inner product function $\mathbf{IP}_n(x)$[5] requires $\Omega(n^2)$ queries. As a corollary, we show that testing the set of Bent functions requires $\Omega(n^2)$ queries.

Let $\mathcal{B}$ denote the set of Bent functions (See [27] for a definition). The following is a consequence of Dickson's lemma (We omit the proof here, but it appears in the full version [6])

Lemma 9. *Let $Q(n)$ denote the the query complexity of $1/4$-testing affine isomorphism to the inner product function. Then $Q^{1/4}(\mathcal{B} \cap \mathcal{Q}) = O(Q(n))$.*

Thus, it is sufficient to lower bound $Q^{1/4}(\mathcal{B} \cap \mathcal{Q})$. In fact, by Observation 2, $Q^{1/4}(\mathcal{B} \cap \mathcal{Q}) \geq Q_2(\mathcal{B} \cap \mathcal{Q})$, and thus we can restrict out attention to lower bounding $Q_2(\mathcal{B} \cap \mathcal{Q})$

Recall from Section 3 that we can associate a graph $G(f)$ with every function $f \in \mathcal{Q}_0$. We now state a criterion for a quadratic function to be Bent that follows from a result due to Rothaus [27].

Lemma 10. *A function $f \in \mathcal{Q}_0$ is Bent iff the adjacency matrix of $G(f)$ is nonsingular.*

We omit the proof due to space constraints and give a proof in the appendix of the full version [6].

Recall from Section 3 that $\mathcal{G}_n \subseteq \{0,1\}^{n^2}$ is the set of graphs on the vertex set $[n]$. Let $\mathcal{P} := \mathcal{B} \cap \mathcal{Q}$, and let $E_n : \mathcal{G}_n \to \{0,1\}$ be a Boolean function such

[5] For the rest of the section we shall assume that the number of variables n is even

that $E_n(G) = 1$ iff the adjacency matrix of G is nonsingular. Due to Lemma 10, $E_{\mathcal{P}}$ turns out to be exactly equal to E_n. Combining with Theorem 5, we have

$$Q_2(\mathcal{P}) \geq R_{\oplus}(E_n) \tag{6}$$

As in the case of E_k, analyzing the decision tree complexity of E_n directly is hard, and we turn to communication complexity. Lemma 2 tells us that $R_{\oplus}(E_n) \geq \frac{1}{2}RCC(E_n(x \oplus y))$.

Let $M_n(\mathbb{F}_2)$ denote the set of $n \times n$ matrices over $\mathbb{F}_2$, and $Det_n : M_n(\mathbb{F}_2) \to \{0,1\}$ be the function such that $Det_n(A) = 1$ iff $A \in M_n(\mathbb{F}_2)$ is nonsingular. The following result from [29] analyzes the communication complexity of Det_n.

Lemma 11. $RCC(Det_n(x \oplus y)) = \Omega(n^2)$

It turns out that the communication complexity of Det_n relates to that of E_n.

Lemma 12. $= RCC(Det_n(x \oplus y)) \leq RCC(E_{2n}(x \oplus y))$

Proof. Let $A \in M_n(\mathbb{F}_2)$. Consider the $2n \times 2n$ matrix A' given by

$$A' := \begin{pmatrix} 0 & A^t \\ A & 0 \end{pmatrix}$$

$A' \in \mathcal{G}_{2n}$ by construction and it can be easily verified that A' is nonsingular iff A is nonsingular.
Now, let the inputs to Alice and Bob be A and B respectively. Since $(A \oplus B)' = A' \oplus B'$, $Det_n(A \oplus B) = 1$ iff $E_{2n}((A \oplus B)') = 1$ iff $E_{2n}(A' \oplus B') = 1$. Thus, to determine if $Det_n(A \oplus B)$ is 1, Alice and Bob can construct A' and B' from A and B respectively, and run the protocol for E_{2n} on A' and B'. This completes the proof.

Thus, using Lemma 11, we have $RCC(E_n(x \oplus y)) = \Omega(n^2)$. Using Lemma 2 and Equation 6, we have that $Q_2(\mathcal{P}) = \Omega(n^2)$.
Thus, based on earlier observations, we can conclude:

Theorem 9 (Formal statement of Theorem 5). *Any adaptive two-sided tester for 1/4-testing affine isomorphism to the inner product function* $\mathbf{IP}_n(x)$ *requires* $\Omega(n^2)$ *queries.*

Corollary 2 tells us that our result is tight. Thus, $\mathbf{IP}_n(x)$ is an example of a function for which the trivial bound for testing affine isomorphism is the best possible.

We have shown that $Q^{1/4}(\mathcal{B} \cap \mathcal{Q}) = \Omega(n^2)$. We now state a result due to Chen et al.(Lemma 2 in [15]) in a form that is suitable for application in our setting:

Lemma 13. *Let* $\mathcal{P}_1$ *and* $\mathcal{P}_2$ *be two properties of Boolean functions that have testers (possibly two-sided)* T_1 *and* T_2 *respectively. Let the query complexity of tester* T_i *be* $q_i(\epsilon, n)$*. Suppose* $dist(\mathcal{P}_1 \backslash \mathcal{P}_2, \mathcal{P}_2 \backslash \mathcal{P}_1) \geq \epsilon_0$ *for some absolute constant* ϵ_0*. Then,* $\mathcal{P}_1 \cap \mathcal{P}_2$ *is* ϵ*-testable with query complexity*

$$O(max\{q_1(\epsilon, n), q_1(\frac{\epsilon_0}{2}, n)\} + max\{q_2(\epsilon, n), q_2(\frac{\epsilon_0}{2}, n)\})$$

In its original form, the lemma has been proven for the case when T_1, T_2 are one-sided, and q_1, q_2 are independent of n, but the proof can be easily adapted to this more general setting.

Another consequence of Dickson's lemma is the following (A proof appears in the full version [6]):

Lemma 14. *Let f, g be Boolean functions. If $f \in \mathcal{B} \setminus \mathcal{Q}$ and $g \in \mathcal{Q} \setminus \mathcal{B}$, then $dist(f, g) \geq 1/4$.*

We are now ready to prove a lower bound for testing Bent functions.

Theorem 10 (Formal statement of Corollary 1). *Any adaptive two-sided tester that $1/8$-tests the set of Bent functions requires $\Omega(n^2)$ queries.*

Proof. It is well known via [2] that $\mathcal{Q}$ is testable with constant number of queries (say $q_1(\epsilon)$). Suppose there is a tester that tests $\mathcal{B}$ using $q_2(\epsilon, n)$ queries. From Lemma 14, we know that $dist(\mathcal{B} \setminus \mathcal{Q}, \mathcal{Q} \setminus \mathcal{B}) \geq \frac{1}{4}$. Thus, by Lemma 13, we have that there is a tester that makes $O(max\{q_1(\epsilon), q_1(\frac{1}{8})\} + max\{q_2(\epsilon, n), q_2(\frac{1}{8}, n)\})$ queries to ϵ-test $\mathcal{B} \cap \mathcal{Q}$.

Setting $\epsilon = \frac{1}{4}$, we have a tester that makes $O(q_1(\frac{1}{8}) + q_2(\frac{1}{8}, n))$ queries to test if a given f is in $\mathcal{B} \cap \mathcal{Q}$, or $1/4$-far from it. Since $Q^{1/4}(\mathcal{B} \cap \mathcal{Q}) = \Omega(n^2)$ and $q_1(\frac{1}{8})$ is a constant, we get $q_2(\frac{1}{8}, n) = \Omega(n^2)$, which completes the proof.

References

1. Alon, N., Fischer, E., Newman, I., Shapira, A.: A combinatorial characterization of the testable graph properties: It's all about regularity. SIAM J. Comput. 39(1), 143–167 (2009)
2. Alon, N., Kaufman, T., Krivelevich, M., Litsyn, S., Ron, D.: Testing low-degree polynomials over gf(2). In: RANDOM-APPROX 2003. pp. 188–199
3. Bhattacharyya, A., Fischer, E., Hatami, H., Hatami, P., Lovett, S.: Every locally characterized affine-invariant property is testable. In: Proceedings of the 45th annual ACM symposium on Symposium on theory of computing. pp. 429–436. STOC '13, ACM, New York, NY, USA (2013), http://doi.acm.org/10.1145/2488608.2488662
4. Bhattacharyya, A., Grigorescu, E., Shapira, A.: A unified framework for testing linear-invariant properties. In: In Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science. pp. 478–487 (2010)
5. Bhrushundi, A.: On testing bent functions. Electronic Colloquium on Computational Complexity (ECCC) 20, 89 (2013)
6. Bhrushundi, A., Chakraborty, S., Kulkarni, R.: Property testing bounds for linear and quadratic functions via parity decision trees. Electronic Colloquium on Computational Complexity (ECCC) 20, 142 (2013)
7. Blais, E.: Testing juntas nearly optimally. In: Proc. ACM symposium on the Theory of computing. pp. 151–158. ACM, New York, NY, USA (2009)
8. Blais, E., Brody, J., Matulef, K.: Property testing via communication complexity. Proc. CCC (2011)
9. Blais, E., Kane, D.M.: Tight bounds for testing k-linearity. In: APPROX-RANDOM. pp. 435–446 (2012)

10. Blais, E., Weinstein, A., Yoshida, Y.: Partially symmetric functions are efficiently isomorphism-testable. In: FOCS. pp. 551–560 (2012)
11. Blum, M., Luby, M., Rubinfeld, R.: Self-testing/correcting with applications to numerical problems. In: STOC. pp. 73–83 (1990)
12. Buhrman, H., García-Soriano, D., Matsliah, A., de Wolf, R.: The non-adaptive query complexity of testing k-parities. CoRR abs/1209.3849 (2012)
13. Chakrabarty, D., Seshadhri, C.: A o(n) monotonicity tester for boolean functions over the hypercube. CoRR abs/1302.4536 (2013)
14. Chakraborty, S., Fischer, E., García-Soriano, D., Matsliah, A.: Junto-symmetric functions, hypergraph isomorphism and crunching. In: IEEE Conference on Computational Complexity. pp. 148–158 (2012)
15. Chen, V., Sudan, M., Xie, N.: Property testing via set-theoretic operations. In: ICS. pp. 211–222 (2011)
16. Fischer, E.: The art of uninformed decisions: A primer to property testing. Science 75, 97–126 (2001)
17. Fischer, E., Kindler, G., Ron, D., Safra, S., Samorodnitsky, A.: Testing juntas. Journal of Computer and System Sciences 68(4), 753 – 787 (2004), special Issue on FOCS 2002
18. Fischer, E., Lehman, E., Newman, I., Raskhodnikova, S., Rubinfeld, R., Samorodnitsky, A.: Monotonicity testing over general poset domains. In: STOC. pp. 474–483 (2002)
19. Goldreich, O.: On testing computability by small width obdds. In: APPROX-RANDOM. pp. 574–587 (2010)
20. Grigorescu, E., Wimmer, K., Xie, N.: Tight lower bounds for testing linear isomorphism. In: APPROX-RANDOM. pp. 559–574 (2013)
21. Huang, W., Shi, Y., Zhang, S., Zhu, Y.: The communication complexity of the hamming distance problem. Inf. Process. Lett. 99(4), 149–153 (2006)
22. Kaufman, T., Sudan, M.: Algebraic property testing: the role of invariance. In: STOC. pp. 403–412 (2008)
23. Lee, T., Shraibman, A.: Lower bounds in communication complexity. Foundations and Trends in Theoretical Computer Science 3(4), 263–398 (2009)
24. MacWilliams, F.J., Sloane, N.J.A.: The Theory of Error-Correcting Codes (North-Holland Mathematical Library). North Holland Publishing Co. (Jun 1988), http://www.worldcat.org/isbn/0444851933
25. Neumann, T.: Bent functions (2006), (Master's thesis)
26. O'Donnell, R.: Analysis of boolean functions (2012), http://www.analysisofbooleanfunctions.org
27. Rothaus, O.: On bent functions. Journal of Combinatorial Theory, Series A 20(3), 300 – 305 (1976), http://www.sciencedirect.com/science/article/pii/0097316576900248
28. Rubinfeld, R., Shapira, A.: Sublinear time algorithms. Electronic Colloquium on Computational Complexity (ECCC) 11(013) (2011)
29. Sun, X., Wang, C.: Randomized communication complexity for linear algebra problems over finite fields. In: STACS. pp. 477–488 (2012)
30. Wimmer, K., Yoshida, Y.: Testing linear-invariant function isomorphism. In: ICALP (1). pp. 840–850 (2013)
31. Zhang, Z., Shi, Y.: On the parity complexity measures of boolean functions. Theor. Comput. Sci. 411(26-28), 2612–2618 (2010)