3-Bit Dictator Testing: 1 vs. 5/8

Ryan O'Donnell
Carnegie Mellon University*
odonnell@cs.cmu.edu

Yi Wu
Carnegie Mellon University
yiwu@cs.cmu.edu

June 27, 2008

Abstract

In the conclusion of his monumental paper on optimal inapproximability results, Håstad [13] suggested that Fourier analysis of Dictator (Long Code) Tests does not seem universally applicable in the study of CSPs. His main open question was to determine if the technique could resolve the approximability of *satisfiable* 3-bit constraint satisfaction problems. In particular, he asked if the "Not Two" (`NTW`) predicate is non-approximable beyond the random assignment threshold of 5/8 on satisfiable instances. Around the same time, Zwick [30] showed that all satisfiable 3-CSPs are 5/8-approximable, and conjectured that the 5/8 is optimal.

In this work we show that Fourier analysis techniques *can* produce a Dictator Test based on `NTW` with completeness 1 and soundness 5/8. Our test's analysis uses the Bonami-Gross-Beckner hypercontractive inequality. We also show a soundness *lower bound* of 5/8 for all 3-query Dictator Tests with perfect completeness. This lower bound for Property Testing is proved in part via a semidefinite programming algorithm of Zwick [30].

Our work precisely determines the 3-query "Dictatorship Testing gap". Although this represents progress on Zwick's conjecture, current PCP "outer verifier" technology is insufficient to convert our Dictator Test into an NP-hardness-of-approximation result.

*Supported in part by NSF CAREER grant CCF-0747250.

1 Introduction: Dictator Testing, and its motivation

1.1 Dictator Testing

In this paper we study a Property Testing problem called *Dictator Testing*. Dictator Testing is strongly motivated by its applications to proving NP-hardness-of-approximation results (in which context it is often called "Long Code Testing"). We describe the Dictator Testing problem informally in this section; for formal definitions, see Section 4.1.

In Dictator Testing, we have black-box query access to an unknown boolean function $f : \{0,1\}^n \to \{0,1\}$. The goal is to test the extent to which f is close to a "dictator" function; i.e., one of the n functions of the form

$$f(x_1, \dots, x_n) = x_i.$$

Recall that a "test" is a randomized algorithm which makes a very small number of queries to f and then either "accepts" or "rejects". It is "nonadaptive" if it determines all query strings before seeing the responses. The Dictator Testing problem was first studied by Bellare, Goldreich, and Sudan [1], with hardness-of-approximation as the motivation. It was later independently introduced, with the "dictator" terminology, by Parnas, Ron, and Samorodnitsky [24].

Definition 1.1. *A Dictator Test has* completeness c *if all* n *dictator functions are accepted with probability at least* c*. We say a Dictator Test has* perfect completeness *if it has completeness* 1*.*

In this paper we consider only nonadaptive Dictator Tests with perfect completeness, unless otherwise specified. As for the "soundness" of a Dictator Test, we briefly discuss several possible criteria one could require:

Local testability: In this model of soundness, any function f which is ϵ-far from every dictator should be accepted with probability at most $1 - \Omega(\epsilon)$. Here we hope to make a very small constant number of queries, such as 3 or 4. Bellare, Goldreich, and Sudan [1] (BGS) implicitly gave a 4-query (indeed, 3-query adaptive) such test; for a simple 3-query construction, see e.g. [22]. Such "Local Dictator Tests" play a useful role in Dinur's proof of the PCP Theorem [4].

Usual Property Testing soundness: In this model, the tester is also given a parameter ϵ; any function f which is ϵ-far from every dictator should be accepted with probability at most, say, 1/3. Here we hope to make few queries as a function of ϵ. Note that by repeating a local test $O(1/\epsilon)$ times, we get an $O(1/\epsilon)$-query Dictator Test with the usual Property Testing soundness.

Rejecting very far functions: In this model, soundness is only concerned with functions f which are $(1/2 - o(1))$-far from every dictator; i.e., have correlation at most $o(1)$ with every dictator. The goal is to accept such functions with as low a probability as possible, while making a very small constant number of queries. For example, the 4-query BGS test accepts these functions with probability at most $17/20 + o(1)$. The major contribution of BGS was showing that Dictator Tests with this kind of soundness can yield strong NP-hardness-of-approximation results for constraint satisfaction problems ("CSPs").

Rejecting "quasirandom" functions. Håstad [12, 13] introduced this relaxation of the above. One can think of it as only requiring soundness for functions f which have correlation at most $o(1)$ with every *"junta"* (function depending on only $O(1)$ coordinates). We refer to such f's as "quasirandom", and such tests as "Dictator-vs.-Quasirandom Tests".

Definition 1.2. (Informal.) *A Dictator-vs.-Quasirandom Test has* soundness *at most s if every quasirandom function is accepted with probability at most $s + o(1)$.*

For example, Håstad [13] gave a (nonadaptive) 3-query Dictator-vs.-Quasirandom Test with soundness 3/4. As Håstad and others have demonstrated, Dictator-vs.-Quasirandom Tests can often be used to prove optimal inapproximability results for CSPs.

1.2 Optimal approximability for k-CSPs

The major motivation for Dictator Testing is proving hardness-of-approximation results for CSPs. We discuss this connection in Section 1.3; however first let us describe k-CSPs. A (boolean) "k-CSP" is a system of constraints over n boolean-valued variables v_i in which each constraint involves at most k of the variables. We also assume each constraint has a nonnegative weight, with the sum of all weights being 1. Given a k-CSP, the natural algorithmic task, called "Max-kCSP", is to find an assignment to the variables such that the total weight of satisfied constraints is as large as possible. We write "Opt" to denote the weight satisfied by the best possible assignment. We also say that a CSP is "satisfiable" if Opt = 1. Our main motivation in this paper is studying the difficulty of *satisfiable Max-3CSP* instances, of which the following is a small example:

weight:	constraint:
1/4	$v_1 \wedge \neg v_3 \wedge v_4$
1/4	`IF` v_3 `THEN` v_4 `ELSE` $\neg v_5$
1/2	$v_2 \neq v_5$

Each constraint in a k-CSP is of a certain "type"; more precisely, it is a certain predicate of arity at most k over the variables. If we specialize Max-kCSP by restricting the type of constraints allowed, we get some of the most canonical NP optimization problems. For example:

- Max-2Sat: only the four predicates $v_i \vee v_j$, $v_i \vee \neg v_j$, $\neg v_i \vee v_j$, $\neg v_i \vee \neg v_j$;
- Max-3Lin: only the two predicates $v_i \oplus v_j \oplus v_k$, $\neg(v_i \oplus v_j \oplus v_k)$;
- Max-Cut: only the predicate $v_i \neq v_j$.

If we restrict the allowed predicates to some set Φ, we call the associated algorithmic problem Max-Φ.

Determining Opt for these CSP problems is NP-hard, but there is an enormous literature on polynomial-time approximation algorithms. To complement this, we can also look for NP-hardness-of-approximation results. As we describe in the next section, all of the best known inapproximability results rely critically on Dictator Testing.

We now have optimal (i.e., matching) approximation algorithms and NP-hardness-of-approximation results for some key problems: Max-kLin(mod q) for $k \geq 3$ [13], Max-3Sat [13, 15, 31], and a few other Max-kCSP problems with $k \geq 3$ [13, 30, 29, 9]. However, many basic problems remain unresolved; for example, we do not know if 90%-approximating Max-Cut is in P or is NP-hard. Similarly, given a satisfiable 3-CSP, we do not know if satisfying 2/3 of the constraint-weight is in P or is NP-hard.

1.3 The connection between Dictator Testing and inapproximability

There is a close connection between CSPs and the Property Testing of boolean functions. To illustrate this, suppose $\mathcal{T}$ is a nonadaptive 3-query Dictator-vs.-Quasirandom Test on functions $f : \{0,1\}^Q \to \{0,1\}$. Imagine we consider all possible random choices of $\mathcal{T}$, and in each case write down the (up to) 3 strings x, y, z queried and the predicate applied to the outcomes to decide accept/reject. The complete behavior of $\mathcal{T}$ might then look like the following:

$$\begin{array}{llccccc} \text{with probability } p_1, & \text{accept iff} & f(x^{(1)}) & \vee & f(y^{(1)}) & \vee & f(z^{(1)}) \\ \text{with probability } p_2, & \text{accept iff} & \neg f(x^{(2)}) & \vee & f(y^{(2)}) & & \\ \text{with probability } p_3, & \text{accept iff} & \neg f(x^{(3)}) & \vee & \neg f(y^{(3)}) & \vee & \neg f(z^{(3)}) \\ & & & \cdots & & & \end{array}$$

This is precisely an instance of Max-3CSP, in which the "variables" are the $f(x)$'s. Note that the weights p_i indeed sum up to 1. More generally, if $\mathcal{T}$ makes at most q nonadaptive queries it can be viewed as an instance of Max-qCSP. Further, suppose that $\mathcal{T}$ "uses the predicate set Φ" — i.e., its decision to accept/reject is always based on applying a predicate from the set Φ to its query responses. Then $\mathcal{T}$ can be viewed as an instance of Max-Φ. The above example illustrates a tester which uses the set of ORs on up to 3 literals; thus it can be viewed as an instance of Max-3Sat.

Suppose that $\mathcal{T}$ is a Dictator Test with completeness at least c. Then the Opt of the associated CSP is at least c; indeed, there are n distinct solutions, the dictators, of value at least c. More crucially, suppose further that $\mathcal{T}$ is a Dictator-vs.-Quasirandom Test with soundness at most s. Taking the contrapositive of Definition 1.2, this means that any solution f satisfying slightly more than weight s of the constraints must be slightly correlated with a junta on constant number of coordinates; i.e., it must "highlight" a small number of dictators. These two properties of the test, taken together, make it useful as a *gadget* in an NP-hardness-of-approximation reduction. Specifically, if $\mathcal{T}$ uses predicate set Φ, it can be used to prove hardness for the Max-Φ problem. Indeed, in the study of inapproximability, one has the following "Rule of Thumb":

Rule of Thumb. *For the Max-Φ problem, to prove that distinguishing* $\mathrm{Opt} \geq c$ *and* $\mathrm{Opt} \leq s + \epsilon$ *is NP-hard, construct a nonadaptive Dictator-vs.-Quasirandom Test using Φ, with completeness c and soundness s.*

For example, the key step in Håstad's famous $(7/8 + \epsilon)$-hardness result for satisfiable Max-3Sat instances was his construction of a 3-query Dictator-vs.-Quasirandom Test using OR tests, with perfect completeness and soundness $7/8$. Actual theorems based on the Rule of Thumb are discussed in Section 3.1.

2 Our contribution: optimal perfect-completeness, 3-query tests

2.1 Satisfiable 3-CSPs

One of the most notable open questions in the area of CSP approximability is that of analyzing satisfiable 3-CSPs:

Question 1: *Given a satisfiable 3-CSP, can we efficiently satisfy constraint-weight at least s?*

Zwick [30] made a comprehensive study of the Max-3CSP problem and gave an efficient algorithm which 5/8-satisfies any satisfiable 3-CSP instance. (This improved and built upon the earlier .514-algorithm of Trevisan [28].) Zwick conjectured that this algorithm is optimal; i.e., obtaining $5/8+\epsilon$ is NP-hard for all constant $\epsilon > 0$. In the language of Probabilistically Checkable Proofs, Zwick's conjecture states that $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,5/8+\epsilon}(O(\log n), 3)$.

Håstad's contemporaneous treatise on optimal inapproximability [11] gave an NP-hardness result for $s > 3/4$, and improving this was left as the main open problem in his work. Almost a decade later, no progress had been made on closing the gap, and Håstad re-posed the problem [14]. Shortly thereafter, Khot and Saket [19] achieved the first improved hardness, showing that satisfiable 3-CSP instances are NP-hard to approximate to any factor better than $20/27 \approx .74$.

In originally posing the problem, Håstad suggested that Fourier analysis of Dictator Tests does not seem universally applicable in the study of CSPs. The associated Dictator-vs.-Quasirandom Property Testing question here is particularly easy to state:

Question 2: *What is the least possible soundness of a nonadaptive* 3*-query Dictator-vs.-Quasirandom Test with perfect completeness?*

Khot and Saket's result yields an upper bound 20/27 for this question; the question of proving a lower bound has not been explicitly considered.

The main result in this paper is an exact answer to Question 2:

Theorem 2.1. *(Main results, informally stated.)*

1. *There is a nonadaptive* 3*-query Dictator-vs.-Quasirandom Test with perfect completeness and soundness* 5/8*. The test uses only the "Not-Two" (*`NTW`*) predicate.*
2. *Every nonadaptive* 3*-query Dictator-vs.-Quasirandom Test with perfect completeness has soundness at least* 5/8*.*

Fourier analysis is the key to the proof of the upper bound.

The `NTW` predicate. `NTW` is the 3-bit predicate which is satisfied if the number of True inputs is either zero, one, or three — i.e., not two. Our test actually uses all eight `NTW` predicates gotten by allowing the query responses to be negated. Although Zwick's algorithm satisfies 5/8 of the constraints in any 3-CSP, even with mixed "types" of constraints, the bottleneck predicate for him is the `NTW` constraint. Håstad's open question more specifically asked whether or not the `NTW` predicate is satisfiable beyond the random-assignment threshold of 5/8 on satisfiable instances.

What we *don't* prove. Unfortunately, the formal theorems behind the Rule of Thumb are *not* sufficient to convert our Dictator Test into a $5/8+\epsilon$ NP-hardness result for satisfiable Max-`NTW` instances, and thus prove Zwick's conjecture. See the discussions in Sections 3.1 and 6. However in an upcoming work building on the present paper, we will show that this result can be obtained assuming Khot's "d-to-1 Conjecture" ([16]) for any constant d.

2.2 Methods

Upper bound. Given the task of constructing an NTW-based Dictator-vs.-Quasirandom Test with perfect completeness, we describe in Appendix D how the correct test is almost "forced" upon us. Thus the main task is in the analysis of this test. For this we use some slightly tricky Fourier analysis, including the hypercontractive inequality [3]. This is the first Dictator Testing result we are aware of that uses the hypercontractive inequality without using the Invariance Principle [21]. Indeed, the Invariance Principle does not seem useful for our result, and the fact that we use the hypercontractive inequality directly gives us an exponentially better tradeoff between soundness and influences than those given by Invariance-based analyses.

Lower bound. The problem of proving lower bounds for Dictator Tests (of the sort needed for inapproximability) does not seem to have been considered until extremely recently, although Samorodnitsky and Trevisan [26] studied the problem for Linearity Tests in 2006. In [23], the present authors observed that because of the "Rule of Thumb" described in Section 1.3, the existence of strong approximation algorithms "ought to" imply Dictator-vs.-Quasirandom Testing lower bounds. As noted by the authors [23] for Max-Cut and independently by Raghavendra [25] for any 2-CSP, a version of this implication can be proved using the work of Khot and Vishnoi [20]. But this version loses perfect completeness, our raison d'etre, and is unproven for 3-CSPs.

In any case, passing through either the Rule of Thumb or Khot-Vishnoi seems like a highly roundabout way to analyze Property Testing algorithms. Instead we work directly, as was done in a much simpler context in [23]. Our proof of the soundness lower bound in Theorem 2.1 involves using Zwick's algorithm to show the existence a quasirandom function passing any given perfect-completeness Dictator Test with probability at least $5/8 - o(1)$; this quasirandom function is either a random threshold function or a random odd parity. Zwick's algorithm in part uses semidefinite programming, and we feel that the use of semidefinite programming in Property Testing lower bounds is an interesting method.

3 Related work

3.1 Formal encapsulations of the Rule of Thumb

There are essentially two known formalizations of the Rule of Thumb from Section 1.3. The original formalization is due to Bellare, Goldreich, and Sudan [1] and was developed by Håstad [13]. It produces the desired NP-hardness-of-approximation result, but requires more than just the Dictator-vs.-Quasirandom Test. Without getting into precise details, it requires a "two-function" version of Dictator-vs.-Quasirandom Tests, in which one has query-access to an unknown $f : \{0,1\}^Q \to \{0,1\}$ and $g : \{0,1\}^{dQ} \to \{0,1\}$, where d should be thought of as polynomially-related to Q. Roughly speaking, one needs a test (using Φ) which checks the extent to which f is close to a dictator i, g is close to a dictator j, and $d(i-1)+1 \le j \le di$. A formal statement appears in, e.g., [10]. The difficulty of generalizing our Dictator-vs.-Quasirandom Test in this way is discussed in Section 6.

The second formalization of the Rule of Thumb appears implicitly in the work of Khot, Kindler, Mossel, and O'Donnell [17]. It gives the desired hardness-of-approximation immediately from the appropriate Dictator-vs.-Quasirandom Test, but there are two deficiencies: First, it requires assuming Khot's Unique Games Conjecture [16]. Second, it has a slight loss in completeness: one only gets hardness of distinguishing $\text{Opt} \ge c - \epsilon$ and $\text{Opt} \le s + \epsilon$, for all constant $\epsilon > 0$. Note that even if we are willing to assume the Unique Games Conjecture, this formalization of the Rule of

Thumb is useless to us — it converts our Dictator-vs.-Quasirandom Test from Theorem 2.1 into a $1-\epsilon$ vs. $5/8+\epsilon$ hardness result for Max-NTW. However Håstad [13] has already shown this result without assuming the Unique Games Conjecture. Further, for the question of nearly-satisfiable Max-3CSPs, Håstad [13] showed $1-\epsilon$ vs. $1/2+\epsilon$ hardness (for Max-3Lin) and Zwick [30] showed that the $1/2$ is optimal.

3.2 Other related work

Besides the connection mentioned in Section 2.2, Raghavendra [25] proves other links between Dictator-vs.-Quasirandom Testing and semidefinite programming (SDP). His main result is that SDP gaps for a CSP can be translated into nearly-equivalent Dictator-vs.-Quasirandom Test. Unfortunately, this does not help us for several reasons. The main reason is again the loss of an ϵ in the completeness, transforming an SDP gap of c vs. s into a Dictator-vs.-Quasirandom Test with completeness $c-\epsilon$ and soundness s. As mentioned in Section 2.2, the distinction between perfect and near-perfect completeness is crucial for us. In addition to this, we are not aware of any explicit SDP gaps for 3-CSPs, certainly not for Max-NTW and not of the strengthened SDP-type needed for Raghavendra's reduction.

As mentioned, progress has been slow for Håstad and Zwick's open problem on the hardness of satisfiable CSPs. An early work of Guruswami, Lewin, Sudan, and Trevisan [10] gave a 3-query *adaptive* Dictator-vs.-Quasirandom Test with soundness $1/2$. This translated into a $1/2+\epsilon$ NP-hardness result for satisfiable CSPs with depth-3-decision-tree constraints. They also gave a 4-query non-adaptive Dictator-vs.-Quasirandom Test with soundness $1/2$. Much later, Engebretsen and Holmerin [5] considered the NP-hardness-of-approximation for satisfiable q-ary Max-kCSP; they generalized Håstad's $3/4$-hardness and [10]'s $1/2$-hardness for 3-CSPs and 4-CSPs, respectively.

Finally, Khot and Saket [19] directly tackled the problem of NP-hardness for satisfiable Max-3CSP. Because they wanted to use the unconditional formalization of the Rule of Thumb described at the beginning of Section 3.1, they had to use a more complicated, p-biased Dictator Test based on multiple predicates. As mentioned, they achieved $20/27+\epsilon$ NP-hardness for satisfiable Max-3CSP with a mix of predicate types.

4 Definitions and preliminaries

The reader is assumed familiar with the basics of Fourier analysis of boolean functions; see, e.g., [27]. As is standard in Fourier analysis, our default representation for bits will be $+1$ and -1 rather than 0 and 1. However the reader is warned that this will change in some places.

4.1 Dictator-vs.-Quasirandom Tests

As mentioned in Section 1.1, to get Dictator Tests useful for optimal inapproximability, Håstad [12] relaxed the usual Property Testing soundness notion. In particular, he considered tests that are only required to reject *quasirandom* functions with high probability. Håstad effectively considered the following definition:

Definition 4.1. *For* $0 \le \epsilon, \delta \le 1$*, we say a function* $f : \{-1,1\}^n \to [-1,1]$ *is* (ϵ,δ)-quasirandom *if*

$$\hat{f}(S)^2 \le \epsilon \quad \text{for all } 0 < |S| \le 1/\delta.$$

Let us make a few comments on this notion. We have chosen the terminology by analogy with quasirandom graphs, which are graphs that have close to the "expected" number of copies of each small subgraph: one can check that an (ϵ, δ)-quasirandom function has covariance at most $\sqrt{2^{1/\delta}\epsilon}$ with every "junta" function on at most $1/\delta$ coordinates. Note also that the definition becomes stricter as ϵ and δ decrease. Some common quasirandom functions include the two constant functions, the Majority function, and parity functions on large numbers of bits. A further short discussion of quasirandom functions — and a related notion of "Gaussianic" functions — appears in Appendix B.

Let us now formally define Dictator-vs.-Quasirandom Tests:

Definition 4.2. *A k-query nonadaptive* Dictator-vs.-Quasirandom Test $\mathcal{T}$ *for functions* $f : \{-1,1\}^n \to \{-1,1\}$*, using predicates from the set* Φ*, is a probability distribution over constraints*

$$\phi(f(x_1), \dots, f(x_{k'})),$$

where $k' \leq k$ *and* $\phi \in \Phi$*. The* completeness c *of* $\mathcal{T}$ *is the minimum probability of passing among dictator functions* $f(x) = x_i$*. We say* $\mathcal{T}$ *has* perfect completeness *if all* n *dictators pass with probability* 1.

The (ϵ, δ)-soundness *of* $\mathcal{T}$ *is the maximum probability* $s(\epsilon, \delta)$ *of passing among* (ϵ, δ)*-quasirandom functions* f*. We say that a family of tests* $(\mathcal{T}_n)$*, parameterized by* n*, has* soundness s *if for all* $\eta > 0$ *there exist* $\epsilon, \delta > 0$ *such that for all sufficiently large* n*, the test* $\mathcal{T}_n$ *has* (ϵ, δ)*-soundness at most* $s + \eta$*.*

4.2 Constraint predicates

For us, a *k-ary predicate* is a function $\phi : \{-1,1\}^k \to \{0,1\}$. We say a string u *satisfies* ϕ if $\phi(u) = 1$, and call the set of strings satisfying ϕ the *support* of ϕ. The convention of having $\{0,1\}$-outputs for predicates is to make the Fourier analysis and semidefinite programming most natural. For inputs, we continue to represent logical "True" and "False" by -1 and 1 respectively; this is done so that logical XOR corresponds to real-valued multiplication. In particular, we have

$$\mathtt{AND}_3(-1,-1,-1) = 1, \quad \mathtt{AND}_3(-1,-1,1) = 0, \quad \mathtt{XOR}_3(-1,-1,-1) = 1, \quad \mathtt{NTW}(-1,1,-1) = 0,$$

etc. For the remained of the paper, whenever we refer to a test using predicates from a set Φ, we *assume* Φ *is closed under input negations.* In particular, we will henceforth not distinguish the predicates $\phi(\pm x_1, \dots, \pm x_k)$. Note that for some problems (e.g., Max-Cut) this distinction *would* be important.

For our Property Testing lower bound we are concerned with general 3-query tests and hence must consider all possible predicates of arity at most 3. As described by Zwick [30], there are twenty-two possible predicates on up to 3 bits:

$$\mathtt{TRU}_0, \mathtt{FLS}_0, \mathtt{IDN}_1, \mathtt{AND}_2, \mathtt{OR}_2, \mathtt{XOR}_2, \{\mathtt{AND}_3, \mathtt{EQU}_3, \mathtt{AXR}_3, \dots, \mathtt{OXR}_3, \mathtt{OR}_3\}.$$

Here the subscript denotes arity. $\mathtt{TRU}_0$ is the always-satisfied 0-ary predicate; $\mathtt{FLS}_0$ is the never-satisfied 0-ary predicate. $\mathtt{IDN}_1$ is the identity predicate. The 2-ary predicates are familiar, as are some of the sixteen 3-ary predicates. We will henceforth drop the subscripts indicating 3-arity; e.g., "XOR" means $\mathtt{XOR}_3$ unless written as $\mathtt{XOR}_2$. The precise definitions of the 3-ary predicates do not much concern us: the main ones we will need to know are NTW, XOR, and EQU. The last of these is the "all-equal" predicate with support on the two strings $\{(-1,-1,-1), (1,1,1)\}$. A point that will be important for us is that the support of NTW is the union of the supports of XOR and EQU.

4.3 Testing averages

A trick introduced in [1, 17] is that of testing *averages* of functions. This is described fully in Appendix C. A special case of this trick is the notion of "folding" from [1]; suffice it to say, this lets us reduce any tester for a function $f : \{-1,1\}^n \to \{-1,1\}$ to a tester for f's "odd part", $f^{\text{odd}} : \{-1,1\}^n \to [-1,1]$ defined by $f^{\text{odd}}(x) = (f(x) - f(-x))/2 = \sum_{|S| \text{ odd}} \hat{f}(S)x_S$. For the definition of testing functions with range $[-1,1]$, again see Appendix C.

5 Dictator-vs.-Quasirandom Testing upper bound with NTW

In this section we prove the upper bound in Theorem 2.1: i.e., we give a family $(\mathcal{T}_n)$ of Dictator-vs.-Quasirandom Tests, using the predicate NTW, with perfect completeness and soundness 5/8. Our test is based on the following mixture distribution over $\{-1,1\}^3$:

$$\mathcal{D}_\delta = (1-\delta)\mathcal{D}_{\texttt{XOR}} + \delta\mathcal{D}_{\texttt{EQU}}. \tag{1}$$

Here $0 < \delta < 1/8$ is a parameter and $\mathcal{D}_\phi$ denotes the uniform distribution on the support of the predicate ϕ. Notice that the support of $\mathcal{D}_\delta$ is the support of the NTW predicate. We can now state our test:

Test $\mathcal{T}_n$ on function $f : \{-1,1\}^n \to [-1,1]$, with parameter $0 < \delta < 1/8$:

1. Form a triple of strings $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ by choosing each bit-triple $(\boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i)$ independently from $\mathcal{D}_\delta$.
2. Test $\texttt{NTW}(f^{\text{odd}}(\boldsymbol{x}), f^{\text{odd}}(\boldsymbol{y}), f^{\text{odd}}(\boldsymbol{z}))$.

Figure 1 *Our test $\mathcal{T}_n$.*

(In this paper we use boldface to denote random variables.)

Here is another way of looking at the test: Essentially, the tester first picks a "random restriction $f_{\boldsymbol{w}}$ with $*$-probability $1-\delta$" (in the sense of [6]), and then runs the BLR linearity test [2] on $f_{\boldsymbol{w}}$. This description differs from our actual test in three small ways: a) our test applies the NTW predicate in the end, rather than XOR, and hence accepts the extra answer $(1,1,1)$; b) the BLR test uses ¬XOR, not XOR, according to our notations; c) we also include the trick of testing only f^{odd}.

In Appendix D we describe why the test given above is almost forced. We also explain intuitively why it should be a good Dictator-vs.-Quasirandom Test. In particular we explain why the prototypical quasirandom functions — constants, Majority, and large parities — all pass with probability only around 5/8.

5.1 The hypercontractive inequality

Our analysis uses the "hypercontractive inequality" for $\{-1,1\}^n$, proved originally by Bonami [3] and independently by Gross [8]:

Theorem 5.1. *Suppose* $0 \le \rho \le 1$ *and* $q \ge 2$ *satisfy* $\rho \le 1/\sqrt{q-1}$*. Then for all* $f : \{-1,1\}^n \to \mathbb{R}$*,*

$$\|T_\rho f\|_q \le \|f\|_2.$$

Here T_ρ is the "noise operator" defined by $T_\rho f = \sum_{S\subseteq[n]} \rho^{|S|}\hat{f}(S)\chi_S$. Equivalently, $(T_\rho f)(x) = \mathbf{E}_{\boldsymbol{y}}[f(\boldsymbol{y})]$, where $\boldsymbol{y}$ is defined to be a random string "ρ-correlated to x"; i.e., $\boldsymbol{y}_i = x_i$ with probability ρ, and $\boldsymbol{y}_i$ is uniformly random otherwise.

We will actually need a strengthening of this inequality which addresses the scenario in which ρ is strictly smaller than $1/\sqrt{q-1}$. In this case, one can obtain an extra fractional power of $\|T_\rho f\|_2$ on the right-hand side:

Corollary 5.2. *Suppose* $0 \le \rho \le 1$*,* $q \ge 2$*, and* $0 \le \lambda \le 1$ *satisfy* $\rho^\lambda \le 1/\sqrt{q-1}$*. Then for all* $f : \{-1,1\}^n \to \mathbb{R}$*,*

$$\|T_\rho f\|_q \le \|T_\rho f\|_2^{1-\lambda}\ \|f\|_2^\lambda.$$

(By taking $\lambda = 1$, one recovers the usual hypercontractive inequality.)

The following short proof of Corollary 5.2 was communicated to us by an anonymous reviewer:

Proof.

$$\begin{aligned}
\|T_\rho f\|_q &= \|T_{\rho^\lambda} T_{\rho^{1-\lambda}} f\|_q^2 \\
&\le \|T_{\rho^{1-\lambda}} f\|_2^2 && \text{(by Theorem 5.1)} \\
&= \sum_{S\subseteq[n]} \rho^{2(1-\lambda)|S|}\hat{f}(S)^2 \\
&= \sum_{S\subseteq[n]} (\rho^{2|S|}\hat{f}(S)^2)^{1-\lambda} \cdot (\hat{f}(S)^2)^\lambda \\
&\le \left(\sum_{S\subseteq[n]} \rho^{2|S|}\hat{f}(S)^2\right)^{1-\lambda} \cdot \left(\sum_{S\subseteq[n]} \hat{f}(S)^2\right)^{\lambda} && \text{(by Hölder's inequality)} \\
&= \|T_\rho f\|_2^{1-\lambda}\ \|f\|_2^\lambda.
\end{aligned}$$

□

5.2 Analyzing our test

We now formally present and prove the testing upper bound stated in our main Theorem 2.1.

Theorem 5.3. *The Dictator-vs.-Quasirandom Test* $\mathcal{T}_n$ *with parameter* δ *described in Figure 1 has perfect completeness and* $(\delta, \delta/\ln(1/\delta))$*-soundness at most* $5/8 + 6\sqrt{\delta}$*.*

Proof. The completeness is clear. Suppose then f is $(\delta, \delta/\ln(1/\delta))$-quasirandom. We may also assume $f = f^{\text{odd}}$. Our goal is to show that f passes $\mathcal{T}_n$ with probability at most $5/8 + 6\sqrt{\delta}$. As usual in such proofs, we arithmetize the probability the test passes using Fourier analysis. Recalling (16), we have

$$\mathbf{E}[\mathtt{NTW}(f(\boldsymbol{x}), f(\boldsymbol{y}), f(\boldsymbol{z}))] = \\ \mathbf{E}\left[\tfrac{5}{8} + \tfrac{1}{8}(f(\boldsymbol{x}) + f(\boldsymbol{y}) + f(\boldsymbol{z})) + \tfrac{1}{8}(f(\boldsymbol{x})f(\boldsymbol{y}) + f(\boldsymbol{y})f(\boldsymbol{z}) + f(\boldsymbol{x})f(\boldsymbol{z})) - \tfrac{3}{8}f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z})\right].$$

Note that the marginal on each of $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$ is uniform; since we are assuming f odd, $\mathbf{E}[f(\boldsymbol{x})] = \mathbf{E}[f(\boldsymbol{y})] = \mathbf{E}[f(\boldsymbol{z})] = 0$. In addition, it is easy to see that the joint distribution on $(\boldsymbol{x}, \boldsymbol{y})$ is that of δ-correlated strings — we get correlation 0 from $\mathcal{D}_{\mathtt{XOR}}$ with probability $1-\delta$ and correlation 1 from $\mathcal{D}_{\mathtt{EQU}}$ with probability δ. Hence

$$\mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{y})] = \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x}) \cdot (T_\delta f)(\boldsymbol{x})] = \sum_S \delta^{|S|}\hat{f}(S)^2 = \sum_{|S| \text{ odd}} \delta^{|S|}\hat{f}(S)^2 \leq \delta^1,$$

where we again used the fact that f is odd. By symmetry, the above holds also for $\mathbf{E}[f(\boldsymbol{y})f(\boldsymbol{z})]$ and $\mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{z})]$. Thus we have established

$$\mathbf{E}[\mathtt{NTW}(f(\boldsymbol{x}), f(\boldsymbol{y}), f(\boldsymbol{z}))] \leq \tfrac{5}{8} + \tfrac{3}{8}\delta - \tfrac{3}{8}\mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z})] \tag{2}$$

and it remains to show that $-\mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z})]$ is small.

By the Fourier decomposition we have

$$\mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z})] = \sum_{A,B,C \subseteq [n]} \hat{f}(A)\hat{f}(B)\hat{f}(C)\mathbf{E}[\boldsymbol{x}_A\boldsymbol{y}_B\boldsymbol{z}_C]. \tag{3}$$

As we have already seen, $\mathbf{E}[\boldsymbol{x}_i] = \mathbf{E}[\boldsymbol{y}_i] = \mathbf{E}[\boldsymbol{z}_i] = 0$ and $\mathbf{E}[\boldsymbol{x}_i\boldsymbol{y}_i] = \mathbf{E}[\boldsymbol{y}_i\boldsymbol{z}_i] = \mathbf{E}[\boldsymbol{x}_i\boldsymbol{z}_i] = \delta$; further, it is clear from (1) that $\mathbf{E}[\boldsymbol{x}_i\boldsymbol{y}_i\boldsymbol{z}_i] = -(1-\delta)$ (the $\mathcal{D}_{\mathtt{XOR}}$ part contributes -1 and the $\mathcal{D}_{\mathtt{EQU}}$ part contributes 0). Hence in (3) we get a contribution only if A, B, C cover each index exactly 0, 2, or 3 times, with contributions of 1, δ, and $-(1-\delta)$, respectively. This means that the sets must be expressible as $S \cup P \cup Q$, $S \cup Q \cup R$, and $S \cup P \cup R$ for disjoint S, P, Q, R. Hence (3) becomes

$$\sum_{S \subseteq [n]} (-1)^{|S|}(1-\delta)^{|S|} \sum_{\substack{P,Q,R \text{ disj} \\ \text{and disj from } S}} \delta^{|P|+|Q|+|R|}\hat{f}(S \cup P \cup Q)\hat{f}(S \cup Q \cup R)\hat{f}(S \cup P \cup R). \tag{4}$$

Since f is odd, the product of Fourier coefficients is 0 unless each of the sets $S \cup P \cup Q$, $S \cup Q \cup R$, $S \cup P \cup R$ have odd cardinality; it is easy to see that this forces $|S|$ to be odd (and also $|P|, |Q|, |R|$ to be even, but we will not use this).

Next, let us write $\overline{S} = [n] \setminus S$ and introduce the function $F_S : \{-1,1\}^{\overline{S}} \to [-1,1]$ defined by

$$F_S(u) = \widehat{f_{u \to \overline{S}}}(S);$$

i.e., $F_S(u)$ is the correlation with parity-on-S of the restriction of f given by substituting u into the coordinates $\overline{S}$. One can check that for any $V \subseteq \overline{S}$,

$$\widehat{F_S}(V) = \hat{f}(S \cup V).$$

Thus we may write (4) as

$$\begin{aligned}
&\sum_{|S| \text{ odd}} (-1)^{|S|}(1-\delta)^{|S|} \sum_{\substack{P,Q,R \text{ disj} \\ \text{and disj from } S}} \delta^{|P|+|Q|+|R|}\widehat{F_S}(P \cup Q)\widehat{F_S}(Q \cup R)\widehat{F_S}(P \cup R) \\
= \; & -\sum_{|S| \text{ odd}} (1-\delta)^{|S|} \sum_{\substack{P,Q,R \text{ disj} \\ \text{and disj from } S}} \widehat{T_{\sqrt{\delta}}F_S}(P \cup Q)\widehat{T_{\sqrt{\delta}}F_S}(Q \cup R)\widehat{T_{\sqrt{\delta}}F_S}(P \cup R) \\
= \; & -\sum_{|S| \text{ odd}} (1-\delta)^{|S|} \mathop{\mathbf{E}}_{\boldsymbol{u}}[(T_{\sqrt{\delta}}F_S)(\boldsymbol{u})^3],
\end{aligned}$$

where in the final expression $\boldsymbol{u}$ has the uniform distribution on $\{-1,1\}^{\overline{S}}$. Summarizing, we have

$$-\mathbf{E}[f(\boldsymbol{x})f(\boldsymbol{y})f(\boldsymbol{z})] = \sum_{|S| \text{ odd}} (1-\delta)^{|S|} \mathop{\mathbf{E}}_{\boldsymbol{u}}[(T_{\sqrt{\delta}}F_S)(\boldsymbol{u})^3]. \tag{5}$$

So far we have made only elementary Fourier manipulations; we now come to the more interesting part of the proof. Recall that

$$T_{\sqrt{\delta}}F_S = \widehat{F_S}(\emptyset) + \sqrt{\delta}\sum_{|V|=1}\widehat{F_S}(V)\chi_V + \delta\sum_{|V|=2}\widehat{F_S}(V)\chi_V + \cdots .$$

For δ very small, one feels that $T_{\sqrt{\delta}}F_S$ should almost be a constant function; specifically, the constant $\widehat{F_S}(\emptyset) = \hat{f}(S)$. If this were correct we would get that (5) is at most

$$\sum_{|S| \text{ odd}} (1-\delta)^{|S|}|\hat{f}(S)|^3. \tag{6}$$

This is precisely the critical term in Håstad's Dictator-vs.-Quasirandom Test based on XOR, so we know it will be small. Specifically,

$$\begin{aligned}(6) &\leq \max_{|S| \text{ odd}}\{(1-\delta)^{|S|}|\hat{f}(S)|\} \cdot \sum_{|S| \text{ odd}} \hat{f}(S)^2 \\ &\leq \max\left\{\max_{0<|S|\leq \ln(1/\delta)/\delta}|\hat{f}(S)|, (1-\delta)^{\ln(1/\delta)/\delta}\right\} \leq \max\{\sqrt{\delta},\delta\} = \sqrt{\delta}, \end{aligned} \tag{7}$$

where we used Parseval, and the second-to-last step used that f is $(\delta, \delta/\ln(1/\delta))$-quasirandom.

We now need to make this more precise and understand the deviation of F_S from its mean. Write $\widetilde{F}_S = F_S - \hat{f}(S)$, so that $\mathbf{E}[\widetilde{F}_S] = 0$. Then

$$\begin{aligned}(5) &= \sum_{|S| \text{ odd}} (1-\delta)^{|S|} \mathop{\mathbf{E}}_{\boldsymbol{u}}[(\hat{f}(S) + (T_{\sqrt{\delta}}\widetilde{F}_S)(\boldsymbol{u}))^3] \\ &\leq \sum_{|S| \text{ odd}} (1-\delta)^{|S|} \cdot 4(|\hat{f}(S)|^3 + \mathop{\mathbf{E}}_{\boldsymbol{u}}[|(T_{\sqrt{\delta}}\widetilde{F}_S)(\boldsymbol{u})|^3]) \qquad (\text{using } (a+b)^3 \leq 4(|a|^3+|b|^3)) \\ &= 4\sum_{|S| \text{ odd}} (1-\delta)^{|S|}|\hat{f}(S)|^3 + 4\sum_{|S| \text{ odd}} (1-\delta)^{|S|}\|T_{\sqrt{\delta}}\widetilde{F}_S\|_3^3 \\ &\leq 4\sqrt{\delta} + 4\sum_{|S| \text{ odd}} (1-\delta)^{|S|}\|T_{\sqrt{\delta}}\widetilde{F}_S\|_3^3, \end{aligned}$$

where we used (7). Combining this with (2), we conclude

$$\mathbf{E}[\mathtt{NTW}(f(\boldsymbol{x}), f(\boldsymbol{y}), f(\boldsymbol{z}))] \leq \tfrac{5}{8} + \tfrac{3}{8}\delta + \tfrac{3}{2}\sqrt{\delta} + \tfrac{3}{2}\sum_{|S| \text{ odd}} (1-\delta)^{|S|}\|T_{\sqrt{\delta}}\widetilde{F}_S\|_3^3, \tag{8}$$

and it remains to analyze the final term,

$$\sum_{|S| \text{ odd}} (1-\delta)^{|S|}\|T_{\sqrt{\delta}}\widetilde{F}_S\|_3^3. \tag{9}$$

To do this we will use Corollary 5.2. First, though, we point out why a more naive bound would not be useful. Suppose we were to use

$$\|T_{\sqrt{\delta}}\widetilde{F}_S\|_3^3 = \mathbf{E}[|T_{\sqrt{\delta}}\widetilde{F}_S|^3] \leq \|T_{\sqrt{\delta}}\widetilde{F}_S\|_\infty \cdot \mathbf{E}[|T_{\sqrt{\delta}}\widetilde{F}_S|^2] \leq 2\|T_{\sqrt{\delta}}\widetilde{F}_S\|_2^2,$$

where we invoked the simple bound

$$\|T_{\sqrt{\delta}}\widetilde{F}_S\|_\infty \le \|T_{\sqrt{\delta}}F_S\|_\infty + |\hat{f}(S)| \le \|F_S\|_\infty + |\hat{f}(S)| \le \|f\|_\infty + \|f\|_\infty \le 2.$$

Now

$$\|T_{\sqrt{\delta}}\widetilde{F}_S\|_2^2 = \sum_{\emptyset \ne V \subseteq \overline{S}} \delta^{|V|}\widehat{F_S}(V)^2 = \sum_{\emptyset \ne V \subseteq \overline{S}} \delta^{|V|}\hat{f}(S \cup V)^2, \tag{10}$$

and thus we would get

$$(9) \le 2 \sum_{|S| \text{ odd}} (1-\delta)^{|S|} \sum_{\emptyset \ne V \subseteq \overline{S}} \delta^{|V|}\hat{f}(S \cup V)^2 = 2 \sum_{U \subseteq [n]} \hat{f}(U)^2 \Pr_{\boldsymbol{S} \subseteq_{1-\delta} U}[|\boldsymbol{S}| \text{ odd and } \boldsymbol{S} \ne U] \tag{11}$$

where $\boldsymbol{S} \subseteq_{1-\delta} U$ means $\boldsymbol{S}$ is a subset of U chosen by including each coordinate with probability $1-\delta$. But for $|U| \gg 1/\delta$,

$$\Pr_{\boldsymbol{S} \subseteq_{1-\delta} U}[|\boldsymbol{S}| \text{ odd and } \boldsymbol{S} \ne U] \approx 1/2,$$

and there is no way we could get an upper bound better than 1 on $\sum_{|U| \gg 1/\delta} \hat{f}(U)^2$, even assuming f is quasirandom. We can also note that it is extremely important not to increase the $\sqrt{\delta}$ in any estimation on $\|T_{\sqrt{\delta}}\widetilde{F}_S\|_3^3$; if we had any $\delta' > \delta$ in the above failed estimate, we wouldn't have a probability distribution on subsets of U, and wouldn't even be able to get an upper bound of 1.

With these issues explained, we now proceed with applying Corollary 5.2 to (9). Let λ be such that

$$\sqrt{\delta}^{\lambda} = 1/\sqrt{3-1} = 1/\sqrt{2}. \tag{12}$$

By our assumption that $0 < \delta < 1/8$ we have $0 < \lambda < 1/3$. Now the corollary implies

$$(9) \le \sum_{|S| \text{ odd}} (1-\delta)^{|S|} \; \|T_{\sqrt{\delta}}\widetilde{F}_S\|_2^{3-3\lambda} \; \|\widetilde{F}_S\|_2^{3\lambda}. \tag{13}$$

We use

$$\|\widetilde{F}_S\|_2^{3\lambda} = (\|\widetilde{F}_S\|_2^2)^{3\lambda/2} = \left(\sum_{\emptyset \ne V \subseteq \overline{S}} \hat{f}(S \cup V)^2\right)^{3\lambda/2} \le (\|f\|_2)^{3\lambda/2} \le 1$$

and

$$\begin{aligned}
\|T_{\sqrt{\delta}}\widetilde{F}_S\|_2^{3-3\lambda} &= \|T_{\sqrt{\delta}}\widetilde{F}_S\|_2^2 \cdot \|T_{\sqrt{\delta}}\widetilde{F}_S\|_2^{1-3\lambda} \\
&\le \left(\sum_{\emptyset \ne V \subseteq \overline{S}} \delta^{|V|}\hat{f}(S \cup V)^2\right) \cdot \sqrt{\delta}^{1-3\lambda} && \text{(using (10) and } \lambda < 1/3\text{)} \\
&= 2^{3/2}\sqrt{\delta} \cdot \sum_{\emptyset \ne V \subseteq \overline{S}} \delta^{|V|}\hat{f}(S \cup V)^2 && \text{(using (12)).}
\end{aligned}$$

Substituting these into (13) and using the calculation in (11) yields

$$(9) \le 2^{3/2}\sqrt{\delta} \cdot \sum_{U \subseteq [n]} \hat{f}(U)^2 \Pr_{\boldsymbol{S} \subseteq_{1-\delta} U}[|\boldsymbol{S}| \text{ odd and } \boldsymbol{S} \ne U] \le 2^{3/2}\sqrt{\delta} \cdot \sum_{U \subseteq [n]} \hat{f}(U)^2 \le 2^{3/2}\sqrt{\delta}.$$

Finally, substituting this into (8) yields

$$\mathbf{E}[\mathtt{NTW}(f(\boldsymbol{x}), f(\boldsymbol{y}), f(\boldsymbol{z}))] \le \tfrac{5}{8} + \tfrac{3}{8}\delta + \tfrac{3}{2}\sqrt{\delta} + 3\sqrt{2}\sqrt{\delta} \le \tfrac{5}{8} + 6\sqrt{\delta},$$

as claimed. □

6 Future work

Despite constructing the required Dictator-vs.-Quasirandom Test, we are unable to prove that satisfiable Max-NTW instances are $(5/8+\epsilon)$-hard to approximate. There is strong evidence that the formal encapsulation of the Rule of Thumb described at the beginning of Section 3.1 cannot be used in our case. The reason is as follows: Our Dictator-vs.-Quasirandom Test from Section 5 has the property that under $\mathcal{D}_\delta$, each pair of bits has small but nonzero correlation δ. As discussed in Appendix D, the correct test seems forced and thus it appears there is no way to avoid this pairwise correlation. But this would seem to rule out constructing the necessary two-function Dictator-vs.-Quasirandom Test. To see this, suppose we try to generalize $\mathcal{D}_\delta$ to a distribution on $(2d+1)$-tuples $(a, \vec{b}, \vec{c}) \in \{-1,1\} \times \{-1,1\}^d \times \{-1,1\}^d$. One can see that the marginal on each triple $(a, \vec{b}_\ell, \vec{c}_\ell)$ must be $\mathcal{D}_\delta$: this is because the g function could always "ignore" all but the ℓth index in each of its index groups J_i. But recall that each $\vec{b}_\ell$ has correlation δ with a. If $d = |J_i| \gg 1/\delta$, then by taking a majority over each of its index groups J_i, the g function can effectively "know" the bit a and hence confound the test.

Thus it seems there is little hope for proving the desired NP-hardness result with today's PCP technology. Further, proving the result based on the Unique Games Conjecture is also out of the question, since Unique-Games inherently has imperfect completeness. Only one window of opportunity remains: trying to prove the hardness result assuming Khot's "d-to-1 Conjecture" [16] for some constant d. This would give us both the perfect completeness we need and let us make $\delta \ll 1/d$ in order to overcome the small pairwise correlation in the test. However, the Dictator-vs.-Quasirandom Testing needed would become significantly more complicated to analyze.

In a forthcoming paper, we will indeed show Zwick's conjecture $\mathrm{NP} \subseteq \mathrm{naPCP}_{1,5/8+\epsilon}(O(\log n), 3)$, assuming the d-to-1 Conjecture for any constant d.

7 Acknowledgment

We are grateful to an anonymous reviewer for significantly simplifying our original proof of Corollary 5.2.

References

[1] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs, and nonapproximability — towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

[2] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computing and Sys. Sci.*, 47(3):549–595, 1993.

[3] A. Bonami. Ensembles $\Lambda(p)$ dans le dual de D^∞. *Ann. Inst. Fourier*, 18(2):193–204, 1968.

[4] I. Dinur. The PCP Theorem by gap amplification. *Journal of the ACM*, 54(3):12, 2007.

[5] L. Engebretsen and J. Holmerin. Three-query PCPs with perfect completeness over non-Boolean domains. *Random Structures & Algorithms*, 27(1):46–75, 2005.

[6] M. Furst, J. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. *Math. Systems Theory*, 17(1):13–27, 1984.

[7] B. Green. Finite field models in additive combinatorics. In *London Mathematical Society Lecture Note Series, Cambridge University Press*, volume 324. 2005.

[8] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97:1061–1083, 1975.

[9] V. Guruswami and S. Khot. Hardness of max 3SAT with no mixed clauses. In *Proc. 20th IEEE Conference on Computational Complexity*, pages 154–162, 2005.

[10] V. Guruswami, D. Lewin, M. Sudan, and L. Trevisan. A tight characterization of NP with 3 query PCPs. *Electronic Colloq. on Comp. Complexity (ECCC)*, 5(34), 1998.

[11] J. Håstad. Some optimal inapproximability results. In *Proc. 29th ACM Symp. on the Theory of Computing*, pages 1–10, 1997.

[12] J. Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Mathematica*, 182(1):105–142, 1999.

[13] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[14] J. Håstad. On nontrivial approximations of CSPs, 2006. http://www.nada.kth.se/∼johanh/approx06.pdf. Invited address at APPROX.

[15] H. Karloff and U. Zwick. A 7/8-approximation algorithm for Max-3Sat? In *Proc. 38th IEEE Symp. on Foundations of Comp. Sci.*, pages 406–415, 1997.

[16] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symp. on the Theory of Computing*, pages 767–775, 2002.

[17] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[18] S. Khot and R. O'Donnell. SDP gaps and UGC-hardness for MaxCutGain. In *Proc. 47th IEEE Symp. on Foundations of Comp. Sci.*, pages 217–226, 2006.

[19] S. Khot and R. Saket. A 3-query non-adaptive PCP with perfect completeness. In *Proc. 21st IEEE Conference on Computational Complexity*, pages 159–169, 2006.

[20] S. Khot and N. Vishnoi. The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *Proc. 46th IEEE Symp. on Foundations of Comp. Sci.*, pages 53–62, 2005.

[21] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th IEEE Symp. on Foundations of Comp. Sci.*, pages 21–30, 2005. To appear, *Annals of Mathematics*.

[22] R. O'Donnell. Analysis of boolean functions lecture notes, 2007. http://www.cs.cmu.edu/ odonnell/boolean-analysis/lecture4.pdf.

[23] R. O'Donnell and Y. Wu. An optimal SDP algorithm for Max-Cut, and equally optimal Long Code tests. In *Proc. 40th ACM Symp. on the Theory of Computing*, pages 335–344, 2008.

[24] M. Parnas, D. Ron, and A. Samorodnitsky. Testing basic boolean formulae. *SIAM Journal on Discrete Mathematics*, 16(1):20–46, 2002.

[25] P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proc. 40th ACM Symp. on the Theory of Computing*, pages 245–254, 2008.

[26] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proc. 38th ACM Symp. on the Theory of Computing*, pages 11–20, 2006.

[27] D. Stefankovic. Fourier transforms in computer science. Master's thesis, University of Chicago, 2002.

[28] L. Trevisan. Approximating satisfiable satisfiability problems. *Algorithmica*, 28(1):145–172, 2000.

[29] L. Trevisan, G. Sorkin, M. Sudan, and D. Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000.

[30] U. Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proc. 9th ACM-SIAM Symp. on Discrete Algorithms*, pages 201–210, 1998.

[31] U. Zwick. Computer assisted proof of optimal approximability results. In *Proc. 13th ACM-SIAM Symp. on Discrete Algorithms*, pages 496–505, 2002.

A Dictator-vs.-Quasirandom Testing lower bound

In this section we prove our Dictator-vs.-Quasirandom Testing lower bound, completing the proof of our main Theorem 2.1. The proof is by viewing any 3-query Dictator-vs.-Quasirandom Test with perfect completeness as a satisfiable instance of Max-3CSP over the "variables" $f(x)$, as discussed in Section 1.3. We then use a variation of Zwick's algorithm [30] to construct a quasirandom (even Gaussianic) function f passing the test with probability at least $5/8 - o(1)$.

Interestingly, the quasirandom function $\boldsymbol{f} : \{-1,1\}^n \to \{-1,1\}$ we construct will always be either:

(a) an odd-size parity, chosen uniformly from among the 2^{n-1} possibilities; or,
(b) a random threshold function, negated if necessary so that $\boldsymbol{f}(1,1,\dots,1) = 1$.

In (b), we mean $\boldsymbol{f}$ is chosen as

$$\boldsymbol{f}(x_1,\dots,x_n) = \mathrm{sgn}(\textstyle\sum_{i=1}^n \boldsymbol{G}_i) \cdot \mathrm{sgn}(\textstyle\sum_{i=1}^n \boldsymbol{G}_i x_i), \tag{14}$$

where the $\boldsymbol{G}_i$'s are independent standard Gaussians.

Let's first note that these two possibilities are indeed quasirandom (and even Gaussianic) with high probability:

Lemma A.1. *Let $\boldsymbol{f} : \{-1,1\}^n \to \{-1,1\}$ be an odd-size parity, chosen uniformly from among the 2^{n-1} possibilities. Assuming $n \geq O(1/\delta)$ we have that $\boldsymbol{f}$ is $(0,\delta)$-quasirandom except with probability at most $2^{-\Omega(n)}$. Assuming $n \geq O(\ln(e/\epsilon)/\delta)$ we have that $\boldsymbol{f}$ is (ϵ,δ)-Gaussianic except with probability at most $2^{-\Omega(n)}$.*

Lemma A.2. *Let $\boldsymbol{f} : \{-1,1\}^n \to \{-1,1\}$ be a random threshold function chosen as in* (14). *Assuming $n \geq O(1/\epsilon^4)$ we have that $\boldsymbol{f}$ is both $(\epsilon, 0)$-quasirandom and $(\epsilon, 0)$-Gaussianic except with probability at most $2^{-n^{\Omega(1)}}$.*

The proofs of these lemmas are straightforward probabilistic considerations and appear in Appendix E. (A variant of the latter appears in the full version of [23].) We can now give the proof of our Dictator-vs.-Quasirandom/Gaussianic testing lower bound:

Theorem A.3. *Let $\epsilon, \delta, \eta > 0$. Then for all $n \geq \mathrm{poly}(1/\epsilon, 1/\delta, \log(1/\eta))$, any dictator-vs.-Gaussianic test $\mathcal{T}$ on functions $f : \{-1,1\}^n \to \{-1,1\}$ with perfect completeness has (ϵ, δ)-soundness at least $5/8 - \eta$.*

Proof. Let the test $\mathcal{T}$ be given. As mentioned, in the end we will consider choosing $\boldsymbol{f}$ to be a random odd parity, and will also consider choosing $\boldsymbol{f}$ to be a random threshold function, negated if necessary so that $\boldsymbol{f}(1, 1, \ldots, 1) = 1$. We will show that at least one of these two options has the property that in expectation, $\boldsymbol{f}$ satisfies at least 5/8 of the constraints tested by $\mathcal{T}$. If we then condition on $\boldsymbol{f}$ being (ϵ, δ)-Gaussianic, Lemmas A.1, A.2 and the condition on n imply that the conditional expectation of the fraction of constraints $\boldsymbol{f}$ satisfies is at least $5/8 - \eta$. Thus there exists an (ϵ, δ)-Gaussianic function passing the test with probability at least $5/8 - \eta$, as necessary.

Let us also mention that with this strategy, every f we might end up choosing is odd and satisfies $f(1, 1, \ldots, 1) = 1$. Given this, we will begin by altering $\mathcal{T}$ so that it "assumes oddness of f"; i.e., we "fold" it. More precisely, whenever a constraint in $\mathcal{T}$ involves some $f(-1, x')$ (for $x' \in \{-1,1\}^{n-1}$), we replace it with $-f(1, -x')$. This preserves the perfect completeness of $\mathcal{T}$ on dictators since they are odd. Further, it does not change the probability of satisfaction for any f we eventually select, since all such f's will be odd. Henceforth, we think of $\mathcal{T}$ as a 3-CSP only over the 2^{n-1} "variables" $f(1, x)$, $x \in \{-1,1\}^{n-1}$.

We emphasize that $\mathcal{T}$ could use any mixture of the twenty-two constraints on up to 3 bits. Actually, $\mathcal{T}$ cannot use $\texttt{FLS}_0$, since it has perfect completeness. As an illustration, $\mathcal{T}$ might query and test...

$$\begin{array}{lllll} \texttt{AND}_2(& -f(x^{(1)}), & -f(y^{(1)}) &) & \text{with probability} \quad p_1, \\ \texttt{AOR}(& -f(x^{(2)}), & -f(y^{(2)}), & f(z^{(2)})) & \text{with probability} \quad p_2, \\ \texttt{MAJ}(& -f(x^{(3)}), & f(y^{(3)}), & -f(z^{(3)})) & \text{with probability} \quad p_3, \\ \texttt{AXR}(& -f(x^{(4)}), & -f(y^{(4)}), & f(z^{(4)})) & \text{with probability} \quad p_4, \\ & & \cdots & & \end{array}$$

Let's say $\mathcal{T}$ uses m different constraints, $\mathcal{C}_1, \ldots, \mathcal{C}_m$.

Similar to Zwick's algorithm, we can begin by quickly eliminating the constraints that "force identities": namely, the constraints $\texttt{TRU}_0$, $\texttt{IDN}_1$, $\texttt{AND}_2$, $\texttt{XOR}_2$, `AND`, `EQU`, `AXR`, `AOR`, and `AOA`. The first seven of these have the following property: Any constraint in $\mathcal{T}$ using one of them is automatically satisfied by any f with $f(1, 1, \ldots, 1) = 1$. As an example, consider the constraint

$$\texttt{AXR}(-f(x^{(4)}), -f(y^{(4)}), f(z^{(4)})) = \neg f(x^{(4)}) \wedge (\neg f(y^{(4)}) \oplus f(z^{(4)})).$$

Since all dictators satisfy this constraint, we must have $x_i^{(4)} = 1$ and $-y_i^{(4)} z_i^{(4)} = -1$ for all i; i.e., it must be the case that $x^{(4)} = (1, \ldots, 1)$, $y^{(4)} = z^{(4)}$. But then $f(x^{(4)}) = 1$ whenever

$f(1,1,\ldots,1) = 1$, and of course $f(y^{(4)}) = f(z^{(4)})$; hence the $\mathtt{AXR}$ constraint is automatically satisfied whenever $f(1,1,\ldots,1) = 1$. Notice that we could *not* have had the constraint, say,

$$\mathtt{AND}_2(-f(x^{(1)}), f(y^{(1)}));$$

by our folding convention $x_1^{(1)} = y_1^{(1)} = 1$, and then the first dictator would not satisfy this constraint. Showing that $\mathtt{TRU}_0$, $\mathtt{IDN}_1$, $\mathtt{AND}_2$, $\mathtt{XOR}_2$, $\mathtt{AND}$, and $\mathtt{EQU}$ constraints are also automatically satisfied by any f with $f(1,1,\ldots,1) = 1$ is easier and is left to the reader.

The other two constraints mentioned, $\mathtt{AOR}$ and $\mathtt{AOA}$, have a related property. Consider a test like

$$\mathtt{AOR}(-f(x^{(2)}, -f(y^{(2)}), f(z^{(2)})) = \neg f(x^{(2)}) \wedge (-\neg f(y^{(2)}) \vee f(z^{(2)})).$$

Again, as all dictators pass this test, $x^{(2)}$ must be $(1,\ldots,1)$. Thus assuming $f(1,1,\ldots,1) = 1$, we may safely replace this constraint with

$$\mathtt{OR}_2(-f(y^{(2)}), f(z^{(2)})).$$

A similar consideration lets us replace all $\mathtt{AOA}$ constraints to $\mathtt{OR}_2$ constraints. In conclusion, after this initial stage, we may assume $\mathcal{T}$ has no $\mathtt{TRU}_0$, $\mathtt{IDN}_1$, $\mathtt{AND}_2$, $\mathtt{XOR}_2$, $\mathtt{AND}$, $\mathtt{EQU}$, $\mathtt{AXR}$, $\mathtt{AOR}$, or $\mathtt{AOA}$ constraints, just as in Zwick's algorithm.

The next step of the proof (based on an idea of [28]) also appears in Zwick's algorithm, albeit at a later stage. To describe this step, we temporarily switch our representation convention for False and True: instead of $1, -1 \in \mathbb{R}$, we view them as $0, 1 \in \mathbb{F}_2$, the field with two elements. Our variable space is now $\mathbb{F}_2^{\{-1,1\}^n}$. Let P be the affine subspace consisting of all 2^{n-1} *odd*-sized parity functions; this includes the dictators, which we denote by χ_i. One can think of P as being $\chi_1 \oplus V$, where V is the the $(n-1)$-dimensional subspace spanned by $\chi_1 \oplus \chi_2$, $\chi_1 \oplus \chi_3$, $\ldots$, $\chi_1 \oplus \chi_n$.

Consider choosing an odd parity $\boldsymbol{f}$ uniformly randomly from P. We claim that in doing so, each constraint $\mathcal{C}_j$ has at least a certain probability of being satisfied based on its type, as shown in the following table:

TWO	XOR	MAJ	XAD	SEL	OAD	XOA	NTW	$\mathtt{OR}_2$	NAE	OXR	OR
3/4	1	1/2	1/2	1/2	5/8	5/8	5/8	3/4	3/4	3/4	3/4

(Zwick has 7/8 in his table for $\mathtt{OR}$; we are using a different argument.) To see this claim, consider some constraint in $\mathcal{T}$, perhaps $\mathcal{C}_j = \phi(f(u), \neg f(v), \neg f(w))$. (Assume for now that $\phi \neq \mathtt{OR}_2$, so ϕ has arity 3.) Since P is an affine subspace and $\boldsymbol{f}$ is chosen uniformly from P, the induced distribution on $(\boldsymbol{f}(u), \neg\boldsymbol{f}(v), \neg\boldsymbol{f}(w))$ will also be uniform on some affine subspace Q of $\mathbb{F}_2^3$. Furthermore, since each choice of $\boldsymbol{f}$ is an odd parity, this Q will be precisely the "affine span" of the triples $\{(u_i, \neg v_i, \neg w_i)\}_{i\in[n]}$. Note also that each of these triples *satisfies* ϕ, since χ_i satisfies $\mathcal{C}_j$. It is now easy to see that for each of the possible constraints (including the 2-ary $\mathtt{OR}_2$), the uniform distribution on any affine span of satisfying assignments leads to an acceptance probability which is at least what's shown in the table. The key point is that for $\mathtt{TWO}$ and $\mathtt{XOR}$, the affine subspace Q *cannot* be all of $\mathbb{F}_2^3$.

This concludes our analysis of choosing $\boldsymbol{f}$ to be a random odd parity; let us revert back to ± 1 notation. We now come to the analysis of choosing $\boldsymbol{f}$ to be a random threshold function. For this, we follow the semidefinite programming part of Zwick's algorithm. As in his algorithm, we formulate the "canonical semidefinite programming relaxation" [15] of the weighted 3-CSP $\mathcal{T}$, but we have a few slight differences. The canonical SDP introduces a unit vector $v_{f(x)}$ for each variable $f(x)$ (with $x \in \{1\} \times \{-1,1\}^{n-1}$). It also introduces a unit vector representing the negation of each variable; for us, we can simply extend the notation $v_{f(x)}$ to each $x \in \{-1\} \times \{-1,1\}^{n-1}$ and include the canonical constraint $v_{f(x)} = -v_{f(-x)}$. The SDP relaxation also introduces a unit vector "v_0" intended to represent False;[1] we will instead take this vector to be $v_{f(e)}$, where e denotes $(1, 1, \ldots, 1)$ (since we want $f(e) = 1$ anyway). Finally, the SDP relaxation includes a scalar variable Z_j for each constraint, intended to represent whether the constraint is satisfied.

The critical observation is this: The fact that all dictators satisfy every constraint in $\mathcal{T}$ implies that setting $v_{f(x)} = x/\sqrt{n}$ for each $x \in \{-1,1\}^n$ and $Z_j = 1$ for each j yields a feasible solution to the SDP relaxation with value 1. (One normally thinks of the vectors as having dimension equal to the number of variables, but it is okay if they have smaller dimension.) To see this, consider an inequality in the SDP arising from the jth constraint in $\mathcal{T}$; it might look something like this:

$$Z_j \leq \frac{3}{4} - \frac{1}{2} v_{f(e)} \cdot v_{f(y)} + \frac{1}{4} v_{f(y)} \cdot v_{f(z)} - \frac{1}{4} v_{f(x)} \cdot v_{f(z)} + \frac{1}{4} v_{f(y)} \cdot v_{f(z)}.$$

(This example occurs when the jth constraint is $\mathtt{MAJ}(-f(x), -f(y), f(z))$.) We would like to show our claimed solution satisfies this inequality; i.e., to show

$$1 \leq \frac{3}{4} - \frac{1}{2} \frac{\sum_{i=1}^n e_i y_i}{n} + \frac{1}{4} \frac{\sum_{i=1}^n y_i z_i}{n} - \frac{1}{4} \frac{\sum_{i=1}^n x_i z_i}{n} + \frac{1}{4} \frac{\sum_{i=1}^n y_i z_i}{n}. \tag{15}$$

But *by design* of the relaxation, the inequality

$$1 \leq \frac{3}{4} - \frac{1}{2} \cdot 1 \cdot b + \frac{1}{4} \cdot b \cdot c - \frac{1}{4} \cdot a \cdot c + \frac{1}{4} \cdot b \cdot c$$

is satisfied whenever (a, b, c) is a triple of bits satisfying the jth constraint. Since all dictators pass this constraint, (x_i, y_i, z_i) is satisfying and hence we conclude

$$1 \leq \frac{3}{4} - \frac{1}{2} e_i y_i + \frac{1}{4} y_i z_i - \frac{1}{4} x_i z_i + \frac{1}{4} y_i z_i$$

for each $i \in [n]$. Averaging this inequality across i confirms 15. The fact that the value of this solution is 1 follows immediately from the fact that all Z_j's are 1.

Given this optimal feasible solution to the SDP relaxation we may apply the randomized rounding analysis of Zwick. This involves choosing a random hyperplane — say with normal $\boldsymbol{G} = (\boldsymbol{G}_1, \ldots, \boldsymbol{G}_n)$, where the $\boldsymbol{G}_i$'s are independent standard Gaussians — and then rounding $v_{f(x)}$ to $\mathrm{sgn}(\boldsymbol{G} \cdot v_{f(e)}) \mathrm{sgn}(\boldsymbol{G} \cdot v_{f(x)})$. In other words, one chooses the solution

$$\boldsymbol{f}(x) = \mathrm{sgn}(\textstyle\sum \boldsymbol{G}_i) \cdot \mathrm{sgn}(\sum \boldsymbol{G}_i x_i),$$

just as promised. Zwick's SDP-rounding analysis shows that the expected fraction of constraints of each type that $\boldsymbol{f}$ satisfies is as shown in the following table:

TWO	XOR	MAJ	XAD	SEL	OAD	XOA	NTW	$\mathtt{OR}_2$	NAE	OXR	OR
.649	1/2	.736	.736	5/6	.824	.824	5/8	.912	.912	3/4	7/8

[1] In [30] it represents True, but his convention is that True is 1.

Finally, we use Zwick's observation that by choosing $\boldsymbol{f}$ to be a random odd parity with probability 1/4 and to be a random threshold as above with probability 3/4, each constraint type is satisfied with probability at least 5/8. Hence at least a 5/8 fraction of all constraints are satisfied in expectation for this mixed distribution on $\boldsymbol{f}$, and hence one of the two distributions indeed is expected to satisfy at least 5/8 of the constraints. □

B Quasirandom and Gaussianic functions

Functions which are $(\epsilon, 0)$-quasirandom — i.e., satisfy $\hat{f}(S)^2 \le \epsilon$ for *every* $S \neq \emptyset$ — are sometimes called "pseudorandom", "regular", or "uniform"; see, e.g., [7]. Functions which are $(\epsilon, 1)$-quasirandom — i.e., satisfy $\hat{f}(i)^2 \le \epsilon$ for all $i \in [n]$ — were called "flat" in [18], and when f is monotone this notion is equivalent to having all "influences" at most $\sqrt{\epsilon}$. In the full version of [23] the present authors introduced a related notion of a function being "(ϵ, δ)-Gaussianic"; this means the function satisfies

$$\sum_{\substack{S \subseteq [n] \\ i \in S}} (1-\delta)^{|S|-1} \hat{f}(S)^2 \le \epsilon \quad \text{for all } i \in [n].$$

Being (ϵ, δ)-Gaussianic is very similar to having all "$(1/\delta)$-low-degree influences" at most ϵ, a condition introduced and studied in [17, 21]. Note that being Gaussianic is essentially a stronger property than being quasirandom: it is easy to check that an (ϵ, δ)-Gaussianic function is also $(e \cdot \epsilon, \delta)$-quasirandom. The converse is not true, though: for example, the function $f(x) = x_1 \oplus \text{Majority}(x_2, \ldots, x_n)$ is $(\Theta(1/\sqrt{n}), 0)$-quasirandom but is not even $(\frac{2}{\pi}\eta, 1-\eta)$-Gaussianic for any $\eta > 0$.

We define Dictator-vs.-Gaussianic tests and their completeness and soundness by analogy with our definitions of Dictator-vs.-Quasirandom Tests.

C Testing averages

In this section we explain the trick of testing *averages* of functions, introduced in [1, 17].

Definition C.1. *Let $\mathcal{F} = \{f_1, \ldots, f_t\}$ be a collection of functions $\{-1,1\}^n \to \{-1,1\}$, and let $\mathcal{T}$ be a Dictator-vs.-Quasirandom Test for such functions. We define the notion of $\mathcal{T}$ testing the* average *of $\mathcal{F}$ as follows: whenever $\mathcal{T}$ is about to query and test $\phi(f(x_1), \ldots, f(x_k))$, it first chooses $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_k$ uniformly and independently from $[t]$ and then queries and tests $\phi(f_{\boldsymbol{i}_1}(x_1), \ldots, f_{\boldsymbol{i}_k}(x_k))$. In the special case that $\mathcal{F} = \{f, f^\dagger\}$, where $f^\dagger(x) := -f(-x)$, we say that $\mathcal{T}$* tests f^{odd}. *Note that $\mathcal{T}$ only needs query-access to f in order to test f^{odd}.*

The trick of testing f^{odd} is precisely the idea of "folding" introduced by [1]. To understand the probability of $\mathcal{F}$ passing test $\mathcal{T}$ we extend all predicates to maps $\phi : [-1,1] \to [0,1]$ via multilinear extension:

$$\texttt{XOR}(a,b,c) = \tfrac{1}{2} - \tfrac{1}{2}abc, \quad \texttt{NTW}(a,b,c) = \tfrac{5}{8} + \tfrac{1}{8}(a+b+c) + \tfrac{1}{8}(ab+bc+ac) - \tfrac{3}{8}abc, \qquad (16)$$

etc. Then by linearity of expectation one can easily check that that $\mathcal{T}$ passes $\mathcal{F}$ with probability

$$\mathop{\mathbf{E}}_{\boldsymbol{\phi}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \sim T}[\boldsymbol{\phi}(\overline{f}(\boldsymbol{x}_1), \ldots, \overline{f}(\boldsymbol{x}_k))],$$

where $\overline{f} : \{-1,1\}^n \to [-1,1]$ is the (pointwise) average of the functions in $\mathcal{F}$. When $\mathcal{F} = \{f, f^\dagger\}$ we write the average $\overline{f}$ as

$$f^{\text{odd}} = \sum_{\substack{S \subseteq [n] \\ |S| \text{ odd}}} \hat{f}(S) \chi_S.$$

Since Dictator-vs.-Quasirandom Testing upper-bound analysis invariably uses Fourier analysis, almost nothing changes if the function f being tested is of the form $f : \{-1,1\}^n \to [-1,1]$. Indeed, all of our tests will allow for such functions, and hence allow for the testing averages.

D Intuition for our Dictator-vs.-Quasirandom Test

In this section, we give an explanation as to how the distribution $\mathcal{D}_\delta$ was chosen, and also explain why the prototypical quasirandom functions — constants, Majority, and large parities — pass our test with probability only with probability about 5/8.

Since dictators are to pass our test with probability 1, clearly whatever distribution $\mathcal{D}_\delta$ we choose should have support only on the five triples in NTW's support. We need to keep symmetry between $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$: otherwise f could "recognize" which string it is operating on. Thus $\mathcal{D}_\delta$ should give the same probability to each of the triples $(1,1,-1)$, $(1,-1,1)$, and $(-1,1,1)$. Thus we are led to look for a distribution of the form:

$\boldsymbol{x}_i$	$\boldsymbol{y}_i$	$\boldsymbol{z}_i$	probability
$+1$	$+1$	$+1$	p
$+1$	$+1$	-1	q
$+1$	-1	$+1$	q
-1	$+1$	$+1$	q
-1	-1	-1	r

We have the constraint $p + 3q + r = 1$. Next, there seems to be no way to reject constant functions except by using the trick of testing f^{odd}, and that trick only works assuming the uniform probability distribution. This leads to the additional constraint $q + r = 1/2$. Together the two constraints imply $q = 1/4 - p/2$, $r = 1/4 + p/2$. Thus we are forced into the distribution $\mathcal{D}_\delta$, except that it is not yet clear that $p = \delta$ should be "small" but nonzero. This smallness is forced by the fact that the prototypical quasirandom function Majority must be rejected with fairly high probability. Note that

$$\mathbf{E}'[\boldsymbol{x}_i\boldsymbol{y}_i] = \mathbf{E}'[\boldsymbol{y}_i\boldsymbol{z}_i] = \mathbf{E}'[\boldsymbol{x}_i\boldsymbol{z}_i] = \delta,$$

where $\mathbf{E}'$ denotes expectation with respect to $\mathcal{D}_\delta$. If $p = \delta$ were quite large then $\text{Maj}(\boldsymbol{x})$, $\text{Maj}(\boldsymbol{y})$, and $\text{Maj}(\boldsymbol{z})$ would be quite correlated, and since $\texttt{NTW}(1,1,1) = \texttt{NTW}(-1,-1,-1) = 1$, this would lead to a high acceptance probability. On the other hand, we also cannot have $p = 0$; in that case, we would be reduced to the BLR test based on XOR predicate, and one can check that the other prototypical quasirandom function parity, on an odd number of bits, would pass with probability 1. Luckily, the extreme noise sensitivity of parity means that even a small δ weight outside the support of XOR is enough to confound large parities.

Now given that we've settled on $\mathcal{D}_\delta$ and hence the test described in Section 5, why should this be a good Dictator-vs.-Quasirandom Test? More specifically, why should it have soundness as low as 5/8? Again, we check the prototypical quasirandom functions:

First, the test evades the constant functions $f \equiv 1$, $f \equiv -1$ by using the trick of testing f^{odd}: the two constant functions each pass with probability 5/8.

More interestingly, the test evades the prototypical odd quasirandom function Majority. To see this, first suppose that $\delta = 0$, so we are picking $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ as in the (negated) BLR linearity test. One can show (using the CLT) that from Majority's point of view, correlating three strings according to `XOR` is not much different from just picking three independent uniform inputs. I.e., $(\mathrm{Maj}(\boldsymbol{x}), \mathrm{Maj}(\boldsymbol{y}), \mathrm{Maj}(-\boldsymbol{x} \circ \boldsymbol{y}))$ has a distribution close to uniform on $\{-1,1\}^3$ and thus $\mathtt{NTW}(\mathrm{Maj}(\boldsymbol{x}), \mathrm{Maj}(\boldsymbol{y}), \mathrm{Maj}(-\boldsymbol{x} \circ \boldsymbol{y}))$ would be true with probability close to 5/8. Now for small $\delta > 0$, the "random restriction" first effectively converts Majority into a different symmetric threshold function; say, $f_{\boldsymbol{w}} = \mathrm{Maj}'$. Its bias will be only $O(\sqrt{\delta})$, and similar reasoning can show (proof omitted) that $\mathtt{NTW}(\mathrm{Maj}(\boldsymbol{x}), \mathrm{Maj}(\boldsymbol{y}), \mathrm{Maj}(-\boldsymbol{x} \circ \boldsymbol{y}))$ is true with probability only $5/8 + O(\delta)$.

Finally, the test evades the other prototypical quasirandom functions, large odd Parities. To see this, again consider the "random restriction" function $f_{\boldsymbol{w}}$. Sometimes $f_{\boldsymbol{w}}$ is again a parity function, and otherwise it is the negation of a parity function. In the former case it will pass the (negated) BLR test (and hence the `NTW` test) with probability 1; however, in the latter case $(f(\boldsymbol{x}), f(\boldsymbol{y}), f(\boldsymbol{z}))$ will have the uniform distribution on the support of $\neg$`XOR` and thus pass `NTW` only with probability 1/4 (when it is $(1,1,1)$). Further, when the original parity f is large enough compared to δ, the random restriction $f_{\boldsymbol{w}}$ will be a parity or a negated parity with roughly equal probability; hence the original parity f will pass with probability roughly $(1/2) \cdot 1 + (1/2) \cdot (1/4) = 5/8$.

E Proofs of quasirandomness and Gaussianicity

Proof. (Lemma A.1.) We prove the latter statement, leaving the former to the reader. When $\boldsymbol{f}$ is the odd parity χ_R we have

$$\sum_{\substack{S \subseteq [n] \\ i \in S}} (1-\delta)^{|S|-1} \hat{\boldsymbol{f}}(S)^2 \leq (1-\delta)^{|R|-1}$$

for each i. Now the random odd set R will have cardinality at least $n/3$ except with probability at most $2^{-\Omega(n)}$. Then

$$(1-\delta)^{|R|-1} \leq (1-\delta)^{n/3-1} \leq (1-\delta)^{\ln(e/\epsilon)/\delta - 1} \leq \epsilon,$$

as necessary. □

Proof. (Lemma A.2.) As mentioned, a variant of this lemma was proven in the full version of [23]; the proof here is mostly the same. By a comment in Section 4.1, it suffices to show that $\boldsymbol{f}$ is $(\epsilon/e, 0)$-Gaussianic except with probability at most $2^{-n^{\Omega(1)}}$. Standard probabilistic arguments show that all of the following simultaneously hold, except with probability at most $2^{-n^{\Omega(1)}}$ over the choice of the vector $\boldsymbol{G} = (\boldsymbol{G}_1, \dots, \boldsymbol{G}_n)$:

$$|\boldsymbol{G}_i| \leq n^{1/4} \quad \text{for all } i, \qquad \text{and,} \quad \tfrac{1}{2}n \leq \|\boldsymbol{G}\|^2 \leq \tfrac{3}{2}n. \tag{17}$$

Fix $\boldsymbol{G} = (G_1, \dots, G_n)$ satisfying these conditions; we will show the associated f is $(\epsilon/e, 0)$-

Gaussianic. We need to show that for each i,

$$\begin{aligned}
\epsilon/e &\geq \sum_{\substack{S\subseteq[n]\\ i\in S}} \hat{f}(S)^2 \\
&= \Pr_{\boldsymbol{x}}[f(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n) \neq f(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1},-\boldsymbol{x}_i,\boldsymbol{x}_{i+1},\ldots,\boldsymbol{x}_n)] \qquad \text{(a well-known identity)} \\
&= \Pr_{\boldsymbol{x}}\Big[\Big|\sum_{j\neq i} G_j \boldsymbol{x}_j\Big| \leq |G_i|\Big].
\end{aligned}$$

By (17),

$$\Pr_{\boldsymbol{x}}\Big[\Big|\sum_{j\neq i} G_j \boldsymbol{x}_j\Big| \leq |G_i|\Big] \leq \Pr_{\boldsymbol{x}}\Big[\Big|\sum_{j\neq i} G_j \boldsymbol{x}_j\Big|/\sigma \leq n^{1/4}/\sigma\Big],$$

where we have written $\sigma = \sqrt{\sum_{j\neq i} G_j^2} = \Theta(\sqrt{n})$ (using (17)). But the Berry-Esseen theorem implies that $(\sum_{j\neq i} G_j \boldsymbol{x}_j)/\sigma$ has distribution close to that of a standard Gaussian, up to error $\max|G_j|/\sigma \leq O(n^{-1/4})$ (using (17) once more). So the probability above is at most $O(n^{1/4}/\sigma) + O(n^{-1/4}) = O(n^{-1/4})$. This is at most ϵ/e by our assumption that $n \geq O(1/\epsilon^4)$. □