Electronic Colloquium on Computational Complexity, Report No. 90 (2007)

Tight lower bounds for adaptive linearity tests

Shachar Lovett[1]

Abstract

Linearity tests are randomized algorithms which have oracle access to the truth table of some function f, which are supposed to distinguish between linear functions and functions which are far from linear. Linearity tests were first introduced by Blum, Luby and Rubenfeld in [BLR93], and were later used in the PCP theorem among other applications. The quality of a linearity test is described by its *correctness* c - the probability it accepts linear functions, its soundness s - the probability it accepts functions far from linear, and its *query complexity* q - the number of queries it makes. The BLR test had $q = 3$ and $s = 1/2$. Linearity tests were studied in order to decrease the soundness of linearity tests, while keeping the query complexity small (for one reason, to improve PCP constructions). Samorodnitsky and Trevisan constructed in [ST00] the Complete Graph Test, which for every $k \in \mathbb{N}$ has $q = \binom{k}{2} + k$ and $s = 2^{-\binom{k}{2}}$. They prove that no Hyper Graph Test can perform better than the Complete Graph Test. Later in [ST06] they prove, among other results, that no non-adaptive linearity test can perform better than the Complete Graph Test. We generalize their result for adaptive tests, and prove that the Complete Graph Test is optimal even against adaptive linearity tests. Our lower bound is actually proven in a more general setting, considering the *Average Query Complexity* of a linearity test. Our proof technique is somewhat different from the one used in [ST06]. In both cases the behavior of linearity tests against quadratic functions are considered, but while [ST06] uses algebraic analysis of the Gowers Norm of certain functions, we use a more direct combinatorial approach, which allows us to also handle the case of adaptive linearity tests.

1 Introduction

We study the relation between the number of queries and soundness of adaptive linearity tests. A linearity test (over the field $\mathbb{F}_2$ for example) is a randomized algorithm which has oracle access to the truth table of a function $f : \{0,1\}^n \to \{0,1\}$, and needs to distinguish between the following two extreme cases:

1. f is linear
2. f is far from linear functions

[1]Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, POB 26, Rehovot 76100, Israel. Email: Shachar.Lovett@weizmann.ac.il. This research was supported by grant 1300/05 from the Israel Science Foundation.

ISSN 1433-8092

A function f is called *linear* if it can be written as $f(x_1, ..., x_n) = a_1x_1 + ... + a_nx_n$, with $a_1, ..., a_n \in \mathbb{F}_2$. The agreement of two functions $f, g : \{0,1\}^n \to \{0,1\}$ is defined as $d(f,g) = |\mathbb{P}_{\mathbf{x}}[f(\mathbf{x}) = g(\mathbf{x})] - \mathbb{P}_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})]|$. f is far from linear functions if it has small agreement with all linear functions (we make this definition precise in Section 2).

Linearity tests were first introduced by Blum, Luby and Rubenfeld in [BLR93]. They presented the following test (coined the BLR test), which makes only 3 queries to f:

1. Choose $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ at random
2. Verify that $f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$.

Bellare et al. [BCH+96] gave a tight analysis of the BLR test. It is obvious that the BLR test always accepts a linear function. They have shown that if the test accepts a function f with probability $1/2 + \epsilon$, then f has agreement at least 2ϵ with some linear function.

For a linearity test, we define that it has *completeness* c if it accepts any linear function with probability of at least c. A test has *perfect completeness* if $c = 1$. A linearity test has *soundness* s if it accepts any function f with agreement at most ϵ with all linear functions, with probability of at most $s + \epsilon'$, where $\epsilon' \to 0$ when $\epsilon \to 0$. We define the *query complexity* q of a test as the maximal number of queries it performs. In the case of the BLR test, it has perfect completeness, soundness $s = 1/2$ (with $\epsilon' = 2\epsilon$) and query complexity $q = 3$.

If one repeats a linearity test with query complexity q and soundness s independently t times, the query complexity grows to $q' = qt$ while the soundness reduces to $s' = s^t$. So, it makes sense to define the *amortized query complexity* $\bar{q}$ of a test as $\bar{q} = q/\log_2(1/s)$. Independent repetition of a test doesn't change it's amortized query complexity. Notice that the BLR test has amortized query complexity $\bar{q} = 3$.

Linearity tests are a key ingredient in the PCP theorem, started in the works of Arora and Safra [AS98] and Arora, Lund, Motwani, Sudan and Szegedy [ALM+98]. In order to improve PCP constructions, linearity tests were studied in order to improve their amortized query complexity.

Samorodnitsky and Trevisan [ST00] have generalized the basic BLR linearity test. They introduced the *Complete Graph Test*. The Complete Graph Test (on k vertices) is:

1. Choose $\mathbf{x}_1, ..., \mathbf{x}_k \in \{0,1\}^n$ independently
2. Verify $f(\mathbf{x}_i + \mathbf{x}_j) = f(\mathbf{x}_i) + f(\mathbf{x}_j)$ for all i, j

This test has perfect completeness and query complexity $q = \binom{k}{2} + k$. They show that all the $\binom{k}{2}$ tests that the Complete Graph Test performs are essentially independent, i.e. that the test has soundness $s = 2^{-\binom{k}{2}}$. This makes this test have amortized query complexity $\bar{q} = 1 + \theta(1/\sqrt{q})$. They show that this test is optimal among the family of Hyper-Graph Tests (see [ST00] for definition of this family of linearity tests), and raise the question of whether the Complete Graph Test is optimal among all linearity tests, i.e. does a test with the same query complexity but with better soundness exist?

They partially answer this question in [ST06], where (among many other results) they show that no non-adaptive linearity test can perform better than the Complete

Graph Test. A test is called *non-adaptive* if it first chooses q locations in the truth table of f, then queries them, and based on the results accept or rejects f. Otherwise, a test is called *adaptive*. An adaptive test may decide on its query locations based on the values of f in previous queries.

In this paper we generalize the lower bound result of [ST06], and show that the Complete Graph Test is indeed optimal among all adaptive tests as well.

1.1 Our techniques

We analyze the behavior of linearity tests on quadratic functions, similar to what has been done in [ST06]. However, we use a more direct combinatorial approach for analysis of linearity tests, while [ST06] used a more algebraic approach (using Gowers Norm).

We model adaptive tests using test trees. A test tree T is a binary tree, where in each inner vertex v there is some label $\mathbf{x}(v) \in \{0,1\}^n$, and the leaves are labeled with either *accept* or *reject*. Running a test tree on a function f is done by querying at each stage f on the label of the current vertex (starting at the root), and following one of the two edges leaving the vertex, depending on the query response. When reaching a leaf, its label (*accept* or *reject*) is the value of that f gets in T . An adaptive test $\mathbb{T}$ can always be modeled as first randomly choosing a test tree from some set $\{T_i\}$, according to some distribution on the test trees, then running the test tree on f.

It turns out that in order to prove the lower bound, it is enough to consider functions f which are quadratic. A function f is quadratic if it can be presented as $f(x_1,...,x_n) = \sum_{i,j} a_{i,j}x_ix_j + \sum_i b_ix_i + c$ for some values $a_{i,j}, b_i, c \in \mathbb{F}_2$. We study the behavior of running test trees on a random linear function, and on a random quadratic function.

The main idea is as follows. Let v be some inner vertex in a test tree T, with the path from the root of T to v being $v_0,...,v_{k-1},v$. If $\mathbf{x}(v)$ is linearly dependent on $\mathbf{x}(v_0),...,\mathbf{x}(v_{k-1})$, then when running T on any linear function, the value of $f(\mathbf{x}(v))$ can be deduced from the already known values of $f(\mathbf{x}(v_0)),...,f(\mathbf{x}(v_{k-1}))$. Therefore, if the vertex v is reached, then the same edge leaving v will always be taken by any linear function. Additionally, if $\mathbf{x}(v)$ is linearly independent of $\mathbf{x}(v_0),...,\mathbf{x}(v_{k-1})$, then either v is never reached running T on linear functions, or the two edges leaving v are taken with equal probability when running T on a random linear function. A similar analysis can be made when running T on quadratic functions, replacing *linear dependence* with a corresponding notion of *quadratic dependence*.

Using this observation, we can define the *linear rank* of a leaf v, marked $l(v)$, as the linear rank of labels on the path from the root to v. Similarly, we define the *quadratic rank* of a leaf v, marked $q(v)$, as the quadratic rank of those labels. We prove that the quadratic rank of any set cannot be much larger than its linear rank, and in particular that $q(v) \leq \binom{l(v)}{2} + l(v)$ for all leaves v. We use this inequality to prove that a test which has completeness c and query complexity q accepts a random quadratic function with a probability of at least $c - 1 + 2^{-q+\phi(q)}$, where $\phi(q)$ is defined as the unique non-negative solution to $\binom{\phi(q)}{2} + \phi(q) = q$.

We use this to show that any linearity test with completeness c and query complexity q must have $s \geq 2^{-q+\phi(q)}$. In particular, the Complete Graph Test on k vertices has

perfect completeness, soundness $s = 2^{-\binom{k}{2}}$ and query complexity $q = \binom{k}{2} + k$. Since $\phi(q) = k$ the Complete Graph Test is optimal among all adaptive tests with the same query complexity.

In fact, we prove a stronger claim. We say that a test $\mathbb{T}$ has *average query complexity* q if for any function f, the average number of queries performed is at most q. In particular any test with query complexity q also has average query complexity q. We prove that for any test with completeness c and average query complexity q, the soundness is at least $s \geq 2^{-q+\phi(q)}$.

We present and analyze linearity tests over $\mathbb{F}_2$. Linearity tests can also be considered over larger fields or groups. Our lower bound actually generalizes easily to any finite field, but for ease of presentation, and since the techniques are exactly the same, we present everything over $\mathbb{F}_2$. We comment further on the modifications required for general finite fields in Section 2.

2 Preliminaries

2.1 Linearity tests

We call a function $f : \{0,1\}^n \to \{0,1\}$ linear if it can be written as $f(x_1, ..., x_n) = a_1x_1 + ... + a_nx_n$ for some $a_1, ..., a_n \in \{0,1\}$ where addition and multiplication are in $\mathbb{F}_2$.

A linearity test is a randomized algorithm with oracle access to the truth table of f, which is supposed to distinguish the following two extreme cases:

1. f is linear (accept)
2. f is ϵ-far from linear functions (reject)

where the agreement of two functions $f, g : \{0,1\} \to \{0,1\}$ is defined as $d(f,g) = |Pr_{\mathbf{x}}[f(\mathbf{x}) = g(\mathbf{x})] - Pr_{\mathbf{x}}[f(\mathbf{x}) \neq g(\mathbf{x})]|$, and f is ϵ-far from linear functions if the agreement it has with any linear function is at most ϵ.

We now follow with some standard definition regarding linearity tests (or more generally, property tests). We say a test has *completeness* c if for any linear function f the test accepts with probability at least c. A test has *perfect completeness* if $c = 1$. We say a test has *soundness* s if for any f which is ϵ-far from linear the test accepts with probability at most $s + \epsilon'$, where $\epsilon' \to 0$ when $\epsilon \to 0$ (in fact, we talk about a family of linearity tests, for $n \to \infty$, but we ignore this subtle point).

A test is said to have *query complexity* q if it accesses the truth-table of f at most q times (for any choice of it's internal randomness). A test is said to have *average query complexity* q if for any function f, the average number of accesses (over the internal randomness of the test) done to the truth table of f is at most q. Obviously, any test with query complexity q is also a test with average query complexity q.

We say a test is *non-adaptive* if it chooses all the locations it's going to query in the truth table of f before reading any of their values. Otherwise, we call the test *adaptive*.

We now turn to model adaptive tests in a way that will be more convenient for our analysis. We first define a test tree and running a test tree on a function.

Definition 1. A *test tree* on functions $\{0,1\}^n \to \{0,1\}$ is a rooted binary tree T. On each inner vertex of the tree v there is a label $x(v) \in \{0,1\}^n$. On each leaf there is a label of either *accept* or *reject*.

Definition 2. *Running a test tree T on a function f* is done as follows. We start at the root of the tree v_0, read the value of $f(x(v_0))$, and according to the value take the left or the right edge leaving v_0. We continue in this fashion on inner vertices of T until we reach a leaf of T. The *value of f in T* is the value of the end leaf (i.e. *accept* or *reject*), and the *depth of f in T* is the depth of the end vertex of f in T.

Using these definitions, we can now model adaptive tests. We identify an adaptive test $\mathbb{T}$ on functions $\{0,1\}^n \to \{0,1\}$ with a distribution of binary trees $\{T_i\}$ (also on functions $\{0,1\}^n \to \{0,1\}$). Running the test $\mathbb{T}$ on a function f is done by randomly choosing one of the trees T_i (according to their distribution), and then running the test tree T_i on f. The result of the function f in the test tree T_i is the result the test $\mathbb{T}$ returns on f.

Notice that a test has query complexity q iff all trees T_i has depth at most q, and has average query complexity q iff for any function f, the average depth reached in a random tree from $\{T_i\}$ is at most q.

In order to define our main theorem, we will define the following function. For $x > 0$ define $\phi(x)$ as the unique real positive solution to $\phi(x)^2/2 + \phi(x)/2 = x$. Notice that for positive integer $\phi(x)$, this is the same as $\binom{\phi(x)}{2} + \phi(x) = x$. The following is the main theorem of this paper:

Theorem 1. *(main theorem) Let $\mathbb{T}$ be an adaptive test with completeness c, soundness s and average query complexity $q \geq 1$. Then $s + 1 - c \geq 2^{-q+\phi(q)}$.*

Notice that for large q, $\phi(q) \approx \sqrt{2q}$, also $\sqrt{q} \leq \phi(q) \leq \sqrt{2q}$, so we get that in particular, $s + 1 - c \geq 2^{-q+\theta(\sqrt{q})}$.

The Complete Graph Test was presented in [ST00]. The test (on a graph with k vertices) can be described as choosing $\mathbf{x}_1, ..., \mathbf{x}_k$ at random, and querying f at $\mathbf{x}_i$ (for $i = 1..k$) and on $\mathbf{x}_i + \mathbf{x}_j$ (for $1 \leq i < j \leq k$). The test accepts f if for any i, j

$$f(x_i) + f(x_j) + f(x_i + x_j) = 0$$

In [ST00] it is proven that the Complete Graph Test has perfect completeness and soundness $s = 2^{-\binom{k}{2}}$. The total number of queries performed is $q = k + \binom{k}{2}$, so by our definitions, $k = \phi(q)$ and $s = 2^{-q+\phi(q)}$. We have the following corollary:

Corollary 2. *The Complete Graph Test is optimal among all adaptive linearity tests.*

Remark. We state and prove all results for functions $f : \{0,1\}^n \to \{0,1\}$. In fact, the lower bound result on adaptive linearity tests holds for functions $f : \mathbb{F}^n \to \mathbb{F}$ for any finite field $\mathbb{F}$, and not just $\mathbb{F}_2$, with only minor adjustments to the definitions and proofs. We need to make the following modifications:

1. Define "ϵ-far from linear functions" for general fields
2. Test trees should have $|F|$ edges leaving each edge instead of 2

3. The proof that random quadratic functions are far from linear, proved in Section 5, should be slightly modified

Since the results follow simply for any finite field, we chose to present the results over $\mathbb{F}_2$ to make the presentation simpler and clearer.

3 Quadratic functions

We will see that in order to prove Theorem 1, it will be enough to limit the functions f to be quadratic. We say a function f is quadratic if it can be written as:

$$f(x_1, ..., x_n) = \sum_{i,j} a_{i,j} x_i x_j + \sum_i b_i x_i + c$$

for some $a_{i,j}, b_i, c \in \mathbb{F}_2$.

In fact, for our usage, we will force our quadratic functions f to have $f(0) = 0$ (equivalently, $c = 0$ in the above description). So, throughout this paper, when speaking of quadratic functions, we actually speak of quadratic functions f with the added condition $f(0) = 0$.

We will study the dynamics of a test tree T in a linearity test $\mathbb{T}$, in two cases - when applied to a uniformly random linear function, and when applied to a uniformly random quadratic function.

The following technical lemma is the key ingredient to the proof of the Theorem 1.

Lemma 3. *Let $\mathbb{T}$ be an adaptive linearity test with completeness c and average query complexity q. Then running $\mathbb{T}$ on a random quadratic function returns* accept *with probability at least $c - 1 + 2^{-q+\phi(q)}$.*

In order to prove Theorem 1, we will also need the following simple lemma:

Lemma 4. *Let f be a random quadratic function. Then the probability that f is not $2^{-\Omega(n)}$-far from linear functions is $2^{-\Omega(n)}$.*

Theorem 1 now follows directly from Lemmas 3 and 4. We sketch now it's proof following the two lemmas.

Proof. (of the main theorem) The average probability that $\mathbb{T}$ returns *accept* on a random quadratic function which is $2^{-\Omega(n)}$-far from linear functions is at least $c - 1 + 2^{-q+\phi(q)} - 2^{-\Omega(n)}$. So, there exists some quadratic function f which is $2^{-\Omega(n)}$-far from linear and on which $\mathbb{T}$ returns *accept* with probability at least $c - 1 + 2^{-q+\phi(q)} - 2^{-\Omega(n)}$. Taking $n \to \infty$ shows that $s + 1 - c \geq 2^{-1+\phi(q)}$. □

The remainder of the paper is organized as follows. Lemma 3 is proved in Section 4, and Lemma 4 is proved in Section 5.

4 Linearity test applied to a random quadratic function

We study tests and test trees applied to linear and quadratic functions, in order to prove Lemma 3. Let $\mathbb{T}$ be an adaptive test with completeness c and average query complexity q. Let T be a some test tree which is a part of the test $\mathbb{T}$.

We start by studying the dynamics of applying T to linear functions. Assume we know that f is a linear function, and we are at some vertex $v \in T$, where the path from the root to v is $v_0, .., v_{k-1}, v$. Assume $\mathbf{x}(v)$ is linearly dependant on $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$. Since we know f is linear, we can deduce the value of $\mathbf{x}(v)$ from $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$, and so we will always follow the same edge leaving v when we apply T to any linear function. On the other hand, if $\mathbf{x}(v)$ is linearly independent of $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$, we know that when we apply T to a random linear function, either we never reach v, or we have equal chances of taking any of the two edges leaving v.

This gives rise to the following formal definition:

Definition 3. Let v be a leaf in T, where the path from the root to v is $v_0, v_1, ..., v_{k-1}, v$. We define the *linear degree* of v, marked $l(v)$, to be the linear rank of $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$.

We define L_T to be the set of leaves of T to which linear functions can arrive. i.e, $v \in L$ if the path from the root to v, $v_0, ..., v_{k-1}, v$ always takes the "correct" edge leaving any vertex v_i with $\mathbf{x}(v_i)$ linearly dependent on $\mathbf{x}(v_0), ..., \mathbf{x}(v_{i-1})$.

The following lemma formalizes the discussion above:

Lemma 5. *For any test tree T:*

1. *For any $v \in L_T$, the probability that a random linear function will arrive to v is $2^{-l(v)}$*
2. $\sum_{v \in L_T} 2^{-l(v)} = 1$

For $v \in L_T$, we define $c(v)$ to be 1 if the value of v is *accept*, and $c(v) = 0$ otherwise. Since the completeness of $\mathbb{T}$ is c, we have that the probability that $\mathbb{T}$ returns *accept* on a random linear function is at least c. On the other hand, for any test tree T in $\mathbb{T}$, the probability that a random linear function will return *accept* is exactly $\sum_{v \in L_T} c(v) 2^{-l(v)}$. So, the following lemma follows:

Lemma 6. $\mathbb{E}_T \sum_{v \in L_T} c(v) 2^{-l(v)} \geq c$

where by $\mathbb{E}_T$ here and throughout the paper we mean the average value of a random test tree T in $\mathbb{T}$.

We now generalize the concept of linear dependence to quadratic functions.

Definition 4. Let $\mathbf{x}_1, ..., \mathbf{x}_k \in \{0,1\}^n$.

1. We say $\mathbf{x}_1, ..., x_k$ are *quadratically dependent* if there are constants $a_1, ..., a_k \in \mathbb{F}_2$, not all zero, s.t. for any quadratic function f we have: $a_1 f(\mathbf{x}_1) + ... + a_k f(\mathbf{x}_k) = 0$. otherwise will call $x_1, ..., x_k$ *quadratically independent.*

2. We say $\mathbf{x}_k$ is *quadratically dependent* on $\mathbf{x}_1, ..., \mathbf{x}_{k-1}$ if there are constants $a_1, ..., a_{k-1} \in \mathbb{F}_2$ s.t. for any quadratic function f we have: $f(\mathbf{x}_k) = a_1 f(\mathbf{x}_1) + ... + a_{k-1} f(\mathbf{x}_{k-1})$. Otherwise we say $\mathbf{x}_k$ is *quadratically independent* of $\mathbf{x}_1, ..., \mathbf{x}_{k-1}$.
3. We define the *quadratic dimension* of $\mathbf{x}_1, ..., \mathbf{x}_k$ to be the size of the largest subset of $\{\mathbf{x}_1, ..., \mathbf{x}_k\}$ which is quadratically independent.

This definition may seem obfuscated, but the following alternative yet equivalent definition will clarify it. The space of quadratic functions over $\{0,1\}^n$ is a linear space over $\mathbb{F}_2$. Let M be it's generating matrix, i.e. the rows of M are a base for the linear space (in particular, the dimensions of M are $(\binom{n}{2} + n) \times 2^n$). A column of M corresponds to an input $\mathbf{x} \in \{0,1\}^n$. Now, $\mathbf{x}_1, ..., \mathbf{x}_k$ are quadratically dependent iff the columns corresponding to them are linearly dependent, and similarly for the other definitions.

Notice that the usual definition of linear dependence is equivalent to this more complex definition, when applied to the linear space of all linear functions.

We now can repeat the informal discussion at the start of this section, except this time for quadratic functions, with all the reasoning left intact. Let $v \in T$ be a vertex, with path from the root being $v_0, ..., v_{k-1}, v$. Assume $\mathbf{x}(v)$ is quadratically dependent on $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$, and f is any quadratic function. The value of $f(\mathbf{x}(v))$ can be deduced from the already known values of $f(\mathbf{x}(v_0)), ..., f(\mathbf{x}(v_{k-1}))$, and so only one edge leaving v will be taken on all quadratic functions. Alternatively, if $x(v)$ is quadratically independent on $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$, then a random quadratic function either never reaches v, or has equal chances of taking each of the two edges leaving v.

This leads to the following definition and lemma for quadratic degree of a vertex $v \in T$, similar to the ones for linear degree.

Definition 5. Let v be a leaf in T, where the path from the root to v is $v_0, v_1, ..., v_{k-1}, v$. We define the *quadratic degree* of v, marked $q(v)$, to be the quadratic rank of $\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})$.

We define Q_T to be the set of leaves of T to which quadratic functions can arrive. Naturally $L_T \subseteq Q_T$. The following lemma on quadratic degree follows from the discussion above:

Lemma 7. *For any test tree T:*

1. *For any $v \in Q_T$, the probability that a random quadratic function will arrive to v is $2^{-q(v)}$*
2. $\sum_{v \in Q} 2^{-q(v)} = 1$
3. *For any $v \in L_T$ we have $q(v) \geq l(v)$*

Last, we mark the depth of a vertex $v \in T$ by $d(v)$. Since $\mathbb{T}$ has average query complexity q, we know that for any function f, the average depth of running a random tree T of $\mathbb{T}$ on f is at most q. So, this also holds for a random linear function. However, the average depth a random linear function arrives on a tree T is exactly $\sum d(v) 2^{-l(v)}$, so the following lemma follows.

Lemma 8. $\mathbb{E}_T \sum_{v \in L_T} d(v) 2^{-l(v)} \leq q$

We now wish to make a connection between $q(v)$ and $l(v)$ for vertices $v \in L_T$. First, we prove that following lemma:

Lemma 9. *For any $\boldsymbol{x}_1, ..., \boldsymbol{x}_k \in \{0,1\}^n$ there are coefficients $a_{i,j}, b_i \in \mathbb{F}_2$ s.t. for any quadratic function f we have:*

$$f(\boldsymbol{x}_1 + ... + \boldsymbol{x}_k) = \sum_{i,j} a_{i,j} f(\boldsymbol{x}_i + \boldsymbol{x}_j) + \sum_i b_i f(\boldsymbol{x}_i)$$

Proof. Let $f(\mathbf{x})$ by some polynomial of degree d. It's derivative in the $\mathbf{y}$ direction is defined to be $f_{\mathbf{y}}(\mathbf{x}) = f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x})$. It's easy to see that the degree of $f_{\mathbf{y}}$ as a function of $\mathbf{x}$ is at most $d-1$. So, taking 3 derivatives from a quadratic function makes it the zero function, and so in particular for any quadratic function f, we we take it's derivatives in directions $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$, and evaluate the result at 0, we get that

$$(((f_{\mathbf{x}})_{\mathbf{y}})_{\mathbf{z}}(0) = 0$$

Opening this expression yields:

$$f(\mathbf{x}+\mathbf{y}+\mathbf{z}) - f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}+\mathbf{z}) - f(\mathbf{y}+\mathbf{z}) + f(\mathbf{x}) + f(\mathbf{y}) + f(\mathbf{z}) - f(0) = 0$$

Since $f(0) = 0$, we can express $f(\mathbf{x}+\mathbf{y}+\mathbf{z})$ as a sum of application of f on an element, or sum of two elements in $\{x, \mathbf{y}, \mathbf{z}\}$. This proves the lemma for $k = 3$. For $k > 3$ we use simple induction. □

Now we can bound $l(v)$ in term of $q(v)$:

Lemma 10. *For any leaf $v \in L_T$, $l(v) \geq \phi(q(v))$*

Proof. Let $v_0, ..., v_{k-1}, v$ be the path from the root of T to v. Let $S \subset \{\mathbf{x}(v_0), ..., \mathbf{x}(v_{k-1})\}$ be a maximal quadratic independent set. $|S| = q(v)$. The linear rank of S is $l(v)$. Let $S' \subset S$ be a maximal set of linearly independent elements of S. $|S'| = l(v)$. Since every $x \in S$ is linearly dependent on S', it can be written as a sum of some of the elements of S'. Assume that $S' = \{x_1, ..., x_{l(v)}\}$. Using Lemma 9, we get that for any $x \in S$ there exists coefficients $a_{i,j}^{(x)}, b_i^{(x)} \in \mathbb{F}_2$ s.t for any quadratic function f:

$$f(\mathbf{x}) = \sum_{1 \leq i < j \leq l(v)} a_{i,j}^{(x)} f(\mathbf{x}_i + \mathbf{x}_j) + \sum_{1 \leq i \leq l(v)} b_i^{(x)} f(\mathbf{x}_i)$$

We have assumed that all the elements of S are quadratically independent. For this to hold, the above equations in the symbolic variables $f(\mathbf{x}_i + \mathbf{x}_j)$ and $f(\mathbf{x}_i)$ must be linearly independent. So the number of equations ($q(v)$) must be at most the number of variables ($\binom{l(v)}{2} + l(v)$). So, we get that:

$$q(v) = |S| \leq \binom{l(v)}{2} + l(v)$$

Reversing this formula, since $\phi(x)$ is monotone, we get that $l(v) \geq \phi(q(v))$.

□

We can now prove our main technical lemma (Lemma 3). We start with some technical lemmas. We define $\psi(x)$ to be $x - \phi(x)$ for $x \geq 1$, and 0 for $x < 1$. Notice that ψ is continuous, and $\psi(x) = x - \phi(x)$ for any non-negative integer x. Hence, using Lemma 10 we get that:

Lemma 11. *For any vertex v in a tree T, $q(v) - l(v) \leq \psi(q(v))$.*

Lemma 12. *ψ is increasing and convex.*

Proof. Since ψ is continuous and constant for $x \leq 1$, it's enough to prove the claim for $x > 1$ (for increasing it's clear, and once we've proved ψ is increasing, it shows it's enough to prove convexity for $x > 1$). We first show ψ is increasing.

For $x > 1$, define $y = \phi(x)$, so $x = y^2/2 + y/2$ and $\psi(y) = y^2/2 - y/2$.

$$\frac{d\psi}{dx} = \frac{d\psi}{dy}\frac{dy}{dx} = \frac{\frac{d\psi}{dy}}{\frac{dx}{dy}} = \frac{y - 1/2}{y + 1/2}$$

If $x > 1$ then $y = \phi(x) > 1$, hence $\frac{d\psi}{dx} > 0$ for $x > 1$, and so ψ is increasing.

To show that ψ in convex,

$$\frac{d^2\psi}{dx^2} = \frac{d\left(\frac{y-1/2}{y+1/2}\right)}{dy}\frac{dy}{dx} = \frac{1}{(y + 1/2)^3} > 0$$

□

We are now finally ready to prove Lemma 3.

Proof. (of Lemma 3) We need to prove that any test $\mathbb{T}$ with completeness c and average query complexity $q \geq 1$ accepts a random quadratic function with probability at least $c - 1 + 2^{-\psi(q)}$. Let us mark the probability the test accepts a random quadratic function by p. Let p_T mark the probability that a tree T accepts a random quadratic function. p_T is at least the probability that a random quadratic function reaches a leaf in L_T which is labeled *accept*. So:

$$p_T \geq \sum_{v \in L_T} c(v) 2^{-q(v)}$$

We now follow to analyze $p = \mathbb{E}_T[p_T]$.

$$p \geq \mathbb{E}_T[\sum_{v \in L_T} c(v) 2^{-q(v)}] = \mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} c(v) 2^{-q(v)+l(v)}]$$

.

We divide the sum in the right side into two parts, $p_0 - p_1$, with $p_0, p_1 \geq 0$, where:

$$p_0 = \mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} 2^{-q(v)+l(v)}]$$

. and

$$p_1 = \mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} (1 - c(v)) 2^{-q(v)+l(v)}]$$

.

We start by analyzing p_1. Since for any v always $q(v) \geq l(v)$ we have:

$$p_1 \leq \mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)}(1 - c(v))]$$

Recall that by Lemma 7 for any tree T we have

$$\sum_{v \in L_T} 2^{-l(v)} = 1$$

and by Lemma 6 we have

$$\mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} c(v)] \geq c$$

so we conclude that:

$$p_1 \leq 1 - c$$

We move to analyze p_0. Since $\mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)}] = 1$ and since the function $X \to 2^X$ is concave, we have by Jensen's inequality that:

$$p_0 \geq 2^{\mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)}(-q(v) + l(v))]}$$

Now, we have that $q(v) - l(v) \leq \psi(q(v))$ by Lemma 12, and also by the same lemma, since $q(v) \leq d(v)$, we get $\psi(q(v)) \leq \psi(d(v))$. So we get:

$$\mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)}(q(v) - l(v))] \leq \mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} \psi(d(v))]$$

.

Since by Lemma 12 ψ is convex, we get that again by Jensen's inequality we get that this is at most $\psi(\mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} d(v)])$. By Lemma 8

$$\mathbb{E}_T[\sum_{v \in L_T} 2^{-l(v)} d(v)] \leq q$$

where q is the average query complexity of $\mathbb{T}$. So, we conclude that $p_0 \geq 2^{-\psi(q)}$, and in total

$$p \geq p_0 - p_1 \geq 2^{-\psi(q)} + c - 1$$

□

5 Random quadratic function is far from linear

In this section we prove Lemma 4, i.e. that a random quadratic function is far from linear. We will use commonly known facts about quadratic functions.

Any quadratic function can be written as:

$$f(\mathbf{x}) = \mathbf{x}^t A\mathbf{x} + < \mathbf{x}, b >$$

The correlation of f with some linear function g is the g-th Fourier coefficient of f. The Fourier coefficients of quadratic functions are well studied. In particular, it is known that all the Fourier coefficients of f have the same absolute value, and that the number of non-zero Fourier coefficients is $2^{rank(A+A^t)}$. So, in order to show that f has no large correlation with some linear function, it's enough to show that $B = A + A^t$ has high rank. In particular, in order to show that f is $2^{-\Omega(n)}$-far from linear functions, we need to show that B has rank $\Omega(n)$. We will show that the probability that a random quadratic function has rank less than $n/4$ is $2^{-\Omega(n)}$. We will use the following lemma:

Lemma 13. *The number of matrices of rank at most k is at most $n^k 2^{nk}$.*

Using Lemma 13, it's easy to prove Lemma 4. The number of matrices of rank at most $n/4$ is at most $2^{n^2/4(1+o(1))}$. For a random quadratic function, B is a random symmetric matrix with zero diagonal, and so the probability that B has rank less than $n/4$ is $2^{-n^2/4(1+o(1))} = 2^{-\Omega(n)}$.

Now we finish by proving Lemma 13.

Proof. Let B be a matrix of rank at most k. There are $\binom{n}{k}$ options to choose k rows which span the row span of the matrix, each other row have at most 2^k options since it must be in the row span of k specific rows. So, the number of possibilities for rank k matrices is at most:

$$\binom{n}{k}(2^k)^{n-k} \leq n^k 2^{nk}$$

□

Acknowledgement. I thank my supervisor, Omer Reingold, for useful comments and for his constant support and interest in the work. I thank Alex Samorodnitsky for helpful discussions.

References

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and hardness of approximation problems. *Journal of the ACM*, 45(3):501555, 1998. Pre- liminary version in Proc. of FOCS '92.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70122, 1998. Preliminary version in Proc. of FOCS '92.

[BCH+96] M. Bellare, D. Coppersmith, J. Hastad, M. Kiwi, and M. Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):17811795, 1996.

[BLR93] M.Blum, M. Luby and R. Rubinfeld. Self testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549-595, 1993. Preliminary version in Proc. of STOC '90.

[ST00] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proceedings of the 32nd ACM symposium on Theory of Computation*, pages 191-199, 2000.

[ST06] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proceedings of the 38nd ACM symposium on Theory of Computation*, pages 11-20, 2006.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/