New algorithms and lower bounds for monotonicity testing

Xi Chen
Department of Computer Science
Columbia University
New York, NY
xichen@cs.columbia.edu

Rocco A. Servedio
Department of Computer Science
Columbia University
New York, NY
rocco@cs.columbia.edu

Li-Yang Tan
Department of Computer Science
Columbia University
New York, NY
liyang@cs.columbia.edu

***Abstract*—We consider the problem of testing whether an unknown Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is monotone versus ε-far from every monotone function. The two main results of this paper are a new lower bound and a new algorithm for this well-studied problem.**

Lower bound: We prove an $\tilde{\Omega}(n^{1/5})$ lower bound on the query complexity of any non-adaptive two-sided error algorithm for testing whether an unknown Boolean function f is monotone versus constant-far from monotone. This gives an exponential improvement on the previous lower bound of $\Omega(\log n)$ due to Fischer *et al.* [1]. We show that the same lower bound holds for monotonicity testing of Boolean-valued functions over hypergrid domains $\{1,\ldots,m\}^n$ for all $m \geq 2$.

Upper bound: We present an $\tilde{O}(n^{5/6})\text{poly}(1/\varepsilon)$-query algorithm that tests whether an unknown Boolean function f is monotone versus ε-far from monotone. Our algorithm, which is non-adaptive and makes one-sided error, is a modified version of the algorithm of Chakrabarty and Seshadhri [2], which makes $\tilde{O}(n^{7/8})\text{poly}(1/\varepsilon)$ queries.

***Keywords*-Boolean functions; Property testing; Monotonicity testing.**

I. Introduction

Monotonicity is a basic and natural property of functions. In the field of property testing, the problem of efficiently testing whether an unknown function f is monotone has been the focus of a long and fruitful line of research, with many works (see e.g. [1]–[17]) studying this problem for functions with various domains and ranges.

In this work we will be concerned with the classical problem of testing monotonicity of *Boolean functions* $f : \{-1,1\}^n \to \{-1,1\}$, which was first posed and considered explicitly by Goldreich *et al.* [3]. Recall a Boolean function f is monotone if $f(x) \leq f(y)$ for all $x \prec y$, where $\prec$ denotes the bitwise partial order on the hypercube. Let

$$\text{dist}(f,g) := \mathbf{Pr}_{\boldsymbol{x}\in\{-1,1\}^n}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})];$$

we say that f is *ε-close to monotone* if $\text{dist}(f,g) \leq \varepsilon$ for some monotone Boolean function g, and that f is *ε-far from monotone* otherwise. We will be interested in query-efficient randomized testing algorithms for the following task: *Given as input a distance parameter $\varepsilon > 0$ and oracle access to an unknown Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, output* Yes *with probability at least $2/3$ if f is monotone, and* No *with probability at least $2/3$ if f is ε-far from monotone.*

The work of Goldreich *et al.* [3] proposed a simple "edge tester" which queries uniform random edges of $\{-1,1\}^n$ hoping to find an edge whose two endpoints violate monotonicity. [3] proved an $O(n^2\log(1/\varepsilon)/\varepsilon)$ upper bound on the query complexity of the edge tester, which was subsequently improved to $O(n/\varepsilon)$ in the journal version [5]. Fischer *et al.* [1] established the first lower bounds shortly after, showing that there exists a constant distance parameter $\varepsilon_0 > 0$ such that $\Omega(\log n)$ queries are necessary for any *non-adaptive* tester (one whose queries do not depend on the oracle's responses to prior queries). This directly implies an $\Omega(\log\log n)$ lower bound for adaptive testers, since any q-query adaptive tester can be simulated by a non-adaptive one that simply carries out all 2^q possible executions. These upper and lower bounds were the best known for more than a decade, until the recent work of Chakrabarty and Seshadhri [2] improved on the linear upper bound of Goldreich *et al.* with an $\tilde{O}(n^{7/8}\varepsilon^{-3/2})$-query tester.

Our main contributions in this work are (i) a new lower bound that improves on the lower bound of [1] by an exponential factor, and (ii) a new algorithm that improves on the upper bound of [2] (in terms of the dependence on n) by a polynomial factor. We now describe these contributions in more detail.

Our lower bound. We give an exponential improvement on the lower bounds of Fischer *et al.* [1]:

Theorem 1. *There exists a universal constant $\varepsilon_0 > 0$ such that any non-adaptive algorithm for testing whether an unknown Boolean function is monotone versus ε_0-far from monotone must make $\Omega(n^{1/5}(\log n)^{-2/5})$ queries. Consequently, any adaptive algorithm must make $\Omega(\log n)$ queries.*

0272-5428/14 $31.00 © 2014 IEEE
DOI 10.1109/FOCS.2014.38

While the aforementioned results of Fischer *et al.* [1] represent the previous best lower bounds on the general testing problem as defined above, additional lower bounds are known for several restricted versions of the problem. In the same paper [1], Fischer *et al.* gave an $\Omega(\sqrt{n})$ lower bound on the query complexity of any non-adaptive *one-sided* tester, i.e. one that always outputs Yes when f is monotone (again, this directly implies an $\Omega(\log n)$ lower bound for adaptive one-sided testers). Restricting further, a *pair tester* is a non-adaptive one-sided tester that independently draws pairs of comparable points $x \prec y$ from some distribution and rejects if and only if some pair that is drawn violates monotonicity. Briët *et al.* [13] proved an $\Omega(n/(\varepsilon \log n))$ lower bound on the query complexity of pair testers whose query complexity can be written as $q(n)/\varepsilon$ for some function q.

In addition to Theorem 1, we show that essentially the same lower bound holds for monotonicity testing of Boolean-valued functions over hypergrid domains $\{1, \ldots, m\}^n$ for $m \geq 2$. (Below and throughout this paper we write $[m]$ to denote $\{1, 2, \ldots, m\}$.) Our most general lower bound is the following:

Theorem 2. *There exists a universal constant $\varepsilon_0 > 0$ such that for all $m \geq 2$, any non-adaptive algorithm for testing whether an unknown function $f : [m]^n \to \{-1, 1\}$ is monotone versus ε_0-far from monotone must make $\tilde{\Omega}(n^{1/5})$ queries.*

To the best of our knowledge, Theorem 2 is the first lower bound for testing monotonicity of *Boolean valued* functions over hypergrid domains. Recent papers of Chakrabarty and Seshadhri [15], [16] and Blais *et al.* [17] essentially closed the problem of testing monotonicity of functions $f : [m]^n \to \mathbb{N}$, showing that $\Theta(n \log m)$ queries are both necessary and sufficient; however, their lower bounds crucially depend on the functions considered having range $\mathbb{N}$ rather than $\{-1, 1\}$.

Our algorithm. We present a new algorithm for monotonicity testing, and prove the following result about its performance:

Theorem 3. *There is a $\tilde{O}(n^{5/6}\varepsilon^{-4})$-query one-sided non-adaptive algorithm for testing whether an unknown n-variable Boolean function is monotone versus ε-far from monotone.*

Recall that the one-sided, non-adaptive tester of Chakrabarty and Seshadhri [2] makes $\tilde{O}(n^{7/8}\varepsilon^{-3/2})$ queries. Thus, while the query complexity of our tester is worse as a function of $1/\varepsilon$ (though still polynomial), its query complexity is polynomially better as a function of n.[1] Like the [2] algorithm, our algorithm is a pair tester, but it evades the $\Omega(n/(\varepsilon \log n))$ lower bound of [13] because its query complexity is not of the form $q(n)/\varepsilon$. Our algorithm builds on the tools developed in [2]; its high-level structure is similar to that of the [2] algorithm, but with an important difference that enables an improved analysis. See Section I-B for more discussion on this point.

A. The lower bound approach

Our lower bound for testing monotonicity builds on previous lower bounds for testing restricted classes of *linear threshold functions* (LTFs). Recall that $f : \{-1, 1\}^n \to \{-1, 1\}$ is a linear threshold function if there exist $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ such that $f(x) = \text{sign}(w \cdot x - \theta)$ for all $x \in \{-1, 1\}^n$.

Background. A *signed majority function* is a linear threshold function of the special form $f(x) = \text{sign}(w \cdot x)$ where $w \in \{-1, 1\}^n$. While [18] showed that the class of all LTFs is ε-testable using $\text{poly}(1/\varepsilon)$ queries (independent of n), in [19] Matulef *et al.* gave an $\Omega(\log n)$ lower bound for non-adaptive algorithms that ε_0-test whether $f : \{-1, 1\}^n \to \{-1, 1\}$ is a signed majority function, where $\varepsilon_0 > 0$ is a universal constant. Like many lower bound arguments in property testing, the proof of [19] employs Yao's minimax principle [20], and works by exhibiting two distributions $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$ over LTFs — more precisely, $\mathcal{D}_{yes}$ is the uniform distribution over all 2^n signed majority functions, and $\mathcal{D}_{no}$ is the uniform distribution over a set of LTFs almost all of which are constant-far from every signed majority function — and arguing that for $q = o(\log n)$, any deterministic q-query algorithm cannot distinguish between the two distributions with non-negligible success probability. (We note that a typical function from $\mathcal{D}_{yes}$ is far from being monotone, and that the same holds for a typical LTF drawn from the $\mathcal{D}_{no}$ distribution of [19].) A key tool in the [19] proof is the Berry–Esséen "central limit theorem (CLT) with error bounds" for sums of independent real-valued random variables.

An *embedded majority function of size k* is an LTF $f : \{-1, 1\}^n \to \{-1, 1\}$ of the form $f(x) = \text{sign}(w \cdot x)$ where $w \in \{0, 1\}^n$ is a vector with exactly k ones. In [21] Blais and O'Donnell showed that for $k = n/2$, any non-adaptive testing algorithm for the class of all

[1] Recall that in property testing the dependence on the size parameter "n" is typically viewed as more important than the dependence on the "closeness" parameter ε. Indeed, ε is often viewed as a constant, so testers with query complexities that are exponential (or worse) as a function of $1/\varepsilon$ but independent of n are commonly referred to as "constant-query testers."

embedded majority functions of size exactly $n/2$ must make $\Omega(n^{1/12})$ queries. Their proof employed a $\mathcal{D}_{yes}$ distribution which is the uniform distribution over all embedded majority functions of size $n/2$, and a $\mathcal{D}_{no}$ distribution which is supported on certain monotone LTFs (which are far from embedded majority functions of size $n/2$). A key technical ingredient in the proofs of [21] is a multidimensional extension of the Berry–Esséen theorem (to independent sums of $\mathbb{R}^q$-valued random variables) which was essentially established in the work of [22], building on ingredients from [23]. Subsequently Ron and Servedio [24] adapted the arguments of [21] to give an improved analysis of the same $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$ distributions from [19] and establish an $\Omega(n^{1/12})$-query lower bound for non-adaptive algorithms that ε_0-test whether $f : \{-1,1\}^n \to \{-1,1\}$ is a signed majority function, thus exponentially improving over the [19] lower bounds for this problem.

This work. Neither the [21] construction nor the [19], [24] construction can be used directly to establish a lower bound for monotonicity testing of functions $f : \{-1,1\}^n \to \{-1,1\}$; as described above, in the [21] construction both the $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$ functions are monotone, and in the [19], [24] construction a typical function from either distribution is far from monotone. Nevertheless, in this work we show that ingredients from [21], [24] can be leveraged to obtain a polynomial lower bound for testing monotonicity of functions $f : \{-1,1\}^n \to \{-1,1\}$. Like these earlier works we employ Yao's principle: we define a $\mathcal{D}_{yes}$ distribution that is supported on monotone LTFs, and a $\mathcal{D}_{no}$ distribution over LTFs that is almost entirely supported on LTFs that are constant-far from every monotone function, and use an analysis which is fairly similar to that of [21], [24], to prove Theorem 1. Using the multidimensional Berry–Esséen theorem of [22] to analyze our $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$ distributions would result in an $\Omega(n^{1/12})$ lower bound. To obtain our improved $\Omega(n^{1/5}\log^{-2/5} n)$ lower bound, we instead adapt a multidimensional CLT of Valiant and Valiant [25] (for Wasserstein distance) to our context.

B. The approach of our algorithm

Our algorithm builds on ingredients from [2], so to explain our approach we first recall the necessary ingredients from that work. Fix a Boolean function[2] $f : \{0,1\}^n \to \{0,1\}$, and let us say that a pair of inputs (x, y) with $x \prec y$ is a *violated edge* if $f(x) = 1, f(y) = 0$ and (x, y) is an edge in $\{0,1\}^n$ (i.e. the Hamming distance between them is 1). [2] establishes a very useful "dichotomy theorem" about Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that are ε-far from monotone: for any $s > 0$, any such function either must have $\Omega(\varepsilon s 2^n)$ violated edges, or must have a *matching* (i.e. a vertex-disjoint set) of $\Omega(\varepsilon 2^n/s)$ violated edges.

To use this dichotomy theorem, Chakrabarty and Seshadhri [2] define a "path tester" which works essentially as follows: it selects a random directed path $\boldsymbol{p}$ of n edges from 0^n up to 1^n, draws two uniform random points $\boldsymbol{x} \prec \boldsymbol{y}$ from the "middle layers" of $\boldsymbol{p}$, and rejects if $\boldsymbol{x}$ and $\boldsymbol{y}$ violate monotonicity, i.e. $f(\boldsymbol{x}) = 1$ and $f(\boldsymbol{y}) = 0$.[3] They prove that if f has a matching of $\Omega(\sigma 2^n)$ violated edges, then their path tester will uncover a violation and reject with probability $\tilde{\Omega}(\sigma^3/\sqrt{n})$. (Roughly speaking, they show that about an $\Omega(\sigma)$ fraction of possible outcomes of $\boldsymbol{y}$, corresponding to the $\sigma 2^n$ upper endpoints of the edges in the matching, are such that with probability $\tilde{\Omega}(\sigma^2/\sqrt{n})$ over the random draw of $\boldsymbol{x}$, the pair $\boldsymbol{y}$ and $\boldsymbol{x}$ together constitute a violation.) On the other hand, if f does not have a matching of this size then (by the dichotomy theorem) it must have $\Omega((\varepsilon^2/\sigma)2^n)$ violated edges, so the edge tester of [3] (querying the endpoints of a uniform random edge) will hit a violated edge with probability $\Omega(\varepsilon^2/(\sigma n))$. Their final algorithm runs their path tester with probability $1/2$ and queries a random edge with probability $1/2$. Choosing σ suitably to equalize the two rejection probabilities, this is a two-query algorithm which succeeds in uncovering a violation for any ε-far-from-monotone function f with probability $\tilde{\Omega}(\varepsilon^{3/2}/n^{7/8})$, giving them a one-sided non-adaptive tester which makes $\tilde{O}(n^{7/8}/\varepsilon^{3/2})$ queries overall.

Our algorithm follows the same high-level framework described above, but differs from [2] by employing a different path tester. After selecting a random path $\boldsymbol{p}$, instead of (essentially) drawing two independent uniform points from the middle layers of the path as is done in [2], our path tester draws a *correlated* pair of points from $\boldsymbol{p}$. More precisely, it selects the first point $\boldsymbol{y}$ uniformly from the middle layers of $\boldsymbol{p}$, and preferentially selects the second point $\boldsymbol{x}$ from $\boldsymbol{p}$ in a way which favors points which are closer to $\boldsymbol{y}$. Via a careful analysis we are able to show that if f has a matching of $\Omega(\sigma 2^n)$ violated edges, then our path tester

[2] For our algorithmic result it will be more convenient to view Boolean functions as mapping $\{0,1\}^n$ to $\{0,1\}$.

[3] Here the "middle layers" of $\boldsymbol{p}$ are the points on the path that have $n/2 \pm O_\varepsilon(\sqrt{n})$ many coordinates which are 1; intuitively, at most an ε-fraction of all points in $\{0,1\}^n$ lie outside these "middle layers" of the hypercube. We note that the above description is a slight simplification of the actual [2] path tester, omitting some details which are not necessary at this stage of our description.

will uncover a violation and reject with probability $\tilde{\Omega}(\sigma^2/\sqrt{n}) \cdot \mathrm{poly}(\varepsilon)$. Roughly speaking, we show that if $\boldsymbol{y}$ is a uniform random upper endpoint of the $\sigma 2^n$ edges in the matching (which occurs with probability about σ), then the probability that our tester selects a point $\boldsymbol{x}$ which gives a violation with $\boldsymbol{y}$ is $\tilde{\Omega}(\sigma/\sqrt{n}) \cdot \mathrm{poly}(\varepsilon)$. Trading this off against the success probability of the edge tester using the dichotomy theorem, we obtain our improved query bound.

Organization of this paper. Our lower bound for the hypercube domain (i.e. Theorem 1) is established in Sections II and III. In Section II we define the two distributions $\mathcal{D}_{yes}$ and $\mathcal{D}_{no}$ and show that unless $q = \Omega(n^{1/5}(\log n)^{-2/5})$, any deterministic q-query algorithm cannot distinguish between the two distributions with non-negligible success probability. The key technical ingredient in our proof of the latter is a lemma that adapts the Valiant–Valiant multidimensional CLT for Wasserstein distance to our context; we prove this lemma in Section III. Theorem 2, showing that the same lower bound of $\tilde{\Omega}(n^{1/5})$ also applies to the query complexity of testers for monotonicity of functions $f : [m]^n \to \{0,1\}$ over general hypergrid domains, is established via a reduction to the $m = 2$ case (Theorem 1); we defer its proof to the full version of the paper.

Our algorithmic result is established in Section IV. In Section IV-A we describe two useful distributions over comparable pairs $(\boldsymbol{x}, \boldsymbol{y})$ from the middle layers of $\{0,1\}^n$ and bound the probability of having both points landing in a fixed set A of size $\sigma 2^n$. Then in Section IV-B we define the *score* of a point x with respect to a set A of points, and use the result of Section IV-A to lower bound the sum of $\mathrm{score}(x, A)$ over all points $x \in A$. Finally in Section IV-C we present our modified path tester as well as the analysis of its success probability, and we combine this tester and the dichotomy theorem of [2] to obtain our improved upper bound.

C. Preliminaries

All probabilities and expectations are with respect to the uniform distribution unless otherwise stated; we will use boldface letters (e.g. $\boldsymbol{x}$ and $\mathbf{X}$) to denote random variables. For a $q \times n$ matrix $Q \in \mathbb{R}^{q\times n}$, we write $Q_{i*} \in \mathbb{R}^n$ to denote its i-th row, $Q_{*j} \in \mathbb{R}^q$ its j-th column, and $Q_{i,j} \in \mathbb{R}$ its entry in the i-th column and j-th row. We use $\prec$ to denote the coordinate-wise partial order on $\{-1,1\}^n$, where $x \prec y$ iff $x_i \le y_i$ for all $i \in [n]$ and $x \ne y$. We also say that $x, y \in \{-1,1\}^n$ are *comparable* if $x \prec y$, $y \prec x$, or $x = y$. Given two functions $f, g : \{-1,1\}^n \to \{-1,1\}$ we will use $\mathrm{dist}(f,g)$ to denote the (normalized Hamming) distance $\mathbf{Pr}_{\boldsymbol{x}\in\{-1,1\}^n}[f(\boldsymbol{x}) \ne g(\boldsymbol{x})]$ between f and g.

Recall that $f : \{-1,1\}^n \to \{-1,1\}$ is monotone if $f(x) \le f(y)$ for all $x, y \in \{-1,1\}^n$ such that $x \prec y$. We say that f is *ε-close to monotone* if $\mathrm{dist}(f,g) \le \varepsilon$ for some monotone $g : \{-1,1\}^n \to \{-1,1\}$, and *$\varepsilon$-far from monotone* otherwise. A linear threshold function (LTF) over $\{-1,1\}^n$ is a function $f : \{-1,1\}^n \to \{-1,1\}$ that can be expressed as $f(x) = \mathrm{sign}(w\cdot x - \theta)$ for some $w_1, \ldots, w_n, \theta \in \mathbb{R}$. Here $\mathrm{sign} : \mathbb{R} \to \{-1,1\}$ is the sign function $\mathrm{sign}(t) = 1$ if $t \ge 0$ and $\mathrm{sign}(t) = -1$ if $t < 0$. For $f(x) = \mathrm{sign}(w\cdot x - \theta)$, an LTF over $\{-1,1\}^n$, it is straightforward to verify that if $w_i \ge 0$ for all $i \in [n]$ then f is monotone.

We need a few standard facts from probability theory:

Fact I.1 (Gaussian anti-concentration)**.** *Let $\mathcal{G}$ be a Gaussian with variance σ^2. Then for all $\varepsilon > 0$ it holds that* $\sup_{\theta\in\mathbb{R}} \left\{ \mathbf{Pr}\left[|\mathcal{G} - \theta| \le \varepsilon\sigma\right] \right\} \le \varepsilon$.

Theorem 4 (Berry–Esséen)**.** *Let $\mathbf{S} = \mathbf{X}_1 + \cdots + \mathbf{X}_n$ where $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent real-valued random variables with $\mathbf{E}[\mathbf{X}_j] = \mu_j$ and $\mathbf{Var}[\mathbf{X}_j] = \sigma_j^2$, and suppose that $|\mathbf{X}_j - \mathbf{E}[\mathbf{X}_j]| \le \tau$ with probability 1 for all $j \in [n]$. Let $\mathcal{G}$ be a Gaussian with mean $\sum_{j=1}^n \mu_j$ and variance $\sum_{j=1}^n \sigma_j^2$, matching those of $\mathbf{S}$. Then for all $\theta \in \mathbb{R}$, we have*

$$\left| \mathbf{Pr}[\mathbf{S} \le \theta] - \mathbf{Pr}[\mathcal{G} \le \theta] \right| \le \frac{O(\tau)}{\left(\sum_{j=1}^n \sigma_j^2\right)^{1/2}}.$$

II. The lower bound: Proof of Theorem 2

Let $\mathcal{D}_{yes}$ be the following distribution over monotone LTFs on $\{-1,1\}^n$: a draw $\boldsymbol{f}_{yes} \sim \mathcal{D}_{yes}$ is $\boldsymbol{f}_{yes}(x) = \mathrm{sign}(\boldsymbol{\sigma}_1 x_1 + \cdots + \boldsymbol{\sigma}_n x_n)$, where each $\boldsymbol{\sigma}_i$ is independently and uniformly chosen from $\{1,3\}$. Let $\mathcal{D}_{no}$ be a similar distribution over LTFs: $\boldsymbol{f}_{no}(x) = \mathrm{sign}(\boldsymbol{\nu}_1 x_1 + \cdots + \boldsymbol{\nu}_n x_n)$, but each $\boldsymbol{\nu}_i$ is independently chosen to be -1 with probability $1/10$, and $7/3$ with probability $9/10$. The following two propositions along with a standard application of Yao's minimax principle [20] yield Theorem 2:

Proposition II.1. *There exists a universal positive constant $\varepsilon_0 > 0$ such that with probability $1 - o_n(1)$, a random LTF $\boldsymbol{f}_{no} \sim \mathcal{D}_{no}$ satisfies $\mathrm{dist}(\boldsymbol{f}_{no}, g) > \varepsilon_0$ for all monotone Boolean functions $g : \{-1,1\}^n \to \{-1,1\}$.*

Proposition II.2. *Let $\mathcal{T}$ be any deterministic non-adaptive two-sided q-query algorithm for testing whether a black-box Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is* *monotone. Then*

$$\left| \Pr_{\boldsymbol{f}_{yes} \sim \mathcal{D}_{yes}} \left[\mathcal{T} \text{ accepts } \boldsymbol{f}_{yes} \right] - \Pr_{\boldsymbol{f}_{no} \sim \mathcal{D}_{no}} \left[\mathcal{T} \text{ accepts } \boldsymbol{f}_{no} \right] \right| = O\left(\frac{q^{5/4} (\log n)^{1/2}}{n^{1/4}} \right).$$

We defer the proof of Proposition II.1 to the full version of the paper; the remainder of this section will be devoted to proving Proposition II.2.

A. Proof of Proposition II.2

Let $\mathcal{T}$ be a deterministic non-adaptive q-query tester. We view its q queries as a $q \times n$ matrix $Q \in \{-1,1\}^{q\times n}$. Following the terminology of [21], we define a "Response Vector" random variable $\mathbf{R}_{yes} \in \{-1,1\}^q$, obtained by drawing $\boldsymbol{f}_{yes} = \text{sign}(\boldsymbol{\sigma}_1 x_1 + \cdots + \boldsymbol{\sigma}_n x_n)$ from $\mathcal{D}_{yes}$ and setting the i-th coordinate of $\mathbf{R}_{yes}$ to be

$$\boldsymbol{f}_{yes}(Q_{i*}) = \text{sign}(\boldsymbol{\sigma}_1 Q_{i,1} + \cdots + \boldsymbol{\sigma}_n Q_{i,n}),$$

and similarly $\mathbf{R}_{no} \in \{-1,1\}^q$ which is obtained by drawing $\boldsymbol{f}_{no} \sim \mathcal{D}_{no}$ and setting the i-th coordinate of $\mathbf{R}_{no}$ to be $\boldsymbol{f}_{no}(Q_{i*})$. By the definition of total variation distance, we can prove Proposition II.2 by showing that

$$d_{\text{TV}}(\mathbf{R}_{yes}, \mathbf{R}_{no}) = O\left(\frac{q^{5/4}(\log n)^{1/2}}{n^{1/4}} \right).$$

Let $\mathbf{S} \in \mathbb{R}^q$ be the random column vector $Q\boldsymbol{\sigma}$ where $\boldsymbol{\sigma}$ is uniform over $\{1,3\}^n$, and $\mathbf{T} \in \mathbb{R}^q$ be the random column vector $Q\boldsymbol{\nu}$ where $\boldsymbol{\nu}$ is drawn from the product distribution over $\{-1, 7/3\}^n$ where $\mathbf{Pr}[\boldsymbol{\nu}_i = -1] = 1/10$ for all $i \in [n]$. The Response Vector $\mathbf{R}_{yes}$ is determined by the orthant of $\mathbb{R}^q$ in which $\mathbf{S}$ lies (as each coordinate of $\mathbf{R}_{yes}$ is simply the sign of the respective coordinate of $\mathbf{S}$), and likewise $\mathbf{R}_{no}$ by the orthant of $\mathbb{R}^q$ in which $\mathbf{T}$ lies. Therefore it suffices for us to prove the following lemma:

Lemma II.3. *Let $\mathbf{S}, \mathbf{T} \in \mathbb{R}^q$ be defined as above. Then for any union $\mathcal{O}$ of orthants in $\mathbb{R}^q$, we have*

$$\big| \mathbf{Pr}[\mathbf{S} \in \mathcal{O}] - \mathbf{Pr}[\mathbf{T} \in \mathcal{O}] \big| = O\left(\frac{q^{5/4}(\log n)^{1/2}}{n^{1/4}} \right).$$

We will need the following multidimensional Berry–Esséen theorem. We defer its proof to Section III.

Theorem 5. *Let $\mathbf{S} = \mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)}$, where $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ are independent $\mathbb{R}^q$-valued random variables such that $|\mathbf{X}_i^{(j)} - \mathbf{E}[\mathbf{X}_i^{(j)}]| \le \tau$ with probability 1 for all $i \in [q], j \in [n]$. Let $\mathcal{G}$ be the q-dimensional Gaussian with the same mean and covariance matrix as $\mathbf{S}$. Let $\mathcal{O}$ be a union of orthants in $\mathbb{R}^q$. Then for all $r > 0$, the difference $|\mathbf{Pr}[\mathbf{S} \in \mathcal{O}] - \mathbf{Pr}[\mathcal{G} \in \mathcal{O}]|$ is at most*

$$O\left(\frac{\tau q^{3/2} \log n}{r} + \sum_{i=1}^{q} \frac{r + \tau}{\left(\sum_{j=1}^n \mathbf{Var}[\mathbf{X}_i^{(j)}] \right)^{1/2}} \right).$$

Proof of Lemma II.3 assuming Theorem 5: We begin by writing $\mathbf{S} = \mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)}$, where $\mathbf{X}^{(j)} = \boldsymbol{\sigma}_j \cdot Q_{*j}$ and $\boldsymbol{\sigma}_j$ is uniform over $\{1,3\}$; i.e. each $\mathbf{X}^{(j)}$ is independently Q_{*j} with probability $1/2$ and $3 \cdot Q_{*j}$ with probability $1/2$. Likewise we may express $\mathbf{T} = \mathbf{Y}^{(1)} + \cdots + \mathbf{Y}^{(n)}$, where $\mathbf{Y}^{(j)} = \boldsymbol{\nu}_j \cdot Q_{*j}$ and $\boldsymbol{\nu}_j$ is -1 with probability $1/10$ and $7/3$ with probability $9/10$.

We claim that the $\mathbf{X}^{(j)}$'s and $\mathbf{Y}^{(j)}$'s have matching means and covariance matrices. It suffices to check this for $\mathbf{X}^{(1)}$ and $\mathbf{Y}^{(1)}$, and we omit the routine calculation due to space considerations. As the $\mathbf{X}^{(j)}$'s and $\mathbf{Y}^{(j)}$'s have matching means and covariance matrices, so do their sums $\mathbf{S}$ and $\mathbf{T}$, and so Theorem 5 gives us a bound on the two differences $|\mathbf{Pr}[\mathbf{S} \in \mathcal{O}] - \mathbf{Pr}[\mathcal{G} \in \mathcal{O}]|$ and $|\mathbf{Pr}[\mathbf{T} \in \mathcal{O}] - \mathbf{Pr}[\mathcal{G} \in \mathcal{O}]|$ for the same q-dimensional Gaussian $\mathcal{G}$.

Recalling that $\mathbf{X}_i^{(j)} = \boldsymbol{\sigma}_j \cdot Q_{i,j}$ and $Q_{i,j} \in \{-1,1\}$, we have that $\mathbf{Var}[\mathbf{X}_i^{(j)}] = 1$ and likewise $\mathbf{Var}[\mathbf{Y}_i^{(j)}] = 1$. Therefore, two applications of Theorem 5 with $\tau := O(1)$ along with the triangle inequality yields the bound

$$\big| \mathbf{Pr}[\mathbf{S} \in \mathcal{O}] - \mathbf{Pr}[\mathbf{T} \in \mathcal{O}] \big| = O\left(\frac{q^{3/2} \log n}{r} + \frac{q(r+\tau)}{\sqrt{n}} \right)$$

for all $r > 0$. Choosing r to be $(qn)^{1/4}(\log n)^{1/2}$ then completes the proof. ■

III. Multidimensional Berry–Esséen via the Valiant–Valiant CLT

In this section, we prove Theorem 5 by adapting a recent multidimensional CLT of Valiant and Valiant [25] which bounds the *Wasserstein distance* between a sum of independent vector-valued random variables and a multidimensional Gaussian.

Definition 6 (Wasserstein distance)**.** The Wasserstein distance between two $\mathbb{R}^q$-valued random variables $\mathbf{S}$ and $\mathbf{T}$, denoted $d_W(\mathbf{S}, \mathbf{T})$, is defined to be:

$$d_W(\mathbf{S}, \mathbf{T}) = \inf_{\mathcal{D}} \left\{ \mathop{\mathbf{E}}_{\mathcal{D}} \left[\|\mathbf{U} - \mathbf{V}\|_2 \right] \right\},$$

where the infimum is taken over all couplings $\mathcal{D}$ of $\mathbf{S}$ and $\mathbf{T}$, i.e. all joint distributions $\mathcal{D}$ of pairs of $\mathbb{R}^q$-valued random variables $(\mathbf{U}, \mathbf{V})$ with marginals distributed according to $\mathbf{S}$ and $\mathbf{T}$ respectively.

Valiant and Valiant [25] recently used Stein's method to prove the following CLT for Wasserstein distance:

Theorem 7 (Valiant-Valiant CLT). *Let* $\mathbf{S} = \mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)}$*, where* $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ *are independent* $\mathbb{R}^q$*-valued random variables, and suppose* $\|\mathbf{X}^{(j)} - \mathbf{E}[\mathbf{X}^{(j)}]\|_2 \le \beta$ *with probability* 1 *for any* $j \in [n]$*. Then*

$$d_W(\mathbf{S}, \mathcal{G}) \le O(\beta q \log n),$$

where $\mathcal{G}$ *is the* q*-dimensional Gaussian with the same mean and covariance matrix as* $\mathbf{S}$*.*

Proof of Theorem 5: We define

$$W_r := \{x \in \mathbb{R}^q \colon |x_i| \le r \text{ for some } i \in [q]\}$$

to be the radius-r region around the orthant boundaries, and partition $\mathcal{O}$ into $\mathcal{O}_{bd} := \mathcal{O} \cap W_r$ (the points in $\mathcal{O}$ that lie close to the orthant boundaries) and $\mathcal{O}_{in} := \mathcal{O} \setminus W_r$ (the points that lie far away from the orthant boundaries). We have

$$\begin{aligned}
&\big|\, \mathbf{Pr}[\mathbf{S} \in \mathcal{O}] - \mathbf{Pr}[\mathcal{G} \in \mathcal{O}] \big| \\
&= \big|(\mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{in}] + \mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{bd}]) \\
&\qquad - (\mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{in}] + \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{bd}])\big| \\
&\le \underbrace{\big|\, \mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{in}] - \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{in}] \big|}_{\Delta} \\
&\qquad + \underbrace{\mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{bd}] + \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{bd}]}_{\Gamma}.
\end{aligned}$$

We next bound the quantities Δ and Γ separately.

For Γ, we have that

$$\begin{aligned}
\Gamma &\le \sum_{i \in [q]} \mathbf{Pr}\left[\mathbf{S}_i \in [-r, r]\right] + \mathbf{Pr}\left[\mathcal{G}_i \in [-r, r]\right] \\
&\le \sum_{i \in [q]} 2\, \mathbf{Pr}\left[\mathcal{G}_i \in [-r, r]\right] \\
&\qquad + \big|\, \mathbf{Pr}\left[\mathbf{S}_i \in [-r, r]\right] - \mathbf{Pr}\left[\mathcal{G}_i \in [-r, r]\right] \big| \\
&\le \sum_{i \in [q]} \frac{O(r)}{\big(\sum_{j=1}^n \mathbf{Var}[\mathbf{X}_i^{(j)}]\big)^{1/2}} + \frac{O(\tau)}{\big(\sum_{j=1}^n \mathbf{Var}[\mathbf{X}_i^{(j)}]\big)^{1/2}} \\
&= \sum_{i \in [q]} \frac{O(r + \tau)}{\big(\sum_{j=1}^n \mathbf{Var}[\mathbf{X}_i^{(j)}]\big)^{1/2}},
\end{aligned}$$

where the first inequality is a union bound over all q dimensions, and the third uses Fact I.1 (Gaussian anti-concentration), the fact that $\mathcal{G}_i$ is a Gaussian of variance $\sum_{j=1}^n \mathbf{Var}[\mathbf{X}_i^{(j)}]$, and Theorem 4 (Berry–Esséen).

For Δ, assume without loss of generality (a symmetrical argument works in the other case) that $\mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{in}] \ge \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{in}]$, so $\Delta = \mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{in}] - \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{in}]$. Let $\mathcal{D}$ be the coupling of $\mathbf{S}$ and $\mathcal{G}$ that achieves the infimum in Definition 6, so $\mathcal{D}$ is the joint distribution of a pair $(\mathbf{U}, \mathbf{V})$ of $\mathbb{R}^q$-valued random variables with marginals distributed according to $\mathbf{S}$ and $\mathcal{G}$ respectively. Since

$$\int_{\mathcal{O}_{in}} \int_{\mathbb{R}^q} \mathcal{D}(u, v)\, dv\, du = \mathbf{Pr}[\mathbf{S} \in \mathcal{O}_{in}]$$

and

$$\begin{aligned}
&\int_{\mathcal{O}_{in}} \int_{\mathcal{O}_{in}} \mathcal{D}(u, v)\, dv\, du \\
&\le \int_{\mathbb{R}^q} \int_{\mathcal{O}_{in}} \mathcal{D}(u, v)\, dv\, du = \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{in}],
\end{aligned}$$

it follows that

$$\begin{aligned}
&\int_{\mathcal{O}_{in}} \int_{\mathbb{R}^q \setminus \mathcal{O}_{in}} \mathcal{D}(u, v)\, dv\, du \qquad (1) \\
&= \int_{\mathcal{O}_{in}} \int_{\mathbb{R}^q} \mathcal{D}(u, v)\, dv\, du - \int_{\mathcal{O}_{in}} \int_{\mathcal{O}_{in}} \mathcal{D}(u, v)\, dv\, du \ge \Delta
\end{aligned}$$

Next we define the quantities

$$\Delta_{\textit{near}}(\mathcal{D}) := \int_{\mathcal{O}_{in}} \int_{\mathcal{O}_{bd}} \mathcal{D}(u, v)\, dv\, du \quad \text{and}$$

$$\Delta_{\textit{far}}(\mathcal{D}) := \int_{\mathcal{O}_{in}} \int_{\mathbb{R}^q \setminus \mathcal{O}} \mathcal{D}(u, v)\, dv\, du.$$

Note that $\Delta_{\textit{near}}(\mathcal{D})$ and $\Delta_{\textit{far}}(\mathcal{D})$ sum to the quantity in (1), and so $\Delta_{\textit{near}}(\mathcal{D}) + \Delta_{\textit{far}}(\mathcal{D}) \ge \Delta$. (In words, since $\mathbf{S}$ places Δ more mass on $\mathcal{O}_{in}$ than $\mathcal{G}$ does, any scheme $\mathcal{D}$ of moving the mass of $\mathbf{S}$ to obtain $\mathcal{G}$ must move at least Δ amount from within $\mathcal{O}_{in}$ to outside it. $\Delta_{\textit{near}}(\mathcal{D})$ is the amount moved from within $\mathcal{O}_{in}$ to $\mathcal{O}$'s boundary $\mathcal{O}_{bd}$, and $\Delta_{\textit{far}}(\mathcal{D})$ is the rest, moved from within $\mathcal{O}_{in}$ to locations entirely out of $\mathcal{O}$.) Since $\|u - v\|_2 \ge r$ for any pair of points $u \in \mathcal{O}_{in}$ and $v \notin \mathcal{O}$, it follows that

$$d_W(\mathbf{S}, \mathcal{G}) \ge r \cdot \Delta_{\textit{far}}(\mathcal{D}).$$

We consider two cases, depending on the relative magnitudes of $\Delta_{\textit{near}}(\mathcal{D})$ and $\Delta_{\textit{far}}(\mathcal{D})$. If $\Delta_{\textit{far}}(\mathcal{D}) \ge \Delta_{\textit{near}}(\mathcal{D})$, we first observe that for all $j \in [n]$ we have $\|\mathbf{X}^{(j)} - \mathbf{E}[\mathbf{X}^{(j)}]\|_2 \le \tau\sqrt{q}$ with probability 1, as each of its q coordinates $i \in [q]$ satisfies $|\mathbf{X}_i^{(j)} - \mathbf{E}[\mathbf{X}_i^{(j)}]| \le \tau$ with probability 1 by the assumption of the theorem. Therefore, we may apply Theorem 7 (Valiant–Valiant CLT), with $\beta := \tau\sqrt{q}$, to get

$$r \cdot \frac{\Delta}{2} \le r \cdot \Delta_{\textit{far}}(\mathcal{D}) \le d_W(\mathbf{S}, \mathcal{G}) = O(\tau q^{3/2} \log n)$$

and hence $\Delta = O((\tau q^{3/2} \log n)/r)$, which along with our upper bound on Γ completes the proof. If on the other hand $\Delta_{\textit{near}}(\mathcal{D}) > \Delta_{\textit{far}}(\mathcal{D})$, then

$$\frac{\Delta}{2} \le \Delta_{\textit{near}}(\mathcal{D}) \le \int_{\mathbb{R}^q} \int_{\mathcal{O}_{bd}} \mathcal{D}(u, v)\, dv\, du = \mathbf{Pr}[\mathcal{G} \in \mathcal{O}_{bd}] \le \Gamma$$

and again our bound on Γ completes the proof. ■

IV. THE ALGORITHM

Throughout the proof of our upper bound, we will assume that $1/n \le \varepsilon \le 1/2$. Note that this is without loss of generality, since if $\varepsilon < 1/n$ then the edge tester alone succeeds with probability $\Omega(\varepsilon/n) = \Omega(\varepsilon^2)$, and if $\varepsilon > 1/2$ then every f is ε-close to one of the two constant functions, both of which are monotone.

For our upper bound it will be more convenient to view Boolean functions as mapping $\{0,1\}^n$ to $\{0,1\}$. Given $x, y \in \{0,1\}^n$ we write $\|x\|_1$ to denote $\sum_{i=1}^n x_i$, the number of 1s in x, and $\|x-y\|_1$ to denote $|\{i \in [n] \colon x_i \ne y_i\}|$, the ℓ_1-distance between x and y. Given $1/n \le \varepsilon \le 1/2$, we fix

$$d(n,\varepsilon) := 2\left\lceil \sqrt{2n\ln(100/\varepsilon)}\right\rceil = O\big(\sqrt{n\ln(1/\varepsilon)}\big),$$

and will denote $d(n,\varepsilon)$ simply by d when the distance parameter ε is clear from the context. For each $i \in \{0,1,\dots,n\}$ we use $L_i := \{x \in \{0,1\}^n \colon \|x\|_1 = i\}$ to denote the i-th layer, and refer to

$$L_{\text{mid}} := \big\{x \in L_i \colon i \in [(n-d)/2, (n+d)/2]\big\}$$

as the middle layers of the hypercube $\{0,1\}^n$. A standard Chernoff bound gives

$$|\{0,1\}^n \setminus L_{\text{mid}}| \le (\varepsilon/50)\cdot 2^n.$$

Finally, by a "path" we always mean a directed path of $n+1$ adjacent vertices from 0^n up to 1^n.

A. Two useful distributions over comparable pairs

Let $\mathcal{D} = \mathcal{D}_{n,\varepsilon}$ denote the following distribution over comparable pairs $(\boldsymbol{x},\boldsymbol{y}) \in L_{\text{mid}} \times L_{\text{mid}}$:

1) First pick a path $\boldsymbol{p}$ uniformly from the collection of all paths going from 0^n to 1^n.
2) Pick $\boldsymbol{x}$ and $\boldsymbol{y}$ independently and uniformly from

$$\boldsymbol{p}_{\text{mid}} := \{z \in \boldsymbol{p} \colon z \in L_{\text{mid}}\}. \tag{2}$$

This distribution is a slight variant of the one induced by the [2] path tester, which takes a parameter σ as input and disallows pairs (x,y) for which $\|x-y\|_1$ is too small relative to σ. Our new tester will *not* sample from $\mathcal{D}$ (see Section IV-C), but we will use $\mathcal{D}$ in our analysis. (Note that $\boldsymbol{x}=\boldsymbol{y}$ with positive probability under $\mathcal{D}$.)

If $\boldsymbol{x}$ and $\boldsymbol{y}$ were chosen independently and uniformly from $\{0,1\}^n$, then the probability that they both land in a fixed set A of $\sigma 2^n$ points, for some $\sigma \in (0,1)$, would be σ^2. The following lemma states that the probability is not much lower for a pair drawn from $\mathcal{D}$ (its proof is essentially identical to that of Claim 2.2.1 of [2], and we omit it in this version):

Lemma IV.1. *Let $A \subseteq L_{\text{mid}}$ with $|A| = \sigma 2^n$. Then*

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\leftarrow\mathcal{D}}[\boldsymbol{x},\boldsymbol{y} \in A] = \Omega\big(\sigma^2 \ln^{-1}(1/\varepsilon)\big).$$

For our analysis the following distribution $\mathcal{D}' = \mathcal{D}'_{n,\varepsilon}$ over comparable pairs $(\boldsymbol{x},\boldsymbol{y}) \in L_{\text{mid}} \times L_{\text{mid}}$ in the middle layers comes in handy:

1) Pick a point $\boldsymbol{x}$ uniformly at random from L_{mid}.
2) Then pick a path $\boldsymbol{p}$ uniformly from the collection of all paths going through 0^n, $\boldsymbol{x}$, and 1^n.
3) Pick $\boldsymbol{y}$ uniformly from $\boldsymbol{p}_{\text{mid}}$ as defined in (2).

Note that $\mathcal{D}'$ is not exactly the same as $\mathcal{D}$, as picking a uniformly random $\boldsymbol{x}$ from the middle layers $\boldsymbol{p}_{\text{mid}}$ of a uniformly random path $\boldsymbol{p}$ does not induce a uniform distribution over L_{mid}. However, the following corollary allows us to switch between these essentially-equivalent distributions at the cost of a $O(1/\varepsilon^4)$ factor; we defer its proof to the full version of the paper.

Corollary IV.2. *Let $A \subseteq L_{\text{mid}}$ with $|A| = \sigma 2^n$. Then*

$$\Pr_{(\boldsymbol{x},\boldsymbol{y})\leftarrow\mathcal{D}'}[\boldsymbol{x},\boldsymbol{y} \in A] = \Omega\big(\sigma^2\varepsilon^4 \ln^{-1}(1/\varepsilon)\big).$$

B. Density and score

We will need the following definition to give a more detailed analysis on the consequence of Corollary IV.2, which is key to the analysis of our monotonicity tester described in Section IV-C.

Definition 8 (density and score)**.** Let $A \subseteq \{0,1\}^n$ be a set of points. For all $x \in \{0,1\}^n$ and $k \in \{0,1,\dots,n\}$, we define the following quantities:

$$\text{dens}_k^{\downarrow}(x,A) := \begin{cases} \Pr\limits_{\substack{\boldsymbol{y}\preceq x\\ \|\boldsymbol{y}-x\|_1=k}}[\boldsymbol{y}\in A] & \text{if } k \le \|x\|_1 \\ 0 & \text{otherwise} \end{cases}$$

and similarly

$$\text{dens}_k^{\uparrow}(x,A) := \begin{cases} \Pr\limits_{\substack{\boldsymbol{y}\succeq x\\ \|\boldsymbol{y}-x\|_1=k}}[\boldsymbol{y}\in A] & \text{if } k \le n-\|x\|_1 \\ 0 & \text{otherwise.} \end{cases}$$

We also define

$$\text{score}^{\downarrow}(x,A) := \sum_{k=0}^{n} \text{dens}_k^{\downarrow}(x,A)$$
$$\text{score}^{\uparrow}(x,A) := \sum_{k=1}^{n} \text{dens}_k^{\uparrow}(x,A)$$

and refer to $\text{score}^{\downarrow}(x,A)$ as the *downward A-score of x* and $\text{score}^{\uparrow}(x,A)$ as its *upward A-score*.

We point out the asymmetry between the definitions of $\text{score}^{\downarrow}(x,A)$ and $\text{score}^{\uparrow}(x,A)$: the first is summed over k starting at 0, whereas the second is summed over k starting at 1. (Note that $\mathrm{dens}_0^{\downarrow}(x, A) = \mathrm{dens}_0^{\uparrow}(x, A) = \mathbf{1}[x \in A]$.) We will need the fact that both the upward and downward A-scores of any $x \in \{0,1\}^n$ are at most $d = d(n, \varepsilon)$ when $A \subseteq L_{\mathrm{mid}}$.

We defer the proofs of the next two lemmas to the full version. The first relates the distribution $\mathcal{D}'$ (more precisely, the distribution over $\boldsymbol{y}$ that is induced by conditioning on a particular outcome of $\boldsymbol{x}$) to the notion of score:

Lemma IV.3. *Let $A \subseteq L_{\mathrm{mid}}$ be a set of $\sigma 2^n$ points and fix a point $x^* \in L_{\mathrm{mid}}$. Then*

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y}) \leftarrow \mathcal{D}'} \left[\boldsymbol{y} \in A \mid \boldsymbol{x} = x^*\right] = \frac{1}{\Theta(\sqrt{n \ln(1/\varepsilon)})} \left(\mathrm{score}^{\downarrow}(x^*, A) + \mathrm{score}^{\uparrow}(x^*, A)\right).$$

The second lower bounds the expected downward A-score of an $\boldsymbol{x}$ drawn uniformly at random from A:

Lemma IV.4. *Let $\varepsilon \geq 1/n$ and $A \subseteq L_{\mathrm{mid}}$ be a set of $\sigma 2^n$ points. Then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x} \in A} \left[\mathrm{score}^{\downarrow}(\boldsymbol{x}, A)\right] = \Omega\left(\frac{\varepsilon^8 \sigma \sqrt{n}}{\sqrt{\ln(1/\varepsilon)}}\right).$$

The conclusion of Lemma IV.4 can be equivalently rewritten as the following sum:

$$\sum_{x \in A} \mathrm{score}^{\downarrow}(x, A) = \Omega\left(\frac{\varepsilon^8 \sigma^2 \sqrt{n}\, 2^n}{\sqrt{\ln(1/\varepsilon)}}\right). \tag{3}$$

We may express the downward A-score $\mathrm{score}^{\downarrow}(x, A)$ as a sum over $m+1$ "buckets" of exponentially increasing size as follows:

$$\mathrm{score}^{\downarrow}(x, A) = \sum_{i=0}^{m} \sum_{k \in B_i} \mathrm{dens}_k^{\downarrow}(x, A) \tag{4}$$

where $B_0 = \{0\}$ and $B_i = \{2^{i-1}, \ldots, 2^i - 1\}$ for each $i \in [m]$ and $m = \lceil \log(n+1) \rceil$. It will be useful for us to focus on a particular bucket $\ell \in \{0, 1, \ldots, m\}$ such that the overall sum of $\mathrm{score}^{\downarrow}(x, A)$ in (3) has a "large" contribution from the ℓ-th bucket. A straightforward argument, exploiting the fact that there are only logarithmically many buckets, lets us achieve this without losing too much in the sum:

Corollary IV.5. *Let $\varepsilon \geq 1/n$ and $A \subseteq L_{\mathrm{mid}}$ be a set of $\sigma 2^n$ points. There exists $\ell \leq m$ such that*

$$\sum_{x \in A} \sum_{k \in B_\ell} \mathrm{dens}_k^{\downarrow}(x, A) = \Omega\left(\frac{\varepsilon^8 \sigma^2 \sqrt{n}\, 2^n}{(\log n)\sqrt{\ln(1/\varepsilon)}}\right). \tag{5}$$

Proof: This follows from (3), (4), and the fact that there are only $m+1$ many buckets. ■

Corollary IV.5 gives a lower bound on the sum of downward A-scores of points $x \in A$ coming from a certain bucket B_ℓ. Our next corollary uses this to give a lower bound on the sum of downward A-scores of points $y \in A_u$ from (essentially) the same bucket B_ℓ, where A_u is an "upper vertex boundary" of A in the following sense: there exists an $|A|$-sized matching M of edges (x, y) where $x \prec y$, $x \in A$ and $y \in A_u$.

Corollary IV.6. *Let $\varepsilon \geq 1/n$ and M be a matching of $\sigma 2^n$ edges in the middle layers. Let*

$$A := \left\{x \in \{0,1\}^n \colon x \prec y \text{ and } (x,y) \in M\right\} \quad \text{and}$$
$$A_u := \left\{y \in \{0,1\}^n \colon y \succ x \text{ and } (x,y) \in M\right\}$$

denote the lower and upper endpoints of edges in M, respectively. For each bucket B_i, $i \in \{0, 1, \ldots, m\}$, we let $B_i' := \{j+1 : j \in B_i\}$. Then there exists an integer $\ell \leq m$ such that

$$\sum_{y \in A_u} \sum_{k \in B_\ell'} \mathrm{dens}_k^{\downarrow}(y, A) = \Omega\left(\frac{2^{\ell+n} \varepsilon^8 \sigma^2}{(\log n)\sqrt{n \ln(1/\varepsilon)}}\right). \tag{6}$$

Proof: By Corollary IV.5, there exists an $\ell \leq m$ such that A satisfies (5). Next for each edge $(x, y) \in M$ we have that

$$\begin{aligned} \mathrm{dens}_{k+1}^{\downarrow}(y, A) &= \mathop{\mathbf{Pr}}_{\substack{\boldsymbol{z} \prec y \\ \|\boldsymbol{z}-y\|_1 = k+1}} [\boldsymbol{z} \in A] \\ &\geq \frac{\binom{\|x\|_1}{k}}{\binom{\|y\|_1}{k+1}} \cdot \mathop{\mathbf{Pr}}_{\substack{\boldsymbol{z} \prec x \\ \|\boldsymbol{z}-x\|_1 = k}} [\boldsymbol{z} \in A] \\ &= \frac{(k+1) \cdot \mathrm{dens}_k^{\downarrow}(x, A)}{\|x\|_1 + 1}. \end{aligned}$$

Therefore, by (5) we have

$$\begin{aligned} \sum_{y \in A_u} \sum_{k \in B_\ell'} \mathrm{dens}_k^{\downarrow}(y, A) &= \sum_{y \in A_u} \sum_{k \in B_\ell} \mathrm{dens}_{k+1}^{\downarrow}(y, A) \\ &\geq \sum_{x \in A_u} \sum_{k \in B_\ell} \frac{(k+1)\, \mathrm{dens}_k^{\downarrow}(x, A)}{\|x\|_1 + 1} \\ &= \Omega\left(\frac{\varepsilon^8 \sigma^2 \sqrt{n}\, 2^n}{(\log n)\sqrt{\ln(1/\varepsilon)}} \cdot \frac{2^\ell}{n}\right). \end{aligned}$$

This completes the proof. ■

C. The weighted path tester and its analysis

Given a Boolean function $f : \{0,1\}^n \to \{0,1\}$, we recall that a pair (x, y) of points is a *violated pair with respect to f if $x \prec y$ and $f(x) > f(y)$*. Our algorithm `weighted-path-tester` for monotonicity testing proceeds as follows:

> `weighted-path-tester`:
>
> 1) Pick a point $\boldsymbol{y}$ uniformly from L_{mid}.
> 2) Pick $\boldsymbol{\ell} \in \{0, 1, \ldots, m = \lceil \log(n+1) \rceil\}$ uniformly, and pick $\boldsymbol{k} \in B'_{\boldsymbol{\ell}}$ uniformly.
> 3) Pick a path $\boldsymbol{p}$ uniformly from the collection of all paths going through $0^n, \boldsymbol{y}$ and 1^n, and set $\boldsymbol{x}$ to be the (unique) point on $\boldsymbol{p}$ that has $\boldsymbol{x} \prec \boldsymbol{y}$ and $\|\boldsymbol{x} - \boldsymbol{y}\|_1 = \boldsymbol{k}$.
> 4) Reject iff $(\boldsymbol{x}, \boldsymbol{y})$ is a violated pair.

Note that an equivalent formulation of step 3) above is that $\boldsymbol{x}$ is drawn uniformly from

$$\big\{z \in \{0,1\}^n \colon z \prec \boldsymbol{y} \text{ and } \|\boldsymbol{y} - z\|_1 = \boldsymbol{k}\big\}.$$

Below we show that if there is a $(\sigma 2^n)$-sized matching M of violated edges of f in the middle layers of the hypercube, then the tester above succeeds in finding a violated pair with probability roughly $\Omega(\sigma^2/\sqrt{n})$.

Proposition IV.7. *Let $f : \{0,1\}^n \to \{0,1\}$ and $\varepsilon \geq 1/n$. Suppose there exists a $(\sigma 2^n)$-sized matching M of violated edges of f all lying in the middle layers of the hypercube. Then* `weighted-path-tester` *above succeeds (i.e., samples $\boldsymbol{x}$ and $\boldsymbol{y}$ that form a violated pair with respect to f) with probability*

$$\Omega\left(\frac{\varepsilon^8 \sigma^2}{(\log^2 n)\sqrt{n \ln(1/\varepsilon)}}\right). \tag{7}$$

Proof: Let A be the set of 1-endpoints of edges in the matching M, and A_u be the 0-endpoints in M, respectively. Let $\mathcal{D}^w$ denote the distribution over comparable pairs $(\boldsymbol{x}, \boldsymbol{y}) \in L_{\mathrm{mid}} \times L_{\mathrm{mid}}$ as induced by our algorithm `weighted-path-tester` above.

We note that every pair $(x, y) \in A \times A_u$ that satisfies $x \prec y$ is a violated pair with respect to f. Therefore, `weighted-path-tester` succeeds with probability at least

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y}) \leftarrow \mathcal{D}^w}\big[\boldsymbol{y} \in A_u, \boldsymbol{x} \in A\big].$$

Applying Corollary IV.6, we know there exists an $\ell^* \in \{0, 1, \ldots, m\}$ such that

$$\sum_{y \in A_u} \sum_{k \in B'_{\ell^*}} \mathrm{dens}_k^{\downarrow}(y, A) = \Omega\left(\frac{2^{\ell^*+n}\varepsilon^8\sigma^2}{(\log n)\sqrt{n \ln(1/\varepsilon)}}\right). \tag{8}$$

Conditioning on the event of $\boldsymbol{y} = y$ and $\boldsymbol{k} = k$, the probability of $\boldsymbol{x} \in A$ is $\mathrm{dens}_k^{\downarrow}(y, A)$. As $\boldsymbol{y}, \boldsymbol{\ell}, \boldsymbol{k}$ are all sampled uniformly, `weighted-path-tester` succeeds with probability at least

$$\begin{aligned}
&\mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y}) \leftarrow \mathcal{D}^w}\big[\boldsymbol{y} \in A_u, \boldsymbol{x} \in A\big] \\
&= \mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y}) \leftarrow \mathcal{D}^w}[\boldsymbol{y} \in A_u] \cdot \mathop{\mathbf{Pr}}_{(\boldsymbol{x},\boldsymbol{y}) \leftarrow \mathcal{D}^w}\big[\boldsymbol{x} \in A \mid \boldsymbol{y} \in A_u\big] \\
&= \frac{|A_u|}{|L_{\mathrm{mid}}|} \cdot \frac{1}{|A_u|} \sum_{y \in A_u} \frac{1}{m+1} \sum_{\ell=0}^{m} \frac{1}{|B'_\ell|} \sum_{k \in B'_\ell} \mathrm{dens}_k^{\downarrow}(y, A) \\
&\geq \frac{1}{(m+1)|L_{\mathrm{mid}}||B'_{\ell^*}|} \cdot \sum_{y \in A_u} \sum_{k \in B'_{\ell^*}} \mathrm{dens}_k^{\downarrow}(y, A) \\
&= \Omega\left(\frac{2^{\ell^*+n}\varepsilon^8\sigma^2}{(\log n)\sqrt{n \ln(1/\varepsilon)}} \cdot \frac{1}{(\log n)2^{\ell^*+n}}\right) \\
&= \Omega\left(\frac{\varepsilon^8\sigma^2}{(\log^2 n)\sqrt{n \ln(1/\varepsilon)}}\right).
\end{aligned}$$

This finishes the proof. ■

Finally we combine Proposition IV.7 with the dichotomy theorem of [2] to prove Theorem 3. To state the latter, we use $v2^n$ to denote the total number of violated edges in f, and use $\sigma 2^n$ to denote the size of the largest matching of violated edges in the middle layers. Then

Theorem 9 (Theorem 2.4 of [2])**.** *For any Boolean f that is ε-far from monotone, $v \cdot \sigma = \Omega(\varepsilon^2)$.*

Proof of Theorem 3: As mentioned at the beginning of Section IV, we may assume without loss of generality that $\varepsilon \geq 1/n$ since otherwise the edge tester alone succeeds with probability $\Omega(\varepsilon/n) = \Omega(\varepsilon^2)$. When $\varepsilon \geq 1/n$, our tester flips a coin, runs the edge tester with probability $1/2$, and runs `weighted-path-tester` with probability $1/2$. Given v and σ as defined above, the success probability of the edge tester is $\Omega(v/n)$; the success probability of `weighted-path-tester` is given in (7). It follows from Theorem 9 that the average of these two is at least

$$\Omega\left(\frac{\varepsilon^4}{n^{5/6}(\log^{2/3} n)(\ln(1/\varepsilon))^{1/6}}\right).$$

This finishes the proof of Theorem 3. ■

Acknowledgment

We thank Eric Blais for a helpful discussion that led to an improvement of our lower bound for general hypergrid domains.

Xi Chen is supported by NSF CCF-1149257 and a Sloan fellowship. Rocco Servedio and Li-Yang Tan are supported by NSF CCF-1115703 and CCF-1319788.

REFERENCES

[1] E. Fischer, E. Lehman, I. Newman, S. Raskhodnikova, R. Rubinfeld, and A. Samorodnitsky, "Monotonicity testing over general poset domains," in *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, 2002, pp. 474–483.

[2] D. Chakrabarty and C. Seshadhri, "A $o(n)$ monotonicity tester for boolean functions over the hypercube," in *Proceedings of the 45th ACM Symposium on Theory of Computing*, 2013, pp. 411–418.

[3] O. Goldreich, S. Goldwasser, E. Lehman, and D. Ron, "Testing monotonicity," in *Proceedings of the 39th IEEE Symposium on Foundations of Computer Science*, 1998, pp. 426–435.

[4] Y. Dodis, O. Goldreich, E. Lehman, S. Raskhodnikova, D. Ron, and A. Samorodnitsky, "Improved testing algorithms for monotonocity," in *Proceedings of the 3rd Workshop on Randomization and Approximation Techniques in Computer Science*, 1999, pp. 97–108.

[5] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samordinsky, "Testing monotonicity," *Combinatorica*, vol. 20, no. 3, pp. 301–337, 2000.

[6] F. Ergün, S. Kannan, S. R. Kumar, R. Rubinfeld, and M. Vishwanthan, "Spot-checkers," *Journal of Computer and System Sciences*, vol. 60, no. 3, pp. 717–751, 2000, earlier version in STOC'96.

[7] E. Fischer, "On the strength of comparisons in property testing," *Information and Computation*, vol. 189, no. 1, pp. 107–116, 2004.

[8] T. Batu, R. Kumar, and R. Rubinfeld, "Sublinear algorithms for testing monotone and unimodal distributions," in *Proceedings of the 36th ACM Symposium on Theory of Computing*, 2004, pp. 381–390.

[9] N. Ailon, B. Chazelle, S. Comandur, and D. Liu, "Estimating the distance to a monotone function," *Random Structures and Algorithms*, vol. 31, no. 3, pp. 371–383, 2007.

[10] S. Halevy and E. Kushilevitz, "Testing monotonicity over graph products," *Random Structures and Algorithms*, vol. 33, no. 1, pp. 44–67, 2008.

[11] R. Rubinfeld and R. A. Servedio, "Testing monotone high-dimensional distributions," *Random Structures and Algorithms*, vol. 34, no. 1, pp. 24–44, 2009.

[12] E. Blais, J. Brody, and K. Matulef, "Property testing lower bounds via communication complexity," *Computational Complexity*, vol. 21, no. 2, pp. 311–358, 2012.

[13] J. Briët, S. Chakraborty, D. García-Soriano, and A. Matsliah, "Monotonicity testing and shortest-path routing on the cube," *Combinatorica*, vol. 32, no. 1, pp. 35–53, 2012.

[14] D. Ron, R. Rubinfeld, M. Safra, A. Samorodnitsky, and O. Weinstein, "Approximating the influence of monotone Boolean functions in $O(\sqrt{n})$ query complexity," *ACM Transactions on Computation Theory*, vol. 4, no. 4, p. 11, 2012.

[15] D. Chakrabarty and C. Seshadhri, "Optimal bounds for monotonicity and lipschitz testing over hypercubes and hypergrids," in *Proceedings of the 45th ACM Symposium on Theory of Computing*, 2013, pp. 419–428.

[16] ——, "An optimal lower bound for monotonicity testing over hypergrids," in *Proceedings of APPROX-RANDOM*, 2013, pp. 425–435.

[17] E. Blais, S. Raskhodnikova, and G. Yaroslavtsev, "Lower bounds for testing properties of functions on hypergrid domains," *Electronic Colloquium on Computational Complexity*, vol. 20, p. 36, 2013.

[18] K. Matulef, R. O'Donnell, R. Rubinfeld, and R. Servedio, "Testing halfspaces," *SIAM Journal on Computing*, vol. 39, no. 5, pp. 2004–2047, 2010.

[19] ——, "Testing ± 1-weight halfspaces," in *Proceedings of APPROX-RANDOM*, 2009, pp. 646–657.

[20] A. Yao, "Probabilistic computations: Towards a unified measure of complexity," in *Proceedings of the 9th Annual ACM Symposium on Theory of Computing*, 1977, pp. 222–227.

[21] E. Blais and R. O'Donnell, "Lower bounds for testing function isomorphism," in *Proceedings of the 25th Annual IEEE Conference on Computational Complexity*, 2010, pp. 235–246.

[22] P. Gopalan, R. O'Donnell, Y. Wu, and D. Zuckerman, "Fooling functions of halfspaces under product distributions," in *Proceedings of the 25th Annual IEEE Conference on Computational Complexity*, 2010, pp. 223–234.

[23] E. Mossel, "Gaussian bounds for noise correlation of functions and tight analysis of Long Codes," in *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, 2008, pp. 156–165.

[24] D. Ron and R. A. Servedio, "Exponentially improved algorithms and lower bounds for testing signed majorities," in *Proceedings of the 24th ACM-SIAM Symposium on Discrete Algorithms*, 2013, pp. 1319–1336.

[25] G. Valiant and P. Valiant, "Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs," in *Proceedings of the 43rd ACM Symposium on Theory of Computing*, 2011, pp. 685–694.