Linearity Testing in Characteristic Two

M. BELLARE * D. COPPERSMITH † J. HÅSTAD ‡ M. KIWI § M. SUDAN ¶

Abstract

Let $\mathsf{Dist}(f,g) = \Pr_u[f(u) \neq g(u)]$ denote the relative distance between functions f, g mapping from a group G to a group H, and let $\mathsf{Dist}(f)$ denote the minimum, over all linear functions (homomorphisms) g, of $\mathsf{Dist}(f,g)$. Given a function $f\colon G \to H$ we let $\mathsf{Err}(f) = \Pr_{u,v}[f(u)+f(v) \neq f(u+v)]$ denote the rejection probability of the BLR (Blum-Luby-Rubinfeld) linearity test. *Linearity testing* is the study of the relationship between $\mathsf{Err}(f)$ and $\mathsf{Dist}(f)$, and in particular the study of lower bounds on $\mathsf{Err}(f)$ in terms of $\mathsf{Dist}(f)$.

The case we are interested in is when the underlying groups are $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$. The corresponding test is used in the construction of efficient PCPs and thence in the derivation of hardness of approximation results, and, in this context, improved analyses translate into better non-approximability results. However, while several analyses of the relation of $\mathsf{Err}(f)$ to $\mathsf{Dist}(f)$ are known, none is tight.

We present a description of the relationship between $\mathsf{Err}(f)$ and $\mathsf{Dist}(f)$ which is nearly complete in all its aspects, and entirely complete (*i.e.* tight) in some. In particular we present functions $L, U\colon [0,1] \to [0,1]$ such that for all $x \in [0,1]$ we have $L(x) \leq \mathsf{Err}(f) \leq U(x)$ whenever $\mathsf{Dist}(f) = x$, with the upper bound being tight on the whole range, and the lower bound tight on a large part of the range and close on the rest.

Part of our strengthening is obtained by showing a new connection between the linearity testing problem and Fourier analysis, a connection which may be of independent interest. Our results are used by Bellare, Goldreich and Sudan to present the best known hardness results for Max3SAT and other MaxSNP problems [7].

* Department of Computer Science & Engineering, Mail Code 0114, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093. mihir@cs.ucsd.edu. This work was done while the author was at the IBM T. J. Watson Research Center.

† Research Division, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA. copper@watson.ibm.com.

‡ Department of Computer Science, Royal Institute of Technology, 10044 Stockholm, Sweden. johanh@nada.kth.se. Part of this work was done while the author was visiting MIT.

§ Dept. of Applied Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139. mkiwi@math.mit.edu. Supported by an AT&T Bell Laboratories PhD Scholarship and NSF Grant CCR–9503322. On leave of absence from Dept. de Ingeniería Matemática, U. de Chile.

¶ Research Division, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598, USA. madhu@watson.ibm.com.

1 Introduction

Linearity testing (and its extension to low degree testing) has come to the fore in the last few years principally due to its crucial role in the construction of efficient PCPs, and thence in the obtaining of (strong) non-approximability results for NP-optimization problems. Yet the problem itself is older, with the basic formulation as we now know it first made in the context of program checking [9]. It also has wider applicability, for example in the testing of linear error-correcting codes.

It is a feature of the area that while tests are easy to specify, they are notoriously hard to analyze, especially to analyze well. Yet, good analyses are, for several reasons, worth striving for. There is, first, the inherent mathematical interest of getting the best possible analysis and understanding of a well-defined combinatorial problem. But, there is a more pragmatic reason: better analyses typically translate into improved (increased) factors shown non-approximable in hardness of approximation results.

The particular problem in linearity testing that we address is a case in point. The BLR (Blum-Luby-Rubinfeld) test is the first ever proposed, and addresses the most basic question, namely testing linearity (as opposed, say, to low-degree). Our focus is the case of most importance in applications, when the underlying function maps between groups of characteristic two. Several analyses have appeared, yet none is tight. With each analysis comes an improved Max3SAT non-approximability factor, but the extent to which the factor can grow remains open. It is a goal of this paper to provide some answers to this question.

We will do this; but in fact do more. Let us begin by describing the problem and past work more precisely.

1.1 The Problem

Although our concern is groups of characteristic two, it will be useful, to discuss past work, to begin more generally, with the problem of linearity testing over arbitrary finite groups. Thus let G, H be finite groups, and recall that a function $g\colon G \to H$ is *linear* if $g(u) + g(v) = g(u+v)$ for all $u, v \in G$. (That is, g is a group homomorphism.) Here are some basic definitions:

- $\mathrm{LIN}(G,H)$ — Set of all linear functions of G to H
- $\mathsf{Dist}(f,g) \stackrel{\text{def}}{=} \Pr_{u \stackrel{R}{\leftarrow} G}[f(u) \neq g(u)]$ — (relative) distance between $f, g\colon G \to H$

– $\mathsf{Dist}(f) \stackrel{\text{def}}{=} \min\{\, \mathsf{Dist}(f,g) : g \in \textsc{Lin}(G,H) \,\}$ — Distance of f to its closest linear function.

We are given oracle access to a function f mapping G to H. (That is, we can specify $u \in G$ and in one step are returned $f(u) \in H$.) We want to test that f is close (in relative distance) to a linear function. We are charged for each oracle call.

THE BLR TEST. The BLR test is the following [9]— Pick $u, v \in G$ at random, query the oracle to obtain $f(u), f(v), f(u+v)$, and reject if $f(u) + f(v) \neq f(u+v)$. Let

$$\mathsf{Err}(f) \stackrel{\text{def}}{=} \Pr_{u,v \stackrel{R}{\leftarrow} G}[\, f(u) + f(v) \neq f(u+v) \,]$$

denote the probability that the BLR test rejects f. The issue in linearity testing is to study how $\mathsf{Err}(f)$ behaves as a function of $x = \mathsf{Dist}(f)$. In particular, one would like to derive good lower bounds on $\mathsf{Err}(f)$ as a function of x.

$\textsc{Rej}(\cdot)$. A convenient way to capture the above issues is via the *rejection probability function* $\textsc{Rej}_{G,H} : [0,1] \to [0,1]$ of the test. It associates to any number x the minimum value of $\mathsf{Err}(f)$, taken over all functions f of distance x from the space of linear functions. Thus, $\textsc{Rej}_{G,H}(x) \stackrel{\text{def}}{=}$

$$\min\{\, \mathsf{Err}(f) : f : G \to H \text{ s.t. } \mathsf{Dist}(f) = x \,\}.$$

The graph of $\textsc{Rej}_{G,H}$ —namely $\textsc{Rej}_{G,H}(x)$ plotted as a function of x— is called the *linearity testing curve*. This curve depends only on the groups G, H.

Thus the most general problem in linearity testing is to determine the function $\textsc{Rej}_{G,H}(\cdot)$ for given G, H. Much of the work that has been done provides information about various aspects of this function.

THE KNEE OF THE CURVE. In particular, one parameter has emerged as an important one in connection with MaxSNP hardness results. This parameter, identified in [2, 6, 7, 8], is a single number, which we call here the *knee* of the curve. It is defined as the minimum rejection probability when the distance (of the function being tested from the space of linear functions) is at least $1/4$:

$$\textsc{Knee}_{G,H} \stackrel{\text{def}}{=} \min\{\, \textsc{Rej}(x) : x \geq 1/4 \,\}\ .$$

Improvements (increases) in the lower bound that can be shown on $\textsc{Knee}_{G,H}$ translate directly into improved (increased) non-approximability factors for MaxSNP problems via [6, 7, 8]. (Exactly how or why this is the case is outside the scope of this paper, and we refer the reader to the works in question.)

1.2 Previous work

The first investigation of the shape of the linearity testing curve, by Blum, Luby and Rubinfeld [9], was in the general context where G, H are arbitrary finite groups. Their analysis showed that $\textsc{Rej}_{G,H}(x) \geq 2x/9$ [9]. (They indicate that this is an improvement of their original analysis obtained jointly with Coppersmith.) Interest in the tightness of the analysis begins with Bellare, Goldwasser, Lund and Russell [6] in the context of improving non-approximability factors for MaxSNP problems. They showed that $\textsc{Rej}_{G,H}(x) \geq 3x - 6x^2$. Meanwhile it was noted that the result of [9] could be used to show that $\textsc{Rej}_{G,H}(x) \geq 2/9$ for $x \geq 1/4$. The last two bounds supersede the first, so that the following theorem captures the state of knowledge.

Theorem 1.1 [6, 9, 10] Let G, H be arbitrary finite groups. Then:

(1) $\textsc{Rej}_{G,H}(x) \geq 3x - 6x^2$.

(2) $\textsc{Knee}_{G,H} \geq 2/9$.

As indicated above, an improved lower bound for the knee would lead to better non-approximability results. But in this general setting, we can do no better: an example of Coppersmith shows that the above value is in fact tight in the case of general groups. (For completeness this example is provided in Appendix A.) This leads into our research. We note that the problem to which linearity testing is applied in the proof system constructions of [2, 6, 7, 8] is that of testing Hadamard codes (in the first three works) and the long code (in the last work). But this corresponds to the above problem in the special case where $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$. For this case, the example of Coppersmith does *not* apply, and we can hope for better results.

1.3 New results and techniques

We look at the performance of the BLR test when the underlying groups are $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$ for some $n \geq 1$. (G is regarded as an additive group in the obvious way. Namely, the elements are viewed as n-bit strings or vectors over $\mathrm{GF}(2)$, and operations are component-wise over $\mathrm{GF}(2)$.) For notational simplicity we now drop the groups G, H from the subscripts, writing $\textsc{Rej}(x)$ and $\textsc{Knee}$— it is to be understood that we mean $G = \mathrm{GF}(2)^n$ and $H = \mathrm{GF}(2)$. We provide two new analyses of $\textsc{Rej}(x)$.

FOURIER ANALYSIS. We establish a new connection between linearity testing and Fourier analysis. We provide an interpretation of $\mathsf{Dist}(f)$ and $\mathsf{Err}(f)$ in terms of the Fourier coefficients of an appropriate transformation of f. We use this to cast the linearity testing problem in the language of Fourier series. This enables us to use Fourier analysis to study the BLR test. The outcome is the following:

Theorem 1.2 $\textsc{Rej}(x) \geq x$.

Apart from lending a new perspective to the linearity testing problem, the result exhibits a feature which distinguishes it from all previous results. Namely, it shows that $\textsc{Rej}(x)$ increases

with x and in fact is $1/2$ at $x = 1/2$.[1] (According to the previous analysis, namely Theorem 1.1, $\text{REJ}(x)$ may have been bounded above by $2/9$ for all $x \geq \alpha$, where α is the larger root of the equation $3z - 6z^2 = 2/9$.) Furthermore we can show that the analysis is tight (to within $o(1)$ factors) at $x = 1/2 - o(1)$.

This result can also be combined with Part (1) of Theorem 1.1 to show that $\text{KNEE} \geq 1/3$. However this is not tight. So we focus next on finding the right value of the knee.

COMBINATORIAL ANALYSIS. The analysis to find the knee is based on combinatorial techniques. It leads us to an isoperimetric problem about a 3-regular hypergraph on the vertices of the n-dimensional hypercube. We state and prove a Summation lemma which provides a tight isoperimetric inequality for this problem. We then use it to provide the following tight bound on the knee of $\text{REJ}(x)$.

Theorem 1.3 $\text{KNEE} = 45/128$.

As the statement indicates we have an equality, not a lower bound— the value of the knee above is tight. This means we have the best possible value from the point of view of applications to MaxSNP hardness. See Section 1.4.

TIGHTNESS OF THE ANALYSIS. We provide examples to indicate that, besides the knee value, the lower bounds on $\text{REJ}(x)$ as indicated by our and previous results are tight for a number of points. In particular, the curve is tight for $x \leq 5/16$, and the bound at $x = 1/2 - o(1)$ is matched up to within $o(1)$ factors (i.e., there exist functions $f_n : \text{GF}(2)^n \to \text{GF}(2)$ such that as n goes to ∞, $\mathsf{Err}(f_n)$ and $\mathsf{Dist}(f_n)$ go to $1/2$).

OTHER RESULTS. The isoperimetric inequality underlying Theorem 1.3 turns out to reveal other facts about $\text{REJ}(x)$ as well. In particular it helps establish a tight *upper bound* on $\mathsf{Err}(f)$ as a function of $\mathsf{Dist}(f)$. This result is presented in Section 3.

Also, while the main focus of this paper has been the BLR test, we also present in Appendix B a more general result about testing for total degree one in characteristic two. The purpose is to further illustrate the strength and elegance of the Fourier analysis technique, as well as its more general applicability to the problem of analyzing program testers.

GRAPH. Figure 1 summarizes the results of this work. The points $\{ (\mathsf{Dist}(f), \mathsf{Err}(f)) : f \}$ lie in the white region of the first graph. The dark shaded region represents the forbidden area before our work, and the lighter shaded region represents what we add to the forbidden area. Note we both extend the lower bound and provide upper bounds. The dots are actual computer constructed examples; they indicate that perhaps the lower bound may be improved, but not by much.[2] Of course, the knee value is tight. Furthermore the upper bound is tight.

The second graph indicates lower bounds on $\text{REJ}(x)$. The parabola is the curve $3x - 6x^2$ representing the result of [6], and the line $2x/9$ represents the result of [9]. The earlier value of the knee appears as the horizontal line at $2/9$. Our additions are the 45 degree line of x and the horizontal line at $45/128$ for the new knee value.

1.4 Application to MaxSNP hardness

Usage of the linearity test in the construction of efficient PCPs, and thence in the derivation of hardness of approximability results for Max-SNP problems, begins in [2] and continues in [6, 7, 8]. In the first three cases, it is used to test the Hadamard code; in the last case, to test the long code. In all cases the underlying problem is the one we have considered above, namely linearity testing with $G = \text{GF}(2)^n$ and $H = \text{GF}(2)$.

The Max-SNP hardness result of [6] used only two things: The lower bound $\text{REJ}(x) \geq 3x - 6x^2$ of Theorem 1.1, and the best available lower bound k on the knee. They were able to express the non-approximability factor for Max-3SAT as an increasing function $g_1(k)$ depending solely on k. Since the only available lower bound on the knee at that time was the $\text{KNEE} \geq 2/9$ of Theorem 1.1, this was the value they used. Their final result was that approximating Max-3SAT within $113/112 \approx 1.009$ is NP-hard.

Improved proof systems were built by [8]. Again, their non-approximability factor had the form $g_2(k)$ for some function g_2 depending only on the best available lower bound k on the knee. They also used $\text{KNEE} \geq 2/9$ to show that approximating Max-3SAT within $74/73 \approx 1.014$ is NP-hard.

Theorem 1.3 would yield direct improvements to the results of [6, 8] with no change in the underlying proof systems or construction. However, better proof systems are now known, namely the ones of [7]. Again, the analysis depends on the best available lower bound on the knee, so that usage of Theorem 1.3 yields a better result than would have been obtained using only Theorem 1.1, and this aspect is now tight. But, interestingly, [7] was also able to exploit Theorem 1.2. Their final conclusion, which uses both our results, was that approximating Max-3SAT within $38/37 \approx 1.027$ is NP-hard. (Using only Theorem 1.1 it would have been $45/44 \approx 1.023$. Using Theorem 1.3 but not Theorem 1.2 it would have been $39/38 \approx 1.026$.)

1.5 Relationship to other work

There are a variety of problems which are studied under the label of *(low-degree) testing*. Furthermore, low-degree tests are used in a variety of ways in proof systems. We briefly explain, first, what are the other problems and results in low degree

[1] Note that $\mathsf{Dist}(f) \leq 1/2$ for all $f\colon G \to H$ because we are working over $\text{GF}(2)$, so only the portion $x \in [0, 1/2]$ of the curve is interesting.

[2] More precisely, we have a randomized procedure that with high probability can construct, for each plotted point, a function f such that $(\mathsf{Dist}(f), \mathsf{Err}(f))$ is arbitrarily close to the point in question.

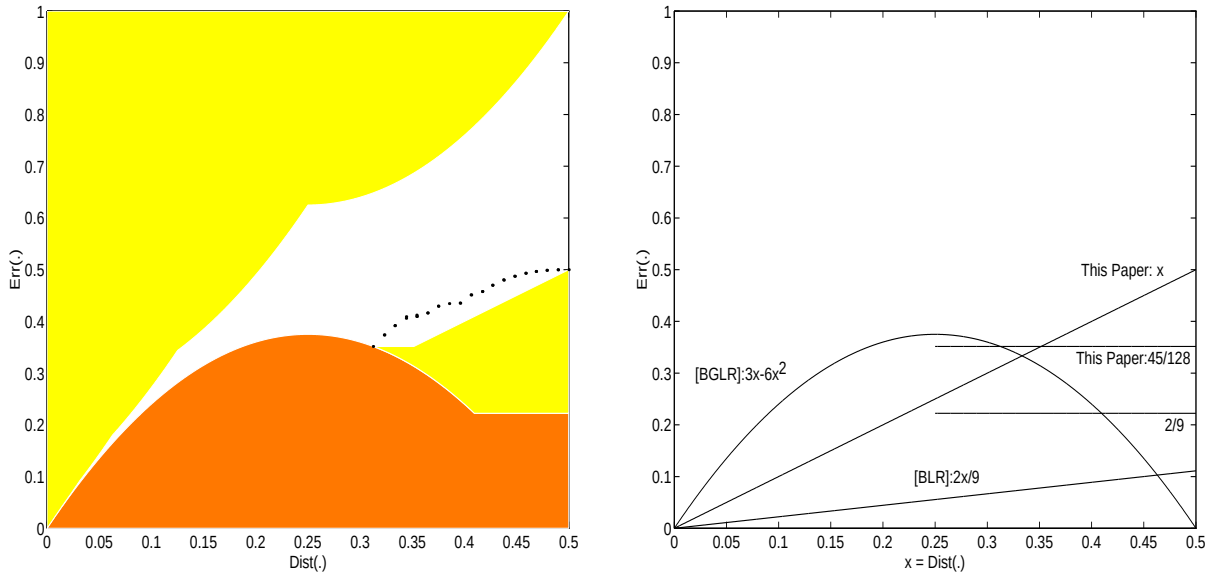

Figure 1: The points $(\mathsf{Dist}(f), \mathsf{Err}(f))$ in the plane, and the successive lower bounds. See text for discussion.

testing and why they differ from ours; second how the usage of these in proof systems is different from the usage of linearity tests.

LOW DEGREE TESTING. We are given an oracle for a function $f\colon F^n \to F$, where F is a field, and we are given a positive integer d. In the low *individual degree* testing problem we are asked to determine whether f is close to some polynomial p of degree d in each of its n variables. When specialized to the case of $d = 1$, this task is referred to as *multi-linearity testing*. In the low *total degree* testing problem we are asked to determine whether f is close to some polynomial p of total degree d in its n variables. Multi-linearity tests were studied by [4, 11]. Low individual degree tests were studied by [3, 5, 12, 16]. Total degree tests were studied by [2, 13, 14, 17].

What we are looking at, namely linearity testing over $\mathrm{GF}(2)$, is a variant of the total degree testing problem in which the degree is $d = 1$, F is set to $\mathrm{GF}(2)$, and the constant term of the polynomial p is forced to 0.[3] Even though a significant amount of work has been put into the analysis of the low degree tests by the above mentioned works, the analysis does not appear to be tight for any case. In particular one cannot use those results to derive the results we obtain here. In fact the tightness of the result obtained here raises the issue as to whether similar techniques can be used to improve the analysis in the above testers.

THE ROLE OF TESTING IN PROOF SYSTEMS. To explain this, first recall that proof systems are built by *recursion* [3]. Each level of recursion will typically use some form of low-degree testing, the kind differing from level to level.

The use of multi-linearity testing was initiated by Babai, Fortnow and Lund [4]. For efficiency reasons, researchers beginning with Babai, Fortnow, Levin and Szegedy [5] then turned to low individual degree testing. This testing is used in the "higher" levels of the recursion. Linearity testing showed up for the first time in the lowest level of the recursion, in the checking of the Hadamard code in [2]. The proof systems we discuss use all these different testers, but, as we explained, the final non-approximability factors obtained can be expressed only in terms of the knee of the linearity testing curve.

1.6 Discussion

The main argument behind the analysis of the BLR test given in [9] is the following: given f taking values from one finite group into another finite group, start by defining a function g whose value at u is $\textsc{Majority}_v\{f(u+v) - f(v)\}$. Then, show that if $\mathsf{Err}(f)$ is sufficiently small, three things happen. First, an overwhelming majority of the values $\{f(u+v) - f(v)\}_v$ agree with $g(u)$, second, g is linear, and last, g is close to f. This argument is *constructive*, since it explicitly builds a function to which f is shown to be close.

The arguments used in all previous works on low-degree testing have been constructive. So far, constructive proof arguments have been unable to show a non-trivial relation between the probability that a given function f fails a test, and the distance from f to any family of low-degree polynomials, when

[3] To illustrate the difference between individual and total degree, note that $f(x_1, \ldots, x_n) = x_1 x_2$ is multi-linear but not linear.

the probability that the test fails is high (*i.e.* larger than $1/2$). Our discrete Fourier analysis approach does not exhibit the constructive properties discussed above, and may be one of the reasons for its success. Further exploration of non-constructive techniques seems to be worth undertaking.

2 Fourier Analysis of the Linearity Test

In this section we prove Theorem 1.2 and discuss how tight it is.

The main result of this section is based on the following observation: If we view f as a real valued function, and let h be the function that at u takes the value $(-1)^{f(u)}$, then, if the distance from f to the nearest linear function is large, the Fourier coefficients of h cannot be very large. Furthermore, the smaller the Fourier coefficients of h are, the higher the probability that f will fail the linearity test.

In the rest of this section, we first review the basic tools of discrete Fourier analysis that we use, and then give a precise formulation of the argument discussed above.

DISCRETE FOURIER TRANSFORM. Consider the family of all real-valued functions on F^n.[4] This collection of functions is a 2^n-dimensional real vector space with the following inner product: $\langle \phi, \theta \rangle = (\sum_{u \in F^n} \phi(u)\theta(u))/|F|^n$.

When one studies a linear space of functions defined on a group, choosing a special basis for the linear space might be very useful. This special basis is the *characters* of the group at hand. In our case the group is F^n. Thus, we chose as basis of our linear space the basis $\{\psi_\alpha\}_{\alpha \in F^n}$, where $\psi_\alpha(u) = (-1)^{\alpha \cdot u}$, and, $\alpha \cdot u = \sum_{i=1}^n \alpha_i u_i$. It can be easily verified that the family $\{\psi_\alpha\}_{\alpha \in F^n}$ forms an orthonormal basis. It follows that any real-valued function ϕ over F^n can be uniquely expressed as a linear combination of the ψ_α's, namely, $\phi = \sum_{\alpha \in F^n} \widehat{\phi}_\alpha \psi_\alpha$. The coefficient $\widehat{\phi}_\alpha$ is referred to as the α-th Fourier coefficient of ϕ. By the ortho-normality property of our chosen basis, we have that $\widehat{\phi}_\alpha = \langle \phi, \psi_\alpha \rangle$. The ortho-normality of the basis also implies Parseval's identity: $\langle \phi, \phi \rangle = \sum_{\alpha \in F^n} (\widehat{\phi}_\alpha)^2$.

The *convolution* of two functions ϕ and θ is denoted by $\phi * \theta$ and defined as follows: $(\phi * \theta)(x) = (\sum_{u+v=x} \phi(u)\theta(v))/|F|^n$. Note, that over the vector space of real-valued functions on F^n the convolution operator is associative, commutative, and distributive with respect to addition.

The following *convolution identity* shows the relationship between the Fourier coefficients of two functions ϕ, θ, and the Fourier coefficients of their convolution: $\widehat{(\phi * \theta)}_\alpha = \widehat{\phi}_\alpha \widehat{\theta}_\alpha$.

LOWER BOUND. To lower bound $\mathsf{Err}(f)$ we use discrete Fourier analysis techniques. We start by establishing a relation between the Fourier coefficients of the function $(-1)^{f(\cdot)}$,[5] and the distance from f to the nearest linear function. More precisely, we show that if the distance from f to the nearest linear function is large the Fourier coefficients of $(-1)^{f(\cdot)}$ are small.

Lemma 2.1 Suppose $f\colon F^n \to F$ and $\alpha \in F^n$. Let $h(\cdot) = (-1)^{f(\cdot)}$. Then $\widehat{h}_\alpha \leq 1 - 2\,\mathsf{Dist}(f)$.

Proof: Let $l_\alpha(u) = \sum_{i=1}^n \alpha_i u_i$. Clearly, $l_\alpha \in$ LIN. Moreover, viewing f and l_α as real valued functions, we have that

$$\begin{aligned} \widehat{h}_\alpha &= \tfrac{1}{|F|} \cdot \textstyle\sum_u (-1)^{f(u)+l_\alpha(u)} \\ &= \Pr_u\left[f(u)=l_\alpha(u)\right] - \Pr_u\left[f(u)\neq l_\alpha(u)\right] \\ &= 1 - 2\,\mathsf{Dist}(f, l_\alpha) \\ &\leq 1 - 2\,\mathsf{Dist}(f). \quad \blacksquare \end{aligned}$$

We will now establish Theorem 1.2.

Proof of Theorem 1.2: Let $f\colon F^n \to F$ be such that $\mathsf{Dist}(f) = x$. Note that if we let $h(\cdot) = (-1)^{f(\cdot)}$, then $(1 - h(u)h(v)h(u+v))/2$ equals 1 if $f(u)+f(v) \neq f(u+v)$, and 0 otherwise. This leads to the following key observation:

$$\mathsf{Err}(f) = \tfrac{1}{2}\left(1 - (h * h * h)(0)\right) .$$

Thus, from the definition of Fourier coefficients and the convolution identity, it follows that:

$$\begin{aligned} \mathsf{Err}(f) &= \tfrac{1}{2}\left(1 - \textstyle\sum_\alpha \widehat{(h * h * h)}_\alpha\right) \\ &= \tfrac{1}{2}\left(1 - \textstyle\sum_\alpha (\widehat{h}_\alpha)^3\right) . \end{aligned}$$

The upper bound for $\widehat{h}_\alpha$ given in Lemma 2.1 and Parseval's identity imply that

$$\mathsf{Err}(f) \geq \tfrac{1}{2}\left(1 - (1-2x)\textstyle\sum_\alpha (\widehat{h}_\alpha)^2\right) = x ,$$

as desired. $\blacksquare$

The next lemma complements Theorem 1.2. To state it we first define the *slack* between functions f and l by

$$\mathsf{sl}(f, l) \stackrel{\text{def}}{=} \Pr_{u,v}\left[f(u)\neq l(u), f(v)\neq l(v), f(u+v)\neq l(u+v)\right] .$$

Lemma 2.2 For all $f\colon F^n \to F$ and all l in LIN,

$$\mathsf{Err}(f) = 3\,\mathsf{Dist}(f, l) - 6\,\mathsf{Dist}(f, l)^2 + 4\,\mathsf{sl}(f, l) .$$

Proof: First observe that $f(u)+f(v)$ and $f(u+v)$ are distinct if and only if f differs from l in exactly one of the points

[4] In the rest of this work, unless explicitly said otherwise, F denotes $\mathrm{GF}(2)$. Furthermore, whenever we write LIN it is to be understood that we are referring to $\mathrm{LIN}(F^n, F)$.

[5] In this section, if the function $f(\cdot)$ appears as an exponent it is to be understood as a real valued function.

$\{u, v, u{+}v\}$ or in all of the points $\{u, v, u+v\}$. Thus $\mathsf{Err}(f)$ equals

$$3\,\mathrm{Pr}_{u,v}\,[\,f(u){\neq}l(u), f(v){=}l(v), f(u{+}v){=}l(u{+}v)\,] + \mathrm{Pr}_{u,v}\,[\,f(u){\neq}l(u), f(v){\neq}l(v), f(u{+}v){\neq}l(u{+}v)\,]\ .$$

But this equals

$$4\,\mathrm{Pr}_{u,v}\,[\,f(u){\neq}l(u), f(v){\neq}l(v), f(u{+}v){\neq}l(u{+}v)\,] + 3\,\mathrm{Pr}_{u,v}\,[\,f(u){\neq}l(u), f(v){=}l(v)\,] - 3\,\mathrm{Pr}_{u,v}\,[\,f(u){\neq}l(u), f(v){\neq}l(v)\,]\ .$$

Observing that the events $\{\ (u,v) : f(u){=}l(u)\ \}$, and $\{\ (u,v) : f(v){=}l(v)\ \}$ are independent, and performing a simple algebraic manipulation, suffices to conclude the proof of the lemma. ∎

TIGHTNESS DISCUSSION. We now discuss how tight the results of this section are. Throughout the rest of this discussion let $x \in [0,1]$ be such that $x\,|F|^n$ is an integer.

If $x > 1/2$, then there is no function $f\colon F^n \to F$ such that $\mathsf{Dist}(f) = x$ (since the expected distance from a randomly chosen linear function to f is $1/2$).

If $x = 1/2$, and we randomly choose f so $f(u) = X_u$, where $\Pr[\,X_u = 1\,] = p$, $\Pr[\,X_u = 0\,] = 1-p$, and $p \in [1/2, 1]$, it follows from a Chernoff bound (see [1, Appendix A]) and Chebyschev's inequality (see [1, Ch. 4]) that with high probability $0 \le x - \mathsf{Dist}(f) \le o(1)$, and $|\mathsf{Err}(f) - \left(3\,p(1-p)^2 + p^3\right)| \le o(1)$, respectively. Thus, if $p = 1/2$, Theorem 1.2 is almost tight in the sense that $\mathrm{REJ}(x)$ is almost x.

If $x \le 5/16$, then Lemma 2.2 is tight, since there are functions f such that $\mathsf{Dist}(f) = x$ and $\mathsf{Err}(f) = 3\,x - 6\,x^2$. In fact, for u in F^n let $\lfloor u \rfloor_k \stackrel{\mathrm{def}}{=} u_1 \cdots u_k$. If $S = \{\ u \in F^n : \lfloor u \rfloor_4 \in \{1000, 0100, 0010, 0001, 1111\}\ \}$, then for any boolean function f which equals 1 in $x\,|F|^n$ elements of S, and 0 otherwise, it holds that $\mathsf{Dist}(f) = \mathsf{Dist}(f,0) = x$, and $\mathsf{sl}(f,0) = 0$, hence, $\mathsf{Err}(f) = 3\,x - 6\,x^2$. Thus, Lemma 2.2, is best possible for x in the interval $[0, 5/16]$.

Figure 1, gives evidence showing that Theorem 1.2 is close to being optimal for x in the interval $[5/16, 1/2]$. But, as the next two sections show, there is room for improvements.

3 The Summation Lemma

This section is devoted to proving a combinatorial result of independent interest, but necessary in the tighter analysis of the linearity test that we give in Section 4. We also apply this result to obtain a tight upper bound on the probability that the BLR test fails.

Loosely stated, we show that given three subsets A, B, C of F^n, the number of triplets (u, v, w) in $A \times B \times C$ such that $u + v + w = 0$, is maximized when A, B and C are the lexicographic smallest $|A|$, $|B|$ and $|C|$ elements of F^n respectively.[6]

The following lemma, independently proved by D. J. Kleitman [15], gives a precise statement of the above discussed fact.

For convenience we introduce the following notation:

$$\Phi(A,B,C) = \{\ (u,v,w) \in A{\times}B{\times}C : u{+}v{+}w = 0\ \}\ .$$

Also, for $S \subseteq F^n$ we let S^* denote the smallest, in lexicographic order, $|S|$ elements of F^n.

Lemma 3.1 [Summation Lemma] For any non-negative integers m_A, m_B and m_C,

$$\max |\Phi(A,B,C)| \ = \ |\Phi(A^*, B^*, C^*)|,$$

where the maximum is taken over all $A, B, C \subseteq F^n$ satisfying $|A| = m_A$, $|B| = m_B$, and $|C| = m_C$.

Proof: We proceed by induction. The case $n = 1$ is trivial. For the inductive step, consider i in $\{1, \ldots, n\}$ and b in F. Let $f_{i,b}$ be the function such that for $u = (u_j)_{j \neq i} \in F^{n-1}$, $(f_{i,b}(u))_j = u_j$ if $j{\neq}i$, and b otherwise, *i.e.* $f_{i,b}$ embeds F^{n-1} onto $\{\ u \in F^n : u_i = b\ \}$ in the natural way. For $S \subseteq F^n$, let $S_b^{(i)} = \{\ (u_j)_{j\neq i} \in F^{n-1} : f_{i,b}(u) \in S\ \}$, *i.e.* $S_b^{(i)}$ is the natural projection into F^{n-1} of the elements of S whose i-th coordinate is b. Furthermore, let

$$S^{(i)} = f_{i,0}\left((S_0^{(i)})^*\right) \bigcup f_{i,1}\left((S_1^{(i)})^*\right)\ .$$

Observe that $|S^{(i)}| = |S|$. Now, given $A, B, C \subseteq F^n$, maximizing $|\Phi(A,B,C)|$, we see that $|\Phi(A,B,C)|$ equals

$$|\Phi(A_0^{(i)}, B_0^{(i)}, C_0^{(i)})| + |\Phi(A_1^{(i)}, B_1^{(i)}, C_0^{(i)})| + |\Phi(A_1^{(i)}, B_0^{(i)}, C_1^{(i)})| + |\Phi(A_0^{(i)}, B_1^{(i)}, C_1^{(i)})|.$$

Applying the inductive hypothesis four times we have that $|\Phi(A,B,C)| \le |\Phi(A^{(i)}, B^{(i)}, C^{(i)})|$. Abusing notation, we let $u \in F^n$ represent the integer with binary expansion u. Then, $A^{(i)} \not\equiv A$, or $B^{(i)} \not\equiv B$, or $C^{(i)} \not\equiv C$, implies that $\sum_{u\in A} u + \sum_{u\in B} u + \sum_{u\in C} u > \sum_{u\in A^{(i)}} u + \sum_{u\in B^{(i)}} u + \sum_{u\in C^{(i)}} u$. Thus, without loss of generality, we can assume that for all i, $A^{(i)} \equiv A$, $B^{(i)} \equiv B$ and $C^{(i)} \equiv C$.

One would like to conclude the proof of the lemma by claiming that, if for all i, $A^{(i)} \equiv A$, $B^{(i)} \equiv B$ and $C^{(i)} \equiv C$, then A, B and C are equal to A^*, B^* and C^* respectively. The latter claim is 'almost' true, in the sense that, if S is a set such that for all i, $S^{(i)} \equiv S$, then, either $S \equiv S^*$, or $S = \{\ u : u_1 = 0$ or $u = 10\cdots0\ \} \setminus \{01\cdots1\}$. The lemma follows by case analysis (omitted). ∎

[6] The lexicographic order in F^n is the total order relation $\le$, such that, $u \le v$ if and only if $\sum_i u_i 2^{-i} \le \sum_i v_i 2^{-i}$ (arithmetic over the reals).

By definition, a subspace V of F^n is such that for every u and v in V, $u+v$ is also in V. This motivates using

$$\tfrac{1}{|S|^2}|\Phi(S,S,S)| \ ,$$

as a measure of how *close* the set $S \subseteq F^n$ is to being a subspace. The larger this quantity is, the closer the set S is off being a subspace. From this point of view, the Summation Lemma implies that the collection of the lexicographic smallest m elements of F^n is the subset of F^n (of cardinality m) that more closely resembles a subspace.

Observing that the slack between f and l is proportional to how close the set $\{\ u : f(u) \neq l(u)\ \}$ is to being a subspace, we obtain the following:

Lemma 3.2 Suppose $f\colon F^n \to F$. Let $x = \mathsf{Dist}(f)$. Let k be the unique integer such that $2^{-k} \leq x < 2^{-k+1}$, and let $\delta = 2^{-k}$. Then

$$\mathsf{Err}(f) \leq 3x - 6x^2 + 4\delta^2 + 12(x-\delta)^2 \ .$$

Proof: Let l be the closest linear function to f, and let $S = \{u : f(u) \neq l(u)\}$. Denote $|\Phi(A,B,C)|/|F|^{2n}$ by $\varphi(A,B,C)$. Then

$$\mathsf{sl}(f,l) = \varphi(S,S,S) \leq \varphi(S^*,S^*,S^*) \ ,$$

where the inequality follows from the Summation Lemma. Now, let V be the smallest, in lexicographic order, $\delta\,|F|^n$ elements of F^n. In particular, V is a subspace. Moreover, $|S^*| = |S| = x\,|F|^n$. Thus

$$\begin{aligned}\varphi(S^*,S^*,S^*) &= \varphi(V,V,V) + 3\,\varphi(S^* \setminus V, S^* \setminus V, V) \\ &= \delta^2 + 3\,(x-\delta)^2 \ .\end{aligned}$$

The claim follows from Lemma 2.2. ∎

Lemma 3.2 is best possible. Indeed, let $x \in [0, 1/2]$, and let $S \subseteq F^n$ be the set of the lexicographic smallest $x\,|F|^n$ elements of F^n. Then, the function f that evaluates to 1 at every $u \in S$, and to 0 elsewhere, is such that $x = \mathsf{Dist}(f)$, and $\mathsf{Err}(f)$ meets the upper bound of Lemma 3.2.

4 Combinatorial analysis of the linearity test

We now prove Theorem 1.3. The tightness discussion of Section 2 already implies that $\textsc{Knee} \leq 45/128$. The main argument used to show that $\textsc{Knee} \geq 45/128$ is the following: Given a function $f\colon F^n \to F$ define a function g_f, from F^n to F, whose value at u is $\textsc{Majority}_v\{f(u+v) - f(v)\}$. Then if $\mathsf{Err}(f)$ is sufficiently small three things occur: (i) An overwhelming majority of the values $\{f(u+v)-f(v)\}_v$ agree with $g_f(u)$, (ii) g_f is linear, (iii) g_f is close to f. This argument was first used in [9] while studying linearity testing over finite groups. We will show how this argument can be tightened in the case of linearity testing over fields of characteristic two.

More precisely, the proof of Theorem 1.3 is a consequence of the following three lemmas:

Lemma 4.1 [9] For all $f\colon F^n \to F$, if g_f is linear, then $\mathsf{Err}(f) \geq \mathsf{Dist}(f, g_f)/2$.

Lemma 4.2 For all $f\colon F^n \to F$, if g_f is linear, then $\mathsf{Err}(f) \geq 2\,\mathsf{Dist}(f,g_f)\cdot[1 - \mathsf{Dist}(f,g_f)]$.

Lemma 4.3 For all $f\colon F^n \to F$, if $\mathsf{Err}(f) < 45/128$, then g_f is linear.

We first show that Theorem 1.3 follows from the above stated results. Assume $\textsc{Knee} < 45/128$, then, there is a function $f\colon F^n \to F$, such that $\mathsf{Err}(f) < 45/128$ and $x = \mathsf{Dist}(f) \geq 1/4$. By Lemma 2.2, $\mathsf{Err}(f) \geq 3x - 6x^2$, thence we need only consider the case in which x is at least $5/16$. Moreover, by Lemma 4.3, g_f is a linear function. Thus, $\mathsf{Dist}(f,g_f) \geq x \geq 5/16$, which together with Lemmas 4.1 and 4.2 imply that $\mathsf{Err}(f) \geq \min_{x\in[5/16,1]} \max\{x/2, 2(1-x)x\} = 3/8$, a contradiction.

The rest of this section is dedicated to proving Lemmas 4.1 through 4.3.

The proofs of Lemmas 4.1 and 4.2 are based on the following observation. For all u:

$$\Pr_v\,[\,f(u+v)-f(v)=g_f(u)\,] \geq 1/2 \ .$$

Hence, if $f(u)\neq g_f(u)$, then $f(u)$ is different from $f(u+v) - f(v)$ at least half of the time. That is:

$$\Pr_{u,v}\,[\,f(u)+f(v)\neq f(u+v)\ \mid f(u)\neq g_f(u)\,] \geq 1/2. \quad (1)$$

Proof of Lemma 4.1: Let $g = g_f$. Simple conditioning says that $\mathsf{Err}(f)$ is at least

$$\Pr_{u,v}\,[\,f(u)+f(v)\neq f(u+v)\ \mid f(u)\neq g(u)\,]\ \mathsf{Dist}(f,g) \ .$$

But by (1) we know this is at least $\mathsf{Dist}(f,g)/2$. ∎

Proof of Lemma 4.2: Let $g = g_f$ and assume it is linear. As observed in the proof of Lemma 2.2, we have that $\mathsf{Err}(f) =$

$$\begin{aligned}&3\,\Pr_{u,v}\,[\,f(u)\neq g(u), f(v)=g(v), f(u+v)=g(u+v)\,] \\ &+\Pr_{u,v}\,[\,f(u)\neq g(u), f(v)\neq g(v), f(u+v)\neq g(u+v)\,] \\ &= 3\,\Pr_{u,v}\,[\,f(u)+f(v)\neq f(u+v)\ \mid f(u)\neq g(u)\,]\ \mathsf{Dist}(f,g) \\ &-2\,\Pr_{u,v}\,[\,f(u)\neq g(u), f(v)\neq g(v), f(u+v)\neq g(u+v)\,] \ .\end{aligned}$$

In this last expression, the first term can be lower bounded, as in the proof of Lemma 4.1, by $3\,\mathsf{Dist}(f,g)/2$. The second term is equal to $2\,\mathsf{sl}(f,g)$. Thus, we have $\mathsf{Err}(f) \geq 3\,\mathsf{Dist}(f,g)/2 - 2\,\mathsf{sl}(f,g)$. Finally, applying Lemma 2.2, we get that $\mathsf{Err}(f) \geq 3\,\mathsf{Dist}(f,g) - 3\,\mathsf{Dist}(f,g)^2 - \mathsf{Err}(f)/2$. The lemma follows. ∎

Proof of Lemma 4.3: By contradiction. Assume g_f is not linear. Then, there are x and y distinct, such that

$g_f(x)+g_f(y)\neq g_f(x+y)$. Without loss of generality, assume that $g_f(x) = g_f(y) = g_f(x+y) = 1$. Furthermore, let $S = span\{x, y\}$. For every $s \in F^n$, define f_s to be the function from S to F, such that $f_s(u) = f(s + u)$. Hence

$$\mathsf{Err}(f) = E_{s,t}\left[p_{s,t}\right] \ , \tag{2}$$

where

$$p_{s,t} = \Pr_{u,v \stackrel{R}{\leftarrow} S}\left[\, f_s(u)+f_t(v)\neq f_{s+t}(u+v)\,\right] \ .$$

But, $p_{s,t}$ depends only on the values that f_s, f_t and f_{s+t} take. That is, on the *trace* of f at s, t and $s + t$, where the trace of f at w is defined as $[f_w(0), f_w(x), f_w(y), f_w(x + y)]$, and denoted by $tr_f(w)$.

To lower bound $p_{s,t}$, the following partition of the elements $s \in F^n$, according to the trace of f at s, plays a crucial role:

$$\begin{aligned}
H_0 &= \{\, s : tr_f(s) \text{ equals } [0,0,0,0] \text{ or } [1,1,1,1]\,\}\\
H_x &= \{\, s : tr_f(s) \text{ equals } [0,0,1,1] \text{ or } [1,1,0,0]\,\}\\
H_y &= \{\, s : tr_f(s) \text{ equals } [0,1,0,1] \text{ or } [1,0,1,0]\,\}\\
H_{x+y} &= \{\, s : tr_f(s) \text{ equals } [0,1,1,0] \text{ or } [1,0,0,1]\,\}\\
H_{odd} &= \{\, s : tr_f(s) \text{ has an odd number of 1's }\,\} \ .
\end{aligned}$$

We also partition $F^n \times F^n$ into six sets as follows:

- $\mathcal{A}$ = Set of all (s,t) such that $\{s,t,s+t\}$ are all in the same set, either H_0 or H_x or H_y or H_{x+y}
- $\mathcal{B}$ = Set of all (s,t) such that two of $\{s,t,s+t\}$ are in the same set H_0 or H_x or H_y or H_{x+y}, and the other one is in H_{odd}
- $\mathcal{C}$ = Set of all (s,t) such that at least two of $\{s,t,s+t\}$ are in H_{odd}
- $\mathcal{D}$ = Set of all (s,t) such that $\{s,t,s+t\} \subset H_0 \cup H_x \cup H_y \cup H_{x+y}$ with exactly two elements from the same set H_0, H_x, H_y or H_{x+y}
- $\mathcal{E}$ = Set of all (s,t) such that one of $\{s,t,s+t\}$ is in H_{odd}, the other two are from different sets in H_0, H_x, H_y and H_{x+y}
- $\mathcal{F}$ = Set of all (s,t) such that $\{s,t,s+t\}$ are from different sets H_0, H_x, H_y, H_{x+y}

We now proceed to show a lower bound for $\mathsf{Err}(f)$ which depends on the relative size of the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{E}$, and $\mathcal{F}$. Indeed, observe that if (s,t) is in $\mathcal{B}$, then $p_{s,t}$ is at least $1/4$. If (s,t) is in $\mathcal{C}$, then $p_{s,t}$ is at least $3/8$. And, if (s,t) is in $\mathcal{D}$, $\mathcal{E}$ or $\mathcal{F}$, then $p_{s,t}$ is equal to $1/2$. Hence, if for a set $S \subseteq F^n \times F^n$ we let $\mu(S) = |S|/|F|^{2n}$, then (2) yields

$$\mathsf{Err}(f) \geq \mu(\mathcal{B})/4 + 3\mu(\mathcal{C})/8 + (\mu(\mathcal{D}) + \mu(\mathcal{E}) + \mu(\mathcal{F})/2 \ .$$

Recalling that

$$\mu(\mathcal{C}) = 1 - (\mu(\mathcal{A}) + \mu(\mathcal{B}) + \mu(\mathcal{D}) + \mu(\mathcal{E}) + \mu(\mathcal{F}))$$

allows us to conclude that

$$\begin{aligned}
\mathsf{Err}(f) \quad \geq \quad & 3/8 - (3\mu(\mathcal{A}) + \mu(\mathcal{B}))/8 \\
& +(\mu(\mathcal{D}) + \mu(\mathcal{E}) + \mu(\mathcal{F}))/8. \qquad (3)
\end{aligned}$$

In what follows we denote, for simplicity's sake, $|H_0|/|F|^n$, $|H_x|/|F|^n$, $|H_y|/|F|^n$, $|H_{x+y}|/|F|^n$ and $|H_{odd}|/|F|^n$, by h_0, h_x, h_y, h_{x+y} and h_{odd} respectively. We now derive from (3) another lower bound for $\mathsf{Err}(f)$ which will depend solely on $h_0, h_x, h_y, h_{x+y}, h_{odd}$ and $\mu(\mathcal{F})$.

We first need the following identities relating the measure of the sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, $\mathcal{D}$, $\mathcal{E}$ and $\mathcal{F}$, to h_0, h_x, h_y, h_{x+y}, and h_{odd}:

$$3\mu(\mathcal{A}) + \mu(\mathcal{B}) + \mu(\mathcal{D}) = 3\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) \ , \tag{4}$$

and

$$\begin{aligned}
&2\mu(\mathcal{D}) + \mu(\mathcal{E}) + 3\mu(\mathcal{F}) = \\
&\quad 3\left((1 - h_{odd})^2 - \left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right)\right). \qquad (5)
\end{aligned}$$

Adding $-1/8$ of (4) and $1/8$ of (5) to (3), gives

$$\begin{aligned}
\mathsf{Err}(f) \quad \geq \quad & \tfrac{3}{8} - \tfrac{3}{4}\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) \\
& +\tfrac{3}{8}(1 - h_{odd})^2 - \tfrac{1}{4}\mu(\mathcal{F}). \qquad (6)
\end{aligned}$$

We now proceed to upper bound $\mu(\mathcal{F})$. We divide the analysis into two cases. But first we assume, without loss of generality, that $h_x \leq h_y \leq h_{x+y}$. Observe also, that for a randomly chosen s, $f_s(x+y)-f_s(0)$ differs from $g_f(x+y)$ at most half of the time. Moreover since we have assumed that $g_f(x+y) = 1$, the following holds:

$$\begin{aligned}
1/2 \quad &\geq \quad \Pr_s\left[\, g_f(x + y)\neq f_s(x + y) - f_s(0)\,\right] \\
&= \quad h_0 + h_{x+y} + h_{odd}/2. \qquad (7)
\end{aligned}$$

It follows that $h_{x+y} \leq 1/2$.

Case 1: $h_x + h_y - h_0 - h_{x+y} > 1/4$.

By assumptions, $h_x \geq h_x + h_y - h_0 - h_{x+y} > 1/4$. So, $h_x, h_y, h_{x+y} \in (1/4, 1/2]$. Now let

$$\varphi(A, B, C) = \frac{|\{\, (u,v,w) \in A \times B \times C : u+v+w = 0\,\}|}{|F|^{2n}} \ .$$

Observe now, that for each element (u,v) of $\mathcal{F}$, $\{u, v, u+v\}$ either contains an element from H_0 or contains one element from each of the sets H_x, H_y and H_{x+y}.

The contribution to $\mathcal{F}$ of the elements (u,v), where $\{u, v, u + v\}$ contain elements from each of the sets H_x, H_y and H_{x+y}, is upper bounded by $6\,\varphi(H_x, H_y, H_{x+y})$. Since

$h_x, h_y, h_{x+y} \in (1/4, 1/2]$, the Summation Lemma implies that $6\,\varphi(H_x, H_y, H_{x+y})$ is

$$\begin{aligned}
&\leq 6\,\varphi(H^*_x, H^*_y, H^*_{x+y}) \\
&= 6\left(\frac{1}{4} - \frac{h_x+h_y+h_{x+y}}{2} + h_x h_y + h_x h_{x+y} + h_y h_{x+y}\right) \\
&= \tfrac{3}{2} - 3\,(h_0+h_{odd})(h_x+h_y+h_{x+y}) - 3\,(h_x^2+h_y^2+h_{x+y}^2).
\end{aligned}$$

Furthermore, the contribution to $\mathcal{F}$ of the elements (u, v), where $\{u, v, u+v\}$ contains an element of H_0 is upper bounded by

$$\begin{aligned}
&3\,\varphi(H_0, H_x, H_y \cup H_{x+y}) + 3\,\varphi(H_0, H_y, H_x \cup H_{x+y}) \\
&\quad + 3\,\varphi(H_0, H_{x+y}, H_x \cup H_y)\ ,
\end{aligned}$$

which is at most $3h_0\,(h_x+h_y+h_{x+y})$. Putting it all together, we have

$$\mu(\mathcal{F}) \leq 3/2 - 3h_{odd}(h_x+h_y+h_{x+y}) - 3(h_x^2+h_y^2+h_{x+y}^2)\ ,$$

which jointly with (6) implies that $\mathsf{Err}(f)$ is

$$\begin{aligned}
\geq\ & \tfrac{3}{8} + \tfrac{3}{8}(1-h_{odd})^2 - \tfrac{3}{4}\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) \\
& -\tfrac{3}{8} + \tfrac{3}{4}h_{odd}(h_x + h_y + h_{x+y}) \\
& +\tfrac{3}{4}(h_x^2 + h_y^2 + h_{x+y}^2) \\
=\ & \tfrac{3}{8} - \tfrac{3}{8}h_{odd}^2 - \tfrac{3}{4}h_0 h_{odd} - \tfrac{3}{4}h_0^2 \\
\geq\ & \tfrac{3}{8} - \tfrac{3}{8}(h_{odd} + 4h_0)^2.
\end{aligned}$$

We conclude the analysis of this case by noting that

$$\begin{aligned}
1/4 &\geq 1 - 3(h_x + h_y - h_0 - h_{x+y}) \\
&\geq 1 - h_x - h_y - h_{x+y} + 3h_0 \\
&= h_{odd} + 4h_0\ ,
\end{aligned}$$

where the first inequality follows by case assumption, and the second one because $h_x \leq h_y \leq h_{x+y}$.

Case 2: $h_x + h_y - h_0 - h_{x+y} \leq 1/4$.

To each element (u, v) in $\mathcal{F}$, associate the unique tuple $(u', v') \in \{u, v, u+v\} \times \{u, v, u+v\}$, such that $(u', v') \in H_0 \times H_{x+y} \cup H_x \times H_y$. This scheme associates to each element of $H_0 \times H_{x+y} \cup H_x \times H_y$ at most 6 elements of $\mathcal{F}$. Thus, $\mu(\mathcal{F}) \leq 6\,(h_0 h_{x+y} + h_x h_y)$. Which jointly with (6) implies

$$\begin{aligned}
\mathsf{Err}(f) \geq\ & \tfrac{3}{8} + \tfrac{3}{8}(1-h_{odd})^2 - \tfrac{3}{2}\,(h_0 h_{x+y} + h_x h_y) \\
& -\tfrac{3}{4}\left(h_0^2 + h_x^2 + h_y^2 + h_{x+y}^2\right) \\
=\ & \tfrac{3}{8} - \tfrac{3}{8}\,(h_x + h_y - h_0 - h_{x+y})^2\ .
\end{aligned}$$

The analysis of this case concludes by observing that

$$\begin{aligned}
1/4 &\geq h_x + h_y - h_0 - h_{x+y} \\
&= 1 - h_{odd} - 2(h_0 + h_{x+y}) \\
&\geq 0\ ,
\end{aligned}$$

where the first inequality is by case assumption, and the latter one follows from (7). ∎

Acknowledgments

J. H. thanks Mike Sipser for making his visit to MIT possible. M. K. thanks Dan Kleitman, Carsten Lund, Mike Sipser and Dan Spielman for several interesting and helpful discussions. We thank Ronitt Rubinfeld for comments on an earlier draft.

References

[1] N. ALON AND J. H. SPENCER. The probabilistic method. John Wiley & Sons, Inc., 1992.

[2] S. ARORA, C. LUND, R. MOTWANI, M. SUDAN AND M. SZEGEDY. Proof verification and intractability of approximation problems. *Proceedings of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992.

[3] S. ARORA AND S. SAFRA. Probabilistic checking of proofs: a new characterization of NP. *Proceedings of the* 33rd *Symposium on Foundations of Computer Science*, IEEE, 1992.

[4] L. BABAI, L. FORTNOW AND C. LUND. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, Vol. 1, 3–40, 1991.

[5] L. BABAI, L. FORTNOW, L. LEVIN AND M. SZEGEDY. Checking computations in polylogarithmic time. *Proceedings of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991.

[6] M. BELLARE, S. GOLDWASSER, C. LUND AND A. RUSSELL. Efficient probabilistically checkable proofs and applications to approximation. *Proceedings of the* 25th *Annual Symposium on Theory of Computing*, ACM, 1993.

[7] M. BELLARE, O. GOLDREICH AND M. SUDAN. Free bits and non-approximability. *Proceedings of the* 36th *Symposium on Foundations of Computer Science*, IEEE, 1995.

[8] M. BELLARE AND M. SUDAN. Improved non-approximability results. *Proceedings of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994.

[9] M. BLUM, M. LUBY AND R. RUBINFELD. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences* Vol. 47, 549–595, 1993.

[10] D. COPPERSMITH. Private communications.

[11] U. FEIGE, S. GOLDWASSER, L. LOVÁSZ, S. SAFRA AND M. SZEGEDY. Approximating clique is almost NP-complete. *Proceedings of the* 32nd *Symposium on Foundations of Computer Science*, IEEE, 1991.

[12] K. FRIEDL, ZS. HÁTSÁGI AND A. SHEN. Low-degree testing. *Proceedings of the* 5th *Annual Symposium on Discrete Algorithms*, ACM-SIAM, 1994.

[13] K. FRIEDL AND M. SUDAN. Some Improvements to Total Degree Tests. *Proceedings of the Third Israel Symposium on Theory and Computing Systems*, IEEE, 1995.

[14] P. GEMMELL, R. LIPTON, R. RUBINFELD, M. SUDAN AND A. WIGDERSON. Self-testing/correcting for polynomials and for approximate functions. *Proceedings of the* 23rd *Annual Symposium on Theory of Computing*, ACM, 1991.

[15] D. J. KLEITMAN. Private communication.

[16] A. POLISHCHUK AND D. SPIELMAN. Nearly Linear Size Holographic Proofs. *Proceedings of the* 26th *Annual Symposium on Theory of Computing*, ACM, 1994.

[17] R. RUBINFELD AND M. SUDAN. Robust characterizations of polynomials and their applications to program testing. IBM Technical Report RC 19156, 1993. To appear in *SIAM Journal on Computing*.

A Tightness of the analysis for the case of general groups

Here is Coppersmith's example. Let m be divisible by 3. Let f be a function from $\mathcal{Z}_m^n$ to $\mathcal{Z}_m$ such that $f(u) = 3k$, if $u_1 \in \{3k-1, 3k, 3k+1\}$. Then, $\mathsf{Dist}(f) = 2/3$. Furthermore, $f(u) + f(v) \neq f(u+v)$ only if $u_1 = v_1 = 1 \pmod 3$, or $u_1 = v_1 = -1 \pmod 3$, *i.e.* $\mathsf{Err}(f) = 2/9$.

B Total degree 1 testing in characteristic two.

Although the main purpose of our work is to give a near optimal analysis of the BLR linearity test, we now describe and analyze a way of testing for total degree 1 in characteristic two. Our purpose is to further illustrate the strength and elegance of the Fourier analysis technique, as well as its more general applicability to the problem of analyzing program testers.

As usual, let $F = \mathrm{GF}(2)$. Recall that a function $p\colon F^n \to F$ is *total degree* 1 if $p(u) + p(v) + p(w) = p(u+v+w)$ for all u, v, w in F^n. In analogy to the case of linearity testing, define

- DEG_1 — Set of all polynomials of total degree 1 of F^n to F
- $\mathsf{Dist}_1(f) \stackrel{\text{def}}{=} \min\{\, \mathsf{Dist}(f,p) : p \in \mathrm{DEG}_1 \,\}$ — Distance of f to its closest polynomial of total degree 1.

Again, assume we are given oracle access to a function f mapping F^n to F. We want to test that f is close to a polynomial of total degree 1 of F^n to F, and make as few oracle queries as possible.

THE TOTAL DEGREE 1 TEST. The test is the following — Pick $u, v, w \in F^n$ at random, query the oracle to obtain $f(u), f(v), f(w), f(u+v+w)$, and reject if $f(u) + f(v) + f(w) \neq f(u+v+w)$. Let $\mathsf{Err}_1(f) \stackrel{\text{def}}{=}$

$$\Pr_{u,v,w \stackrel{R}{\leftarrow} F^n} \left[f(u) + f(v) + f(w) \neq f(u+v+w) \right] ,$$

be the probability that the test rejects f. Also let $\mathrm{REJ}_1(x) \stackrel{\text{def}}{=}$

$$\min\{\, \mathsf{Err}_1(f) : f\colon F^n \to F \text{ s.t. } \mathsf{Dist}_1(f) = x \,\} .$$

In order to understand how good this test is we need to lower bound $\mathsf{Err}_1(f)$ in terms of $x = \mathsf{Dist}_1(f)$. The techniques discussed in this work gives us tools for achieving this goal. In fact, observe that if $h(\cdot) = (-1)^{f(\cdot)}$ (f viewed as a real valued function), then $|h_\alpha| \leq 1 - 2x$, for all α in F^n. Indeed, note that all functions in DEG_1 are of the form $l_\alpha(\cdot) + \beta$, where β is in F and l_α denotes the function that sends u to $\sum_{i=1}^n \alpha_i u_i$ (arithmetic over F). Then, as in Lemma 2.1, we have that $\widehat{h}_\alpha = 1 - 2\mathsf{Dist}(f, l_\alpha) \leq 1 - 2x$. Moreover, since $\mathsf{Dist}(f, l_\alpha) + \mathsf{Dist}(f, l_\alpha + 1) = 1$, we also have that $\widehat{h}_\alpha = 2\mathsf{Dist}(f, l_\alpha + 1) - 1 \geq 2x - 1$, which proves the claim.

The crucial observation is now the following:

$$\mathsf{Err}(f) = \tfrac{1}{2}\left(1 - (h * h * h * h)(0)\right) .$$

Our previous claim and an argument similar to the one sketched in the proof of Theorem 1.2 yield

$$\begin{aligned} \mathsf{Err}_1(f) &= \tfrac{1}{2}\left(1 - \textstyle\sum_\alpha (\widehat{h}_\alpha)^4\right) \\ &\geq \tfrac{1}{2}\left(1 - (1-2x)^2 \textstyle\sum_\alpha (\widehat{h}_\alpha)^2\right) \\ &= 2x(1-x). \end{aligned}$$

Finally, note that $f(u) + f(v) + f(w)$ and $f(u+v+w)$ are distinct if and only if f differs from every $p \in \mathrm{DEG}_1$ in exactly one of the points $\{u, v, w, u+v+w\}$, or in exactly three of the points $\{u, v, w, u+v+w\}$. This observation leads to a generalization of Lemma 2.2 that allows to show that $\mathsf{Err}_1(f) \geq 8x(1-x)(1/2-x)$. This, coupled with our previous derivations yields:

Lemma B.1 $\mathrm{REJ}_1(x)$ is lower bounded by

$$\max\{\, 8x(1-x)(1/2-x) \,,\, 2x(1-x) \,\} .$$