Testing Real-Valued Modularity and Submodularity

Pooya Hatami* Jan Vondrák†

Abstract

We study the question of testing whether a function $f : \{0,1\}^n \to \mathbb{R}$ is modular/submodular or ε-far from it (with respect to Hamming distance). We provide two results: First, it is possible to test using $O(\frac{n}{\varepsilon \log n})$ queries whether f is modular (equivalently affine, or linear with a constant term). For constant ε, this improves upon a simple tester that uses $O(n)$ queries. Second, we prove that testing submodularity requires $\Omega(\frac{n^2}{\log n})$ queries, thus separating the two problems. This improves on a linear lower bound due to Seshadhri and Vondrák.

*Institute for Advanced Study, Princeton, NJ, USA. email: pooyahat@math.ias.edu Research supported by the National Science Foundation grant No. CCF-1412958

†Dept. of Mathematics, Stanford University, Stanford, CA 94305. email: jvondrak@stanford.edu

1 Introduction

Property testing is concerned with the problem of deciding whether a given function $f : D \to R$ possesses a certain property or is "ε-far" from it, by a small number of queries (ideally depending only on ε) to an oracle providing $f(x)$ for a given $x \in D$. The notion of ε-far depends on the context. In this work we consider the Hamming distance model where a function f is ε-far from g if they differ on at least an ε-fraction of the domain D.

Property testing was originally motivated by applications in complexity theory, in seminal papers by Babai-Fortnow-Lund [BFL91], Blum-Luby-Rubinfeld [BLR93a], Rubinfeld-Sudan [RS96] and Bellare et al. [BCH+96]. These works were concerned with algebraic properties of functions, which ended up playing a crucial role in the development of probabilistically checkable proofs [AS98, ALM+98]. More recently, the area has blossomed with many works on testing for combinatorial, graph-theoretic and algebraic properties [AS08a, AS08b, BLR93b, AFKS00, AKK+05, KS08, BFH+13].

In this paper, we revisit the question of testing for basic algebraic properties of functions on the boolean cube $\{0,1\}^n$. Perhaps the most basic question is whether a given function on the boolean cube is linear or not. This question for functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a classical problem, which is resolved by the BLR test in $O(1/\varepsilon)$ queries [BLR93a]. A similar tester also allows testing whether a function is modular (affine) in $O(1/\varepsilon)$ queries. Linearity and homomorphism testing are studied in the more general setting of functions $f : G \to H$ for groups G and H, see for example [GLR+91, BLR93a, OY15]. Curiously, the same question for real-valued functions $f : \{0,1\}^n \to \mathbb{R}$ in the Hamming distance model has not received much attention, even though property testing for other properties of real-valued functions such as convexity and submodularity have been investigated [PRR03, SV14]. Convexity and submodularity can be viewed as degree-2 properties of real-valued functions, while modularity and linearity are degree-1 properties. The main question we ask in this paper is, what can one say about modularity testing of real-valued functions and how does its complexity compare to degree-2 properties?

1.1 Our results

More formally, we are interested in testing properties of real-valued functions $f : \{0,1\}^n \to \mathbb{R}$ in the Hamming distance model. The Hamming distance of $f : \{0,1\}^n \to \mathbb{R}$ from $g : \{0,1\}^n \to \mathbb{R}$ is defined to be $\frac{1}{2^n} |\{x : f(x) \neq g(x)\}|$, namely the fraction of the cube where f needs to be modified in order to obtain g. The Hamming distance of $f : \{0,1\}^n \to \mathbb{R}$ from a property $\mathcal{P}$ is the smallest distance of f from a function in $\mathcal{P}$.

A property $\mathcal{P} \subseteq \{f : \{0,1\}^n \to \mathbb{R} | n \in \mathbb{Z}\}$ is said to be $q(n,\varepsilon)$-query testable if there is a tester that queries a given function $f : \{0,1\}^n \to \mathbb{R}$ on $q \leqslant q(n,\varepsilon)$ (possibly adaptively) randomly selected points and either accepts or rejects f with the following guarantees.

- If $f \in \mathcal{P}$, it always accepts.
- If f is ε-far from $\mathcal{P}$, it rejects with probability at least 2/3.

Modularity. A function $f : \{0,1\}^n \to \mathbb{R}$ is said to be *modular* if for every $x, y \in \{0,1\}^n$, $f(x \vee y) + f(x \wedge y) = f(x) + f(y)$, where $\vee$ and $\wedge$ are the bitwise logical 'or' and 'and' respectively. It is easy to see that modularity is equivalent to *affineness*, where a function $f : \{0,1\}^n \to \mathbb{R}$ is said to be *affine* if there exist $\alpha_1, ..., \alpha_n \in \mathbb{R}$ and $c \in \mathbb{R}$ such that $f(x) = \sum_{i=1}^n \alpha_i x_i + c$ for any $x \in \{0,1\}^n$. A function is said to be *linear* if it is affine with $c = 0$.

There is a simple tester for modularity, and similarly for linearity, with $n + O(\frac{1}{\varepsilon})$ queries. Given query access to $f : \{0,1\}^n \to \mathbb{R}$, the tester queries $f(0)$ and $f(e_i)$ for all $i \in [n]$, and defines

$$g(x) := f(0) + \sum_{i=1}^{n} (f(e_i) - f(0)) x_i,$$

where $e_i \in \{0,1\}^n$ is the vector with only the i-th entry equal to 1. (In the case of linearity, $f(0)$ should be equal to 0 and we can reject immediately if that is not the case.) Next, we estimate the Hamming distance of f from g by repeating the following test $O(1/\varepsilon)$ times.

- Pick $x \in \{0,1\}^n$ uniformly at random and reject if $f(x) \neq g(x)$.

The tester accepts if f was not rejected in any of the above steps. It is easy to see that a modular function f will always be accepted, since in this case $f = g$. Otherwise, if f is ε-far from modularity, it in particular is ε-far from g which is modular and the above tester will reject with the desired probability. Note that the tester in fact does more, and learns the modular function.

Although this tester seems inefficient compared to linearity/affineness testers known for the case of functions over the group $\mathbb{F}_2$, breaking the n barrier in the real-valued functions with Hamming distance setting seems to be nontrivial. As our main result, we give a sublinear tester for modularity.

Theorem 1.1. *Modularity for functions $f : \{0,1\}^n \to \mathbb{R}$ is testable using $O(\frac{n}{\varepsilon \log n})$ queries.*

The main tool in the analysis of our tester is a strengthening of the KKL inequality due to Talagrand [Tal94].

It is an intriguing question whether $\tilde{\Theta}(n)$ queries are necessary for testing modularity of real-valued functions. Also, our tester does not seem to extend easily to linearity testing.

Submodularity. Our second result is on testing submodularity over $\{0,1\}^n$. A function $f : \{0,1\}^n \to \mathbb{R}$ is said to be *submodular* if for every $x, y \in \{0,1\}^n$, $f(x \vee y) + f(x \wedge y) \leqslant f(x) + f(y)$. Equivalently, f is submodular if for all $x \preceq y$ and $i \in [n]$ such that $y_i = 0$, $f(x + e_i) - f(x) \leqslant f(y + e_i) - f(y)$, where $x \preceq y$ if $x_j \leqslant y_j$ for all $j \in [n]$. Seshadhri and Vondrak [SV14] proved an $\Omega(n)$ lower-bound for testing submodularity, via a reduction from monotonicity testing. Here we give a near-quadratic lower bound for testing submodularity, showing that testing submodularity is in fact harder than (real-valued) modularity.

Theorem 1.2. *Any tester for submodularity for constant distance $\varepsilon \leqslant 1/4$ requires $\Omega(\frac{n^2}{\log n})$ queries.*

We prove Theorem 1.2 via a reduction from communication complexity of set-disjointness, a technique introduced in [BBM12].

It is tempting to conjecture that $\tilde{\Theta}(n^2)$ is the right answer for testing submodularity (due to its being a degree-2 property). But a quadratic tester for submodularity is not known and in fact no polynomial or quasipolynomial tester is known ([SV14] gives a submodularity tester using $1/\varepsilon^{O(\sqrt{n}\log n)}$ queries). This wide gap remains open.

2 Preliminaries

KKL Edge-Isoperimetric Theorem. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. The *(ℓ_2^2) influence* of the i-th variable on f is defined to be

$$\mathrm{Inf}_i[f] = \mathop{\mathbf{E}}_{x\in\{0,1\}^n}\left[(f(x\oplus e_i) - f(x))^2\right],$$

where x is chosen uniformly at random and $\oplus$ is the Boolean exclusive or. The *total (ℓ_2^2) influence* of f is defined as

$$\mathbf{I}[f] := \sum_{i=1}^{n} \mathrm{Inf}_i[f] = n \cdot \mathop{\mathbf{Pr}}_{x\in\{0,1\}^n, i\in[n]}[f(x) \neq f(x\oplus e_i)].$$

We will need the following edge-isoperimetric inequality due to Talagrand [Tal94] which is a strengthening of the well-known KKL theorem of Kahn, Kalai, and Linial [KKL88].

Theorem 2.1 (KKL Edge-Isoperimetric Theorem)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be nonconstant and let $\mathbf{I}'[f] := \mathbf{I}[f]/\mathbf{Var}[f] \geqslant 1$. Then there exists $i \in [n]$, such that*

$$\mathrm{Inf}_i[f] \geqslant \left(\frac{9}{\mathbf{I}'[f]^2}\right) \cdot 9^{-\mathbf{I}'[f]}.$$

Communication complexity. The two-party communication model as introduced by Andrew Yao [Yao79] is the following problem involving two parties called Alice and Bob. Suppose $f : \{0,1\}^n \times \{0,1\}^n \to S$ is a function where S is some set. Alice receives a vector $x \in \{0,1\}^n$ and Bob receives a vector $y \in \{0,1\}^n$, and they wish to compute $f(x,y)$ with the least amount of communication between them. A joint strategy of Alice and Bob is called a communication protocol. In a randomized protocol Alice and Bob have access to a shared random string R which they can use to select what messages to exchange. The communication cost of a protocol is the worst-case maximum number of bits sent, taken over all possible inputs and values of R. The *bounded-error communication complexity* of f, denoted by $R(f)$ is the minimum cost of a randomized protocol that on all inputs (x,y) computes f with success probability $2/3$.

We will work with the well-known *Disjointness* function $\mathrm{DISJ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, defined as

$$\mathrm{DISJ}_n(x,y) = \vee_{i=1}^n (x_i \wedge y_i),$$

namely, if we think of $x, y \in \{0,1\}^n$ as subsets of $[n]$, $\mathrm{DISJ}_n(x,y) = 1$ if $x \cap y \neq \emptyset$. The disjointness function is proved to have linear randomized communication complexity.

Theorem 2.2 ([KS92, Raz92, BYJKS02])**.** *$R(\mathrm{DISJ}_n) = \Omega(n)$.*

We will prove Theorem 1.2 via a reduction from the communication problem of disjointness. This technique was introduced by Blais et al. [BBM12], where they used reductions from communication complexity to prove tight lower-bounds for several property testing problems.

3 A sublinear tester for modularity

Here we prove Theorem 1.1. We will use a combination of three different tests on a given function $f : \{0,1\}^n \to \mathbb{R}$.

Test 1. Pick $i \in [n]$ uniformly at random, and $x = (x_1, \ldots, x_n) \in \{0,1\}^n$ with $x_i = 0$ uniformly at random.

- Reject if $f(x + e_i) - f(x) \neq f(e_i) - f(0)$.

Test 2. Pick $y = (y_1, ..., y_n) \in \{0,1\}^n$ at random where for each i independently $y_i = 1$ with probability $\frac{1}{\sqrt{n}}$. Pick uniformly at random $m = \min\{\sum_{i=1}^n y_i, \sqrt{n}\}$ coordinates $i_1, ..., i_m \in [n]$ such that $y_{i_j} = 1$ for all $j \in [m]$.

- Reject if there exists $j \in [m]$ such that $f(y) - f(y - e_{i_j}) \neq f(e_{i_j}) - f(0)$.

Test 3. Pick $x, y \in \{0,1\}^n$ according to the following distribution: For each $i \in [n]$ independently sample

$$(x_i, y_i) = \begin{cases} \text{(0,0) with probability } \frac{1}{2} \\ \text{(0,1) with probability } \frac{1}{\sqrt{n}} \\ \text{(1,0) with probability } \frac{1}{2} - \frac{1}{\sqrt{n}} \end{cases}$$

- Reject if $f(x + y) \neq f(x) + f(y) - f(0)$.

Modularity/affineness Tester.

1. Run **Test 1.** $O(\frac{n}{\varepsilon \log n})$ times
2. Run **Test 2.** $O(n^{1/4})$ times
3. Run **Test 3.** $O(n^{3/4})$ times,
4. Accept if none of the above have rejected.

It is clear that the above tester always accepts a modular function, as neither of the three tests reject a modular function. It is left to show that a function that is ε-far from modularity is rejected with high probability.

Intuition. A modular function is characterized by the fact that its discrete derivatives $f(x + e_i) - f(x)$ are constant. We need to identify a deviation from this property if it occurs, without learning the n partial derivatives (since that would take $\Omega(n)$ queries). Test 1 guarantees that if it passes, then most of the partial derivatives are mostly constant. Still, it is not clear if this guarantees that the function must be close to modular.

Hence, we argue as follows. If the function is modular, then there is a linear function, $g(x) = \sum_i (f(e_i) - f(0))x_i$, such that $f - g$ is constant. We can try to test whether $f - g$ is constant without ever learning a full description of g. The key observation is that if $f - g$ is far from a constant, then it has different level sets of significant size and by an isoperimetric argument, there are many edges of the hypercube between different level sets of $f - g$ (either many edges overall, or significantly many in a particular coordinates; a quantitative expression of this is the KKL Edge-Isoperimetric Theorem we mentioned earlier). Let's call such edges problematic. If there are many problematic edges overall, it will be detected by Test 1. However, it could be that only 1 (or a small number) of coordinates have problematic edges, and again we cannot afford to test each coordinate separately.

It turns out we can handle this case by considering random pairs of correlated points differing in $O(\sqrt{n})$ coordinates, effectively trying to find a problematic coordinate in batches of $O(\sqrt{n})$. It is important that such pairs of points can be chosen so that the distribution of each is still roughly uniform. We need a careful combination of two tests, Test 2 and Test 3, to handle this case: Test 2 checks that discrete derivatives look individually constant for most random batches of $O(\sqrt{n})$ coordinates, and assuming that this is the case, Test 3 finally discovers a problematic edge there is a significant number of them in some coordinate.

Analysis.

Lemma 3.1. *Let $f : \{0,1\}^n \to \mathbb{R}$ be ε-far from modularity. The modularity tester above rejects with probability $\Omega(1)$.*

Proof. Define $g : \{0,1\}^n \to \mathbb{R}$ as $g(x) := \sum_{i=1}^n (f(e_i) - f(0))x_i$, and $h(x) := f(x) - g(x)$.

Claim 3.2. *h is ε-far from being constant.*

Proof. Assume to the contrary that $h = f - g$ is ε-close to a constant $c \in \mathbb{R}$. This means that f is ε-close to $g + c$ a modular function, a contradiction. □

Choose a Boolean function $h^B : \{0,1\}^n \to \{0,1\}$ such that

(i) $h^B(x) \neq h^B(y)$ only if $h(x) \neq h(y)$

(ii) $\mathbf{Var}(h^B) = \Omega(\varepsilon)$.

For example, we pick λ such that $\min\{\mathbf{Pr}[h(x) < \lambda], \mathbf{Pr}[h(x) \geqslant \lambda]\} \geqslant \varepsilon$, which is possible since h is ε-far from constant. Then we define $h^B(x) = 0$ if $h(x) < \lambda$ and $h^B(x) = 1$ if $h(x) \geqslant \lambda$. We have $\min\{\mathbf{Pr}[h_B(x) = 0], \mathbf{Pr}[h^B(x) = 1]\} \geqslant \varepsilon$, hence $\mathbf{Var}(h^B) = \Omega(\varepsilon)$.

We will break down the remaining analysis into two cases, based on the maximum influence of a variable in h^B.

Case (i), $\forall i \in [n]$, $\mathbf{Inf}_i[h^B] \leqslant n^{-1/4}$. In this case by Theorem 2.1,

$$\left(\frac{9}{\mathbf{I}'[h^B]^2}\right) \cdot 9^{-\mathbf{I}'[h^B]} \leqslant n^{-1/4},$$

therefore $\mathbf{I}'[h^B] = \Omega(\log n)$ and $\mathbf{I}[h^B] = \mathbf{Var}[h^B]\,\mathbf{I}'[h^B] = \Omega(\varepsilon \log n)$. In particular,

$$\Pr_{i\in[n],x\in\{0,1\}^n:x_i=0}[h^B(x) \neq h^B(x+e_i)] = \Omega\left(\frac{\varepsilon \log n}{n}\right).$$

By construction, $h^B(x) \neq h^B(x+e_i)$ implies $h(x) \neq h(x+e_i)$ and thus

$$f(x+e_i) - f(x) \neq g(x+e_i) - g(x) = f(e_i) - f(0).$$

Note that this is exactly what **Test 1.** checks and thus $O(\frac{n}{\varepsilon \log n})$ independent runs of **Test 1.** will reject f with positive probability.

Case (ii), $\exists i^* \in [n]$, $\mathrm{Inf}_{i^*}[h^B] \geqslant n^{-1/4}$. Let $D_{x,y}$ be the distribution over $\{0,1\}^n \times \{0,1\}^n$ as described in **Test 3.**, and note that the marginal distribution on y, D_y, is used by **Test 2.**. Assume for contradiction that f is rejected with only probability $o(1)$ in step 2 and step 3 of our modularity tester. We will obtain a contradiction by showing that in this case $O(n^{3/4})$ independent runs of **Test 3.** reject f with positive probability. Note that,

$$\begin{aligned}\Pr_{(x,y)\sim D_{xy}}[f(x+y) \neq f(x)+f(y)-f(0)] &\geqslant \frac{1}{\sqrt{n}} \Pr_{(x,y)\sim D_{xy}}[f(x+y) \neq f(x)+f(y)-f(0)|y_{i^*}=1] \\ &= \frac{1}{\sqrt{n}} \Pr_{(x,y)\sim D_{xy}}[f(x+y) - f(x+y-e_{i^*}) \neq f(x)+f(y)-f(0)-f(x+y-e_{i^*})|y_{i^*}=1]. \quad (1)\end{aligned}$$

Under D_{xy}, $x+y$ is uniformly distributed over $\{0,1\}^n$, and thus

$$\Pr_{x,y\sim D_{xy}}[f(x+y) - f(x+y-e_{i^*}) \neq f(e_{i^*}) - f(0)|y_{i^*}=0] \geqslant \mathrm{Inf}_{i^*}[h^B] \geqslant n^{-1/4}. \quad (2)$$

Moreover, since f passed step 2 with $1-o(1)$ probability,

$$\Pr_{y\sim D_y}[f(y) - f(y-e_{i^*}) \neq f(e_{i^*}) - f(0)|y_{i^*}=0] = o(n^{-1/4}); \quad (3)$$

here we have used the fact that under D_y, $\sum_{i=1}^n y_i$ is concentrated around $\sqrt{n}$, and thus for a typical y, $f(y) - f(y-e_{i^*}) \neq f(e_{i^*}) - f(0)$ would be detected if it holds. Under the assumption that step 3 does reject with only $o(1)$ probability,

$$\Pr_{x,y\sim D_{xy}}[f(x) + f(y-e_{i^*}) \neq f(x+(y-e_{i^*})) + f(0)|y_{i^*}=0] = o(n^{-3/4}). \quad (4)$$

Thus combining Eqs. (2) to (4), by a union bound on the RHS of Eq. (1), we have

$$\Pr_{(x,y)\sim D_{xy}}[f(x+y) \neq f(x)+f(y)-f(0)|y_{i^*}=1] \geqslant n^{-1/4} - o(n^{-3/4}) - o(n^{-1/4}) = \Omega(n^{-1/4}),$$

thus proving

$$\Pr_{(x,y)\sim D_{xy}}[f(x+y) \neq f(x)+f(y)-f(0)] \geqslant \frac{1}{\sqrt{n}} \cdot \Omega(\frac{1}{n^{1/4}}) = \Omega(n^{-3/4}).$$

Thus $O(n^{3/4})$ runs of **Test 3.** will reject f, a contradiction. $\square$

4 Lower bound for submodularity testing

Here we prove Theorem 1.2 via a reduction from the Disjointness problem. Suppose there is a $q(n,\varepsilon)$ query tester for testing submodularity. We show that this implies $R(\mathrm{DISJ}_{n^2}) \leqslant q(n,\varepsilon)\cdot \log n$. This combined with Theorem 2.2 shows that $q(n,\varepsilon)=\Omega(n^2/\log n)$.

Let $N := n^2$. Alice and Bob are given $x=(x_{i,j})_{1\leqslant i,j\leqslant n}, y=(y_{i,j})_{1\leqslant i,j\leqslant n}\in\{0,1\}^N$ respectively, which we think of as n by n matrices. Define $h_{xy}:\{0,1\}^n\to\mathbb{R}$ as

$$h_{xy}(z) := 2|z|(n-|z|) + \sum_{i,j:\, x_{i,j}=1}\chi_{\{i,j\}}(z) + \sum_{i,j:\, y_{i,j}=1}\chi_{\{i,j\}}(z),$$

where $|z|=\sum_{i=1}^n z_i$ and $\chi_S(z)=(-1)^{\sum_{i\in S} z_i}$ for $S\subseteq[n]$. Note that the first term contributes -4 to each expression in the form $h_{xy}(z)+h_{xy}(z+e_i+e_j)-h_{xy}(z+e_i)-h_{xy}(z+e_j)$, which should be nonpositive for a submodular function. The remaining terms depend on x and y, and we analyze them in the following.

Claim 4.1. *Let $x,y\in\{0,1\}^N$ be such that $\mathrm{DISJ}_N(x,y)=1$, then h_{xy} is $\frac{1}{4}$-far from submodularity.*

Proof. Let $i,j\in[n]$ be such that $x_{i,j}=y_{i,j}=1$. Let $z\in\{0,1\}^n$ be any vector such that $z_i=z_j=0$. Note that,

$$\begin{aligned}
&h_{xy}(z)+h_{xy}(z+e_i+e_j)-h_{xy}(z+e_i)-h_{xy}(z+e_j)\\
&\qquad = -4+2\chi_{\{i,j\}}(z)+2\chi_{\{i,j\}}(z+e_i+e_j)-2\chi_{\{i,j\}}(z+e_i)-2\chi_{\{i,j\}}(z+e_j) = 8\chi_{\{i,j\}}(z)-4>0,
\end{aligned}$$

violating submodularity. Thus, in order to obtain a submodular function, the value of h_{xy} needs to be changed on at least one of the points z, $z+e_i$, $z+e_j$, and $z+e_i+e_j$, for each z such that $z_i=z_j=0$. □

Claim 4.2. *Let $x,y\in\{0,1\}^N$ be such that $\mathrm{DISJ}_N(x,y)=0$, then h_{xy} is submodular.*

Proof. Suppose $x,y\in\{0,1\}^N$ are such that $\mathrm{DISJ}_N(x,y)=0$. Therefore at most one of the bits $x_{i,j},y_{i,j}$ is nonzero for any given $i,j\in[n]$. Suppose $z\in\{0,1\}^n$ with $z_i=z_j=0$; then

$$\begin{aligned}
&h_{xy}(z)+h_{xy}(z+e_i+e_j)-h(z+e_i)-h(z+e_j)\\
&\qquad \leqslant -4+\chi_{\{i,j\}}(z)+\chi_{\{i,j\}}(z+e_i+e_j)-\chi_{\{i,j\}}(z+e_i)-\chi_{\{i,j\}}(z+e_j)\leqslant 0.
\end{aligned}$$

Hence h_{xy} is submodular. □

The protocol. Given their inputs $x,y\in\{0,1\}^N$, Alice and Bob use a submodularity tester as follows. Bob will respond to queries of the submodularity tester to h_{xy}. Alice and Bob can compute a query $h_{xy}(z)$ within $O(\log n)$ bits of communication. Thus, after $O(q(n,\varepsilon)\cdot\log n)$ bits of communication, the tester outputs whether h_{xy} is ε-far from submodularity or is submodular. At this point by Claims 4.1 and 4.2, Bob knows the correct answer to $\mathrm{DISJ}_N(x,y)$ with probability 2/3. Thus

$$R(\mathrm{DISJ}_N)\leqslant O(\log n\cdot q(n,\varepsilon)),$$

for $\varepsilon\leqslant 1/4$, which combined with Theorem 2.2 implies $q(n,\varepsilon)=\Omega(n^2/\log n)$ finishing the proof of Theorem 1.2.

References

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy, *Efficient testing of large graphs*, Combinatorica **20** (2000), no. 4, 451–476. MR 1804820 (2002f:05143) 2

[AKK+05] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron, *Testing Reed-Muller codes*, IEEE Trans. Inform. Theory **51** (2005), no. 11, 4032–4039. 2

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy, *Proof verification and the hardness of approximation problems*, J. ACM **45** (1998), no. 3, 501–555. 2

[AS98] Sanjeev Arora and Shmuel Safra, *Probabilistic checking of proofs: A new characterization of NP*, J. ACM **45** (1998), no. 1, 70–122. 2

[AS08a] Noga Alon and Asaf Shapira, *A characterization of the (natural) graph properties testable with one-sided error*, SIAM J. on Comput. **37** (2008), no. 6, 1703–1727. 2

[AS08b] ———, *Every monotone graph property is testable*, SIAM J. on Comput. **38** (2008), no. 2, 505–522. 2

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef, *Property testing lower bounds via communication complexity*, Computational Complexity **21** (2012), no. 2, 311–358. 3, 5

[BCH+96] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos Kiwi, and Madhu Sudan, *Linearity testing over characteristic two*, IEEE Trans. Inform. Theory **42** (1996), no. 6, 1781–1795. 2

[BFH+13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett, *Every locally characterized affine-invariant property is testable*, Proceedings of the 45th annual ACM symposium on Symposium on theory of computing (New York, NY, USA), STOC '13, ACM, 2013, pp. 429–436. 2

[BFL91] László Babai, Lance Fortnow, and Carsten Lund, *Nondeterministic exponential time has two-prover interactive protocols*, Comput. Complexity **1** (1991), no. 1, 3–40. MR 1113533 (92h:68031) 2

[BLR93a] Manuel Blum, Michael Luby, and Ronitt Rubinfeld, *Self-testing/correcting with applications to numerical problems*, J. Comput. Syst. Sci. **47** (1993), no. 3, 549–595. 2

[BLR93b] Manuel Blum, Michael Luby, and Ronitt Rubinfeld, *Self-testing/correcting with applications to numerical problems*, Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (Baltimore, MD, 1990), vol. 47 (3), 1993, pp. 549–595. MR 1248868 (94m:65020) 2

[BYJKS02] Ziv Bar-Yossef, TS Jayram, Ravindra Kumar, and D Sivakumar, *An information statistics approach to data stream and communication complexity*, Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on, IEEE, 2002, pp. 209–218. 4

[GLR+91] Peter Gemmell, Richard Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson, *Self-testing/correcting for polynomials and for approximate functions*, STOC, vol. 91, Citeseer, 1991, pp. 32–42. 2

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial, *The influence of variables on boolean functions*, Foundations of Computer Science, 1988., 29th Annual Symposium on, IEEE, 1988, pp. 68–80. 4

[KS92] Bala Kalyanasundaram and Georg Schintger, *The probabilistic communication complexity of set intersection*, SIAM Journal on Discrete Mathematics **5** (1992), no. 4, 545–557. 4

[KS08] Tali Kaufman and Madhu Sudan, *Algebraic property testing: the role of invariance*, Proc. 40th Annual ACM Symposium on the Theory of Computing, 2008, pp. 403–412. 2

[OY15] Kenta Oono and Yuichi Yoshida, *Testing properties of functions on finite groups*, arXiv preprint arXiv:1509.00930 (2015). 2

[PRR03] Michal Parnas, Dana Ron, and Ronitt Rubinfeld, *On testing convexity and submodularity*, SIAM J. Comput. **32** (2003), no. 5, 1158–1184. 2

[Raz92] Alexander A. Razborov, *On the distributional complexity of disjointness*, Theoretical Computer Science **106** (1992), no. 2, 385–390. 4

[RS96] Ronitt Rubinfeld and Madhu Sudan, *Robust characterizations of polynomials with applications to program testing*, SIAM J. Comput. **25** (1996), no. 2, 252–271. MR 1379300 (97h:68050) 2

[SV14] Comandur Seshadhri and Jan Vondrák, *Is submodularity testable?*, Algorithmica **69** (2014), no. 1, 1–25. 2, 3, 4

[Tal94] Michel Talagrand, *On Russo's approximate zero-one law*, Ann. Probab. **22** (1994), no. 3, 1576–1587. MR 1303654 3, 4

[Yao79] Andrew Chi-Chih Yao, *Some complexity questions related to distributive computing (preliminary report)*, Proceedings of the eleventh annual ACM symposium on Theory of computing, ACM, 1979, pp. 209–213. 4