Improved Testing Algorithms for Monotonicity

Yevgeniy Dodis* Oded Goldreich† Eric Lehman‡ Sofya Raskhodnikova§
Dana Ron¶ Alex Samorodnitsky‖

December 20, 1999

Abstract

We present improved algorithms for testing monotonicity of functions. Namely, given the ability to query an unknown function $f : \Sigma^n \mapsto \Xi$, where Σ and Ξ are finite ordered sets, the test always accepts a monotone f, and rejects f with high probability if it is ϵ-far from being monotone (i.e., every monotone function differs from f on more than an ϵ fraction of the domain). For any $\epsilon > 0$, the query and time complexities of the test are $O((n/\epsilon) \cdot \log|\Sigma| \cdot \log|\Xi|)$. The previous best known bound was $\tilde{O}((n^2/\epsilon) \cdot |\Sigma|^2 \cdot |\Xi|)$.

We also present an alternative test for the boolean range $\Xi = \{0, 1\}$ whose complexity is independent of alphabet size $|\Sigma|$. This test has query complexity $O((n/\epsilon)\log^2(n/\epsilon))$ and time complexity $O((n/\epsilon)\log^3(n/\epsilon))$.

*Lab for Computer Science, MIT, 545 Technology Sq. Cambridge, MA 02139. email: yevgen@theory.lcs.mit.edu.

†Dept. of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, ISRAEL. e-mail: oded@wisdom.weizmann.ac.il.

‡Lab for Computer Science, MIT, 545 Technology Sq. Cambridge, MA 02139. email: e_lehman@theory.lcs.mit.edu.

§Lab for Computer Science, MIT, 545 Technology Sq. Cambridge, MA 02139. email: sofya@theory.lcs.mit.edu.

¶Dept. of EE – Systems, Tel Aviv University, Ramat Aviv, ISRAEL. e-mail: danar@eng.tau.ac.il.

‖DIMACS Center, Rutgers University, Piscataway, NJ 08854. email: salex@av.rutgers.edu.

1 Introduction

Property Testing (cf., [13, 9]) is a general formulation of computational tasks in which one is to determine whether a given object has a predetermined property or is "far" from any object having the property. Thus, property testing captures a natural notion of approximation, where the measure approximated is the object's "distance" to having the property. Typically one aims at performing this task within complexity smaller than the size of the object, while employing a randomized algorithm and given oracle access to a natural encoding of the object (as a function). Thus, we are talking of determining with high probability whether a function, to which we have oracle access, belongs to some class or is "far" from this class (i.e., one needs to modify the function value at many places so to obtain a function in the class).

Much work in this area was devoted to testing algebraic properties of functions such as linearity (e.g., [5, 1, 4, 3]) and low-degree properties (e.g., [5, 7, 13, 12, 2]). Recently, some attention was given to testing combinatorial properties of functions; firstly, for functions representing graphs [9, 10, 11], and more recently for functions per se [6, 8]. The most natural combinatorial property of functions is monotonicity, and indeed [8] focuses on testing monotonicity. The basic problem studied there is the following. Given a distance parameter ϵ and oracle access to a function $f : \{0,1\}^n \mapsto \{0,1\}$, determine whether f is monotone or is "ϵ-far" from being monotone. Monotonicity is defined in the natural manner: One considers the standard partial order $\prec$ on binary strings (i.e., $x_1x_2\cdots x_n \prec y_1y_2\cdots y_n$ iff $x_i \leq y_i$ for every i and $x_i < y_i$ for some i), and f is said to be monotone if $f(x) \leq f(y)$ for every $x \prec y$. The definition extends naturally to functions defined on the standard partial order of strings over an arbitrary alphabet, Σ, and having an arbitrary range Ξ. That is,

Definition 1 (monotone functions and testing): *Let Σ and Ξ be sets with total order $\leq_\Sigma$ and $\leq_\Xi$, respectively. We consider the partial order, $\prec$, defined on equal-length strings over Σ by $x_1x_2\cdots x_n \prec y_1y_2\cdots y_n$ iff $x_i \leq_\Sigma y_i$ for every i and $x_i \neq y_i$ for some i.*

- *A function $f : \Sigma^n \mapsto \Xi$ is* monotone *if $f(x) \leq_\Xi f(y)$ holds for every $x \prec y$.*
- *A* relative distance of $f : \Sigma^n \mapsto \Xi$ from the class of monotone functions, *$\epsilon_{\mathrm{M}}(f)$, is the minimum over all monotone functions $g : \Sigma^n \mapsto \Xi$ of $\mathrm{dist}(f,g) \stackrel{\mathrm{def}}{=} |\{x \in \Sigma^n : f(x) \neq g(x)\}| \,/\, |\Sigma|^n$.*
- *A function $f : \Sigma^n \mapsto \Xi$ is ϵ-*far from monotone *if $\epsilon_{\mathrm{M}}(f) \geq \epsilon$.*
- *A probabilistic oracle machine M is said to be a* tester of monotonicity *if*

$$\Pr[M^f(\epsilon, n) = 1] \geq \frac{2}{3} \qquad \text{for any monotone function } f, \tag{1}$$

$$\Pr[M^f(\epsilon, n) = 0] \geq \frac{2}{3} \qquad \text{for } f \text{ which is } \epsilon\text{-far from monotone.} \tag{2}$$

Note that all notions are defined w.r.t. Σ and Ξ, and so at times we prefer terms which explicitly mention this dependence.

The main result of [8] is a tester of monotonicity for the case $\Sigma = \Xi = \{0,1\}$ having query and time complexities of the form $\mathrm{poly}(n)/\epsilon$. Specifically, the analysis of the query complexity in [8] yields a bound of $\tilde{O}(n^2/\epsilon)$, and it was also shown that $\Omega(n/\epsilon)$ is a lower bound on the query complexity of their algorithm. For general Σ and Ξ, the bounds obtained in [8] were proportional to $|\Sigma|^2 \cdot |\Xi|$. Here we improve both the algorithm and the analysis in [8] to obtain the following.

Theorem 1 (main result): *There exists a tester of monotonicity with query and time complexity*

$$q(\epsilon, n) \stackrel{\text{def}}{=} O\left(\frac{n \cdot (\log|\Sigma|) \cdot (\log|\Xi|)}{\epsilon}\right).$$

The tester works by selecting independently $q(\epsilon, n)/2$ pairs of n-long strings over Σ, and comparing the two f-values obtained for the elements of each pair.[1]

Thus, the global feature of being monotone or far from it, is determined by a sequence of many independent random local checks. Each local check consists of selecting a pair, (x, y), so that (w.l.o.g) $x \prec y$, according to some fixed distribution and checking whether $f(x) \leq_\Xi f(y)$. If we ever find a pair for which this does not hold (i.e., local violation of monotonicity), then we reject. Otherwise we accept. Thus, we never reject a monotone function, and the challenge is to analyze the dependence of rejection probability on the distance of the given function from being monotone.

The only thing left unspecified in the above description of the testing algorithm is the distribution by which the pairs are selected. In case $\Sigma = \{0, 1\}$ there seems to be a very natural choice. Uniformly select $i \in [n] \stackrel{\text{def}}{=} \{1, ..., n\}$, independently and uniformly select $z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n \in \{0, 1\}$, and set $x = z_1 \cdots z_{i-1} 0 z_{i+1} \cdots z_n$ and $y = z_1 \cdots z_{i-1} 1 z_{i+1} \cdots z_n$. Our improvement over [8], in this case (where $\Sigma = \Xi = \{0, 1\}$), comes from a better (and in fact tight for $\Xi = \{0, 1\}$) analysis of this test: Let $\delta_{\mathrm{M}}(f)$ denote the fraction of pairs (x, y) as above for which $f(x) > f(y)$. We show that $\delta_{\mathrm{M}}(f) \geq \epsilon_{\mathrm{M}}(f)/n$ improving on the bound $\delta_{\mathrm{M}}(f) \geq \epsilon_{\mathrm{M}}(f)/n^2 \log(1/\epsilon_{\mathrm{M}}(f))$ in [8] (whereas by [8] there exist functions f for which $\delta_{\mathrm{M}}(f) = 2\epsilon_{\mathrm{M}}(f)/n$).

In case of non-binary $\Sigma = \{1, \ldots, d\}$ there seem to be several natural possibilities: Even if we restrict ourselves (as above) to select only pairs of strings which differ on a single coordinate i, there is still the question of how to select the corresponding pair of symbols. We study several natural possibilities of randomly selecting pairs $(k, \ell) \in \Sigma \times \Sigma$.

1. Distribution p_0: Select uniformly a pair $(k, k+1)$ with $k \in \{1, \ldots, d-1\}$.
2. Distribution p_1: Select uniformly a pair (k, ℓ) from a particular subset of size $O(d \log d)$ of the set of all pairs.
3. Distribution p_2: Select uniformly a pair (k, ℓ) with $k < \ell$.

A key result of this work is the reduction of the analysis of testing algorithms as above for any n and Σ, and for $\Xi = \{0, 1\}$, to their behavior in the special case of $n = 1$ (where we simply select pairs (k, ℓ) according to one of the above distributions and check the order between $f(k)$ and $f(\ell)$). Using this reduction we derive the following theorem.

Theorem 2 (Monotonicity Testing of Boolean functions): *There exist efficiently samplable distributions on pairs $(x, y) \in \Sigma^n \times \Sigma^n$ with $x \prec y$ so that for every function $f : \Sigma^n \mapsto \{0, 1\}$ the following holds:*

1. *If (x, y) is drawn according to one distribution (derived from p_1) then*

$$\Pr[f(x) > f(y)] \;=\; \Omega\left(\frac{\epsilon_{\mathrm{M}}(f)}{n \cdot (\log|\Sigma|)}\right).$$

[1] Since the algorithm is comparison-based, its complexity depends only on the size of the image of the function. Thus, one may replace Ξ in the above bound by $\Xi_f = \{f(x) : x \in \Sigma^n\}$. In particular, $\log|\Xi_f| \leq n \cdot \log|\Sigma|$, so our bound is never worse than $O(n^2 \cdot (\log|\Sigma|)^2/\epsilon)$.

2. If (x, y) is drawn according to another distribution (derived from p_2) then

$$\Pr[f(x) > f(y)] \;=\; \Omega\left(\frac{\epsilon_{\mathrm{M}}(f)^2}{n^2}\right).$$

We note that the first item of the theorem can also be derived by applying our reduction and using an alternative distribution on pairs in Σ^2 which was previously suggested in [6], and analyzed for the case $n = 1$. The second item leads to an algorithm having complexity $O(n^2/\epsilon^2)$.

The reader may be tempted to say that since our algorithm is comparison-based, the analysis should also hold for non-boolean functions. However, this is false. For example, by Item (2) above, boolean functions over Σ may be tested for monotonicity within complexity independent of $|\Sigma|$. In contrast, a lower bound in [6] asserts that arbitrary functions over Σ (e.g., with $\Xi = \Sigma$) cannot be tested for monotonicity within complexity independent of $|\Sigma|$ (but rather require complexity $\Omega(\log|\Sigma|)$ for some fixed distance parameter $\epsilon > 0$). Thus, a natural question arises: *Under what conditions and at what cost can results regarding testing of monotonicity of boolean functions be transformed to results for testing monotonicity of arbitrary functions*? Our most general result is the following.

Theorem 3 (Monotonicity Testing – Range Reduction): *Consider the task of testing monotonicity of functions defined over any partially ordered set S (with p.o. $\prec_S$). Suppose that for some distribution on pairs $(x, y) \in S \times S$ with $x \prec_S y$ and for every function $f : S \mapsto \{0, 1\}$*

$$\Pr[f(x) > f(y)] \;\geq\; \frac{\epsilon_{\mathrm{M}}(f)}{C}\;,$$

where C depends on S only. Then, for every Ξ and every function $f : S \mapsto \Xi$ for pairs selected according to the same distribution

$$\Pr[f(x) > f(y)] \;\geq\; \frac{\epsilon_{\mathrm{M}}(f)}{C \cdot \log_2|\Xi|}\;.$$

Theorem 1 follows by combining Part 1 of Theorem 2 and Theorem 3 with $C = O(n \cdot \log|\Sigma|)$.

An Alternative Algorithm for Boolean Functions. This algorithm, suggested by Noga Alon, works by picking $i \in \{1, \ldots, n\}$ uniformly, and then querying f on several strings that differ only on the i^{th} coordinate. The analysis of this algorithm also reduces to the $n = 1$ case, proving the following theorem.

Theorem 4 (Alternative tester for Boolean functions): *There exists a tester of monotonicity for Boolean functions with query complexity*

$$O\left(\frac{n}{\epsilon}\log^2\left(\frac{n}{\epsilon}\right)\right)$$

and time complexity

$$O\left(\frac{n}{\epsilon}\log^3\left(\frac{n}{\epsilon}\right)\right).$$

Organization:

We start with some preliminaries in Section 2. In Section 3 we show how the analysis of our algorithm in the boolean-range case for arbitrary n and Σ, reduces to the case $n = 1$. The algorithms for the case $n = 1$ (each corresponding to a different distribution on pairs in $\Sigma \times \Sigma$), are provided in Subsection 3.3, and the proof of Theorem 2, in Subsection 3.4. In Section 4 we prove a general reduction from an arbitrary range to the boolean range, and derive Theorem 3. Finally, in Section 5 we present a different kind of algorithm and prove Theorem 4.

2 Preliminaries

Let Σ and Ξ be sets with total order $\leq_\Sigma$ and $\leq_\Xi$, respectively. We consider the partial order, $\prec$, defined on equal-length strings over Σ as in the introduction, and shorthand $\leq_\Xi$ by $\leq$.

For any pair of functions $f, g : \Sigma^n \mapsto \Xi$, we define the *distance* between f and g, denoted $\mathrm{dist}(f, g)$, to be the fraction of instances $x \in \Sigma^n$ on which $f(x) \neq g(x)$. As in the introduction, we let $\epsilon_{\mathrm{M}}(f)$ denote the minimum distance between f and any monotone function $g : \Sigma^n \mapsto \Xi$. Let us formally define an algorithmic schema playing a dominant role [2] in this paper. The schema uses an arbitrary probability distribution $p : \Sigma \times \Sigma \mapsto [0, 1]$. Without loss of generality, we assume that the support of p is restricted to pairs (k, ℓ) with $k < \ell$. The function t referred to below, depends on p.

ALGORITHMIC SCHEMA: Given parameters ϵ, n, Σ, Ξ, and oracle access to an arbitrary function $f : \Sigma^n \mapsto \Xi$, repeat the following steps up to $t(\epsilon, n, |\Sigma|, |\Xi|)$ times:

1. Uniformly select dimension $i \in [n]$, prefix $\alpha \in \Sigma^{i-1}$, and suffix $\beta \in \Sigma^{n-i}$.
2. Select (k, ℓ) according to p. Let $x = \alpha\, k\, \beta$, $y = \alpha\, \ell\, \beta$.
3. If $f(x) > f(y)$ (i.e., (x, y) witnesses that f is not monotone), then reject.

If all iterations were completed without rejecting then accept.

We focus on the analysis of a single iteration of the above test. Such an iteration is fully specified by the distribution, denoted $D_p^n : \Sigma^n \times \Sigma^n \mapsto [0, 1]$, by which pairs (x, y) are selected. That is, $D_p^n(x, y) = \frac{p(k,\ell)}{n \cdot |\Sigma|^{n-1}}$ if $x = \alpha\, k\, \beta$ and $y = \alpha\, \ell\, \beta$, for some α, β, and $D_p^n(x, y) = 0$ otherwise. Observe that $D_p^n(x, y) > 0$ only if $x \prec y$. Let $\textsc{Detect}(f, D_p^n)$ be the probability that a pair (x, y) selected according to D_p^n witnesses that f is not monotone; that is,

$$\textsc{Detect}(f, D_p^n) \stackrel{\mathrm{def}}{=} \Pr_{(x,y)\sim D_p^n}[f(x) > f(y)] \tag{3}$$

(where the above definition can of course be applied to any distribution D on pairs $x \prec y$). Our goal is to find distributions D_p^n (determined by the distributions p) for which $\textsc{Detect}(f, D_p^n)$ is "well" lower-bounded as a function of $\epsilon_{\mathrm{M}}(f)$. If D_p^n is such that $\textsc{Detect}(f, D_p^n) \geq \delta(\epsilon, n, |\Sigma|, |\Xi|)$ *for any* $f : \Sigma^n \mapsto \Xi$ with $\epsilon_{\mathrm{M}}(f) \geq \epsilon$, then setting $t(\epsilon, n, |\Sigma|, |\Xi|) = \Theta(1/\delta(\epsilon, n, |\Sigma|, |\Xi|))$ yields a tester for monotonicity.

THE PARTIAL ORDER GRAPH: It will be convenient to view the partial order over Σ^n as a directed (acyclic) graph, denoted G_Σ^n. The vertices of G_Σ^n are the strings in Σ^n and directed edges correspond

[2] This schema describes a general comparison-based algorithm in which pairs of points in Σ^n are chosen independently. In section 5 we deal with a different type of algorithms and another algorithmic schema is defined.

to comparable pairs (i.e. (x, y) is an edge iff $x \prec y$). An edge (x, y) is said to be violated by f if $f(x) > f(y)$. We denote by $\textsc{Viol}(f)$ the set of violated edges of f. We remark that most of the definitions in this section naturally extend to any partially ordered set S in place of Σ^n.

3 Dimension Reduction for Boolean Functions

In this section we restrict our attention to boolean functions $f : \Sigma^n \mapsto \{0, 1\}$. Without loss of generality assume $\Sigma = \{1 \dots d\}$, so $|\Sigma| = d$. In what follows we reduce the analysis of the performance of our algorithmic schema for any n and Σ (and $\Xi = \{0, 1\}$) to its performance for the case $n = 1$ (the "line"). In Subsection 3.3 we describe and analyze several algorithms for the line. Recall that by our algorithmic schema any such algorithm is determined by a probability distribution p on pairs $(k, \ell) \in \Sigma \times \Sigma$. We conclude this section by combining the reduction with these algorithms to derive Theorem 2.

3.1 A Sorting Operator

We begin with a few definitions. For $i \in [n]$, any choice of $\alpha \in \Sigma^{i-1}$ and $\beta \in \Sigma^{n-i}$ defines a *line $f_{i,\alpha,\beta}$ of f in direction i* (or an *i-line of f*) by setting $f_{i,\alpha,\beta}(x) = f(\alpha, x, \beta)$, for $x \in \Sigma$. Thus $f_{i,\alpha,\beta} : \Sigma \mapsto \{0, 1\}$ is a one-dimensional boolean function. For any $i \in [n]$, we say that a function f is *monotone in dimension i*, if for every $\alpha \in \Sigma^{i-1}$, $\beta \in \Sigma^{n-i}$ the line $f_{i,\alpha,\beta}$ is a monotone one-dimensional function. For a set of indices $T \subseteq [n]$, we say that f is *monotone in dimensions T*, if for every $i \in T$, the function f is monotone in dimension i. In what follows we describe *sort* operators which can transform any boolean function over Σ^n into a monotone function (as we prove below).

Definition 2 *For every $i \in [n]$, the function $S_i[f] : \Sigma^n \mapsto \{0, 1\}$ is defined as follows: For every $\alpha \in \Sigma^{i-1}$ and every $\beta \in \Sigma^{n-i}$, we let $S_i[f](\alpha\,1\,\beta)$,..., $S_i[f](\alpha\,d\,\beta)$ be assigned the values of $f(\alpha\,1\,\beta), \dots, f(\alpha\,d\,\beta)$, in* sorted *order. In other words, S_i acts on f by sorting its i-lines.*

For any $i \in [n]$ let

$$\epsilon^i_{\mathrm{M}}(f) \stackrel{\mathrm{def}}{=} \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})),$$

where the expectation is taken uniformly over $\alpha \in \Sigma^i$, $\beta \in \Sigma^{n-i}$. Namely, $\epsilon^i_{\mathrm{M}}(f)$ is the average (relative) distance of an i-line of f from being monotone.

In the next lemma we show that by sorting in one dimension we do not increase the average distance from being monotone in any other dimension.

Lemma 5 *For every $f : \Sigma^n \mapsto \{0, 1\}$ and $j \in [n]$, we have:*

1. *If f is monotone in dimensions $T \subseteq [n]$ then $S_j[f]$ is monotone in dimensions $T \cup \{j\}$;*
2. *For every $i \in [n] \setminus \{j\}$,*
$$\epsilon^i_{\mathrm{M}}(S_j(f)) \leq \epsilon^i_{\mathrm{M}}(f).$$

Proof: The important observation is that in order to prove both items we may consider the function f restricted at all dimensions but the two in question. Furthermore, proofs of both items boil down to asserting claims about sorting zero-one matrices.

Item 1. Let i be some index in T, and assume without loss of generality that $i < j$. We fix any $\alpha \in \Sigma^{i-1}$, $\beta \in \Sigma^{j-i-1}$ and $\gamma \in \Sigma^{n-j}$, and consider the function $f' : \Sigma^2 \mapsto \{0, 1\}$ defined by

$f'(\sigma\tau) \stackrel{\text{def}}{=} f(\alpha\,\sigma\,\beta\,\tau\,\gamma)$. Clearly, f' is monotone in its first dimension (as f is monotone in dimension i). We need to show that so is $S_2[f']$ (as by definition of S_2 we have that $S_2[f']$ is monotone in dimension 2). Our claim thus amounts to saying that if one sorts the rows of a d-by-d zero-one matrix which is column-sorted then the columns remain sorted (the matrix we consider has its (σ,τ)-entry equal to $f'(\sigma\,\tau)$).

Let M denote a (d-by-d zero-one) matrix in which each column is sorted. We observe that the number of 1's in the rows of M is monotonically non-decreasing (as each column contributes a unit to the 1-count of row k only if it contributes a unit to the 1-count of row $k+1$). That is, if we let o_k denote the number of 1's in the k^{th} row then $o_k \leq o_{k+1}$ for $k = 1, ..., d-1$. Now suppose we sort each row of M resulting in a matrix M'. Then the k^{th} row of M' is $0^{d-o_k}1^{o_k}$, and it follows that the columns of M' remain sorted (as the $k+1^{\text{st}}$ row of M' is $0^{d-o_{k+1}}1^{o_{k+1}}$ and $o_k \leq o_{k+1}$).

Item 2. Fixing $i, j, \alpha, \beta, \gamma$ and defining f' as above, here we need to show that $\epsilon_{\text{M}}(S_2(f')) \leq \epsilon_{\text{M}}(f')$. This claim amounts to showing that if we sort the rows of a zero-one $d \times d$ matrix M, then the sum (or average) of the distances to monotone of the columns cannot increase.

The first observation is that it suffices to prove the above for a $d \times 2$ matrix. This is true because we can sort the rows of M by first sorting the d subrows of length two that correspond to the first two columns, and then sort the d subrows that correspond the columns 2 and 3, and so on. Thus we only need to show that in each such step the sum of the distances to monotone of the two columns in question does not increase.

Let the columns before sorting be A and B, and after sorting A' and B'. Let C be the monotone column (vector) closest to A, and let D be the monotone column closest to B. We will define monotone columns C' and D' such that $\text{dist}(A', C') + \text{dist}(B', D') \leq \text{dist}(A, C) + \text{dist}(B, D)$ so that $\epsilon_{\text{M}}(A') + \epsilon_{\text{M}}(B') \leq \epsilon_{\text{M}}(A) + \epsilon_{\text{M}}(B)$.

Let k be the first index on which C switches from 0 to 1, and let ℓ be the corresponding index in D.

Consider two cases. The first is $\ell \leq k$. Then we may set $C' = C$ and $D' = D$. To verify this observe that for any $j < \ell$, if $A[j] = 1$ and $B[j] = 0$, in which case j was counted in the distance between A and C (but not in the distance between B and D), then now $A'[j] = 0$ and $B'[j] = 1$, so that j is counted in the distance between B' and D (but not in the distance between A' and C). All other configurations (00, 01, and 11) don't change and so contribute as before. Similarly, for $j \geq k$, if $A[j] = 1$ and $B[j] = 0$, then we count this in the distance between B and D, and after sorting, in the distance between A' and C. Finally (for this case), a 10 in $\ell \leq j < k$ is counted in both distances before sorting, while after sorting there is no contribution.

The second case is $k < \ell$. Now we set $C' = D$ and $D' = C$ (so that the boundaries between 0's and 1's "switch sides"). Similarly to the above, for indices $j < k$ and $j \geq \ell$ the sum of the contributions remains the same. For $k \leq j < \ell$, we see the following. If we had a 00 then it contributed only to the distance between A and C, and now it contributes to the distance between B' and $D' = C$. A 11 contributed to $\text{dist}(B, D)$ and now to $\text{dist}(A', C' = D)$. A 01 contributed to both and now contributes to neither, and a 10 contributed to neither, and after being sorted to 01 also contributed to neither. ■

3.2 Dimension Reduction

With Lemma 5 at our disposal, we are ready to state and prove that the analysis of the algorithmic schema (for any n) reduces to its analysis for the special case $n = 1$. Let A denote one iteration of the algorithmic schema, p be any distribution on pairs $(k, \ell) \in \Sigma \times \Sigma$ such that $k < \ell$, and D_p^n

be the corresponding distribution induced on edges of G^n_Σ. The dimension reduction lemma upper bounds $\epsilon_{\mathrm{M}}(f)$ and lower bounds $\textsc{Detect}(f, D^n_p)$ by the corresponding quantities for $n = 1$.

Lemma 6 (Dimension Reduction for Boolean Range) *Let $f : \Sigma^n \mapsto \{0,1\}$. Then the set of lines $\left\{f_{i,\alpha,\beta} : \Sigma \mapsto \{0,1\} \ : \ i \in [n],\ \alpha \in \{0,1\}^{i-1},\ \beta \in \{0,1\}^{n-i}\right\}$ of f satisfies the following properties (all expectations below are taken uniformly over $i \in [n]$, $\alpha \in \{0,1\}^{i-1}$ and $\beta \in \{0,1\}^{n-i}$):*

1. $\textsc{Detect}(f, D^n_p) = \mathbf{E}_{i,\alpha,\beta}(\textsc{Detect}(f_{i,\alpha,\beta},\ p))$.
2. $\epsilon_{\mathrm{M}}(f) \leq 2n \cdot \mathbf{E}_{i,\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta}))$.

We prove the lemma momentarily, but first consider its implication on the relatively simple case of $\Sigma = \{0,1\}$. First observe that in this case there is only one possible distribution p – the one assigning all weight to the pair $(0,1)$. Also, for any $f' : \{0,1\} \mapsto \{0,1\}$, Algorithm A rejects with probability exactly $2\epsilon_{\mathrm{M}}(f')$. Thus, the lemma implies that in the binary (domain and range) case, for any $f : \{0,1\}^n \mapsto \{0,1\}$,

$$\begin{aligned}
\mathbf{E}_{i,\alpha,\beta}(\textsc{Detect}(f_{i,\alpha,\beta},\ p)) &= \mathbf{E}_{i,\alpha,\beta}(2 \cdot \epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})) \\
&= \frac{2}{n} \cdot \sum_i \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})) \\
&\geq \frac{1}{n} \cdot \epsilon_{\mathrm{M}}(f)
\end{aligned}$$

and we obtain a testing algorithm whose complexity is $O(n/\epsilon)$. Note that the algorithm is very simple – it uniformly picks an edge whose endpoints differ in exactly one coordinate. Let us now prove the lemma.

Proof: The first item of the lemma follows immediately from the definition of $\textsc{Detect}(f, D^n_p)$.

We go on to the second item. For $i = 1, \ldots, n+1$, we define $f_i \stackrel{\text{def}}{=} S_{i-1} \cdots S_1[f]$. Thus, $f_1 \equiv f$, and we have that f_{n+1} is monotone. It follows that

$$\epsilon_{\mathrm{M}}(f) \ \leq \ \mathrm{dist}(f, f_{n+1}) \ \leq \ \sum_{i=1}^{n} \mathrm{dist}(f_i, f_{i+1}). \tag{4}$$

Next, for $i = 1 \ldots n$, $\alpha \in \{0,1\}^{i-1}$ and $\beta \in \{0,1\}^{n-i}$, define the function $g_{i,\alpha,\beta} : \Sigma \mapsto \{0,1\}$, by $g_{i,\alpha,\beta}(x) = f_i(\alpha\, x\, \beta)$, for $x \in \Sigma$. Observe that the functions $\{g_{i,\alpha,\beta}\}$ are the set of lines of f_i in the i'th direction.

Throughout the proof, $\sum_{\alpha,\beta}$ refers to summing over all (α,β)'s in $\Sigma^{i-1} \times \Sigma^{n-i}$, and $\mathbf{E}_{\alpha,\beta}$ refers to expectation over uniformly distributed $(\alpha,\beta) \in \Sigma^{i-1} \times \Sigma^{n-i}$. We claim that

$$\mathrm{dist}(f_i, f_{i+1}) \ \leq \ 2 \cdot \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(g_{i,\alpha,\beta})). \tag{5}$$

This inequality is proven (below) by observing that f_{i+1} is obtained from f_i by sorting, separately, the elements in each $f_{i,\alpha,\beta}$. (The factor of 2 is due to the relationship between the distance of a vector to its sorted form and its distance to monotone.) We have,

$$\begin{aligned}
d^n \cdot \mathrm{dist}(f_i, f_{i+1}) &= \sum_{\alpha,\beta} |\{x \in \Sigma : f_i(\alpha\, x\, \beta) \neq f_{i+1}(\alpha\, x\, \beta)\}| \\
&= \sum_{\alpha,\beta} |\{x \in \Sigma : g_{i,\alpha,\beta}(x) \neq S_i[g_{i,\alpha,\beta}](x)\}| \\
&\leq \sum_{\alpha,\beta} 2d \cdot \epsilon_{\mathrm{M}}(g_{i,\alpha,\beta}) = 2d^n \cdot \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(g_{i,\alpha,\beta}))
\end{aligned}$$

where the inequality is justified as follows. Consider a vector $v \in \{0,1\}^d$, and let $S(v)$ denote its sorted version. Then $S(v) = 0^z 1^{d-z}$, where z denotes the number of zeros in v. Thus, for some $e \geq 0$, the vector v has e 1-entries within its z-prefix and e 0-entries in its $(d-z)$-suffix. So the number of locations on which v and $S(v)$ disagree is exactly $2e$. On the other hand, consider an arbitrary perfect matching of the e 1-entries in the prefix and the e 0-entries in the suffix. To make v monotone one must alter at least one entry in each matched pair; thus, $\epsilon_{\mathrm{M}}(v) \geq e/d$.

By the second part of Lemma 5, for any $i \in [n]$, $\alpha \in \Sigma^{i-1}$, $\beta \in \Sigma^{n-i}$ holds

$$\mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})) \geq \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(g_{i,\alpha,\beta})). \tag{6}$$

Combining this with Eq. (4) and (5), the second item of the lemma follows:

$$\epsilon_{\mathrm{M}}(f) \;\leq\; \sum_{i=1}^{n} \mathrm{dist}(f_i, f_{i+1}) \leq 2 \cdot \sum_{i=1}^{n} \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(g_{i,\alpha,\beta})) \leq 2 \cdot \sum_{i=1}^{n} \mathbf{E}_{\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})) = 2n \cdot \mathbf{E}_{i,\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})).$$

■

3.3 Testing Monotonicity on a Line (the $n = 1$ case)

In this section we design algorithms for the case $n = 1$, for any Σ and Ξ. In accordance with our algorithmic schema, the design of such algorithms amounts to the design of a probability distribution $p : \Sigma^2 \mapsto [0,1]$ (with support only on pairs (k,ℓ) with $k < \ell$).

Note that for $n = 1$, we have $D_p^n \equiv p$. We present three such distributions, denoted p_0, p_1, and p_2, and provide bounds on $\textsc{Detect}(f, p_j)$, for each j.

The following lemma[3], which relates $\epsilon_{\mathrm{M}}(f)$ to $\textsc{Viol}(f)$, will be used in our analysis of various algorithms. Recall that a *matching* of a graph is a collection of edges that share no common endpoint.

Lemma 7 *For any $f : \Sigma \mapsto \Xi$ the graph $G' = (\Sigma, \textsc{Viol}(f))$ has a matching of size $\epsilon_{\mathrm{M}}(f) \cdot |\Sigma|/2$.*

Proof: Recall that a *vertex cover* of a graph is a subset of vertices such that every edge of the graph has at least one of its endpoints in the subset. We claim that a minimum vertex cover of G' has size at least $\epsilon_{\mathrm{M}}(f) \cdot |\Sigma|$. The lemma directly follows as the size of a maximum matching is at least $1/2$ of the size of the minimum vertex cover. Let $U \subseteq \Sigma$ be any vertex cover of G'. We next show that by modifying the value of f only on points in U, we obtain a monotone function, implying that $|U| \geq \epsilon_{\mathrm{M}}(f) \cdot |\Sigma|$, as claimed.

Let $T = \Sigma \backslash U$. By definition of U, there are no violated edges between pairs of vertices in T. Consider the following iterative process, where in each step we modify the value of f on a single $y \in U$, remove y form U and add it to T. We maintain the property (which holds initially) that following each step there are no violated edges between vertices in T. The process ends when $U = \emptyset$ and $T = \Sigma$, so that the final function is monotone. To redefine the value of f on y, we consider the following two subsets of T: $T_1 = \{x \in T : \; (x,y) \in \textsc{Viol}(f)\}$ and $T_2 = \{z \in T : \; (y,z) \in \textsc{Viol}(f)\}$. By transitivity of the partial order, and the fact that there are no violated edges (x,z), for $x, z \in T$, at most one of these subsets is non-empty. If T_1 is non-empty then we let $f(y) = \max_{x \in T_1}\{f(x)\}$, and if T_2 is non-empty, then $f(y) = \min_{z \in T_2}\{f(y)\}$. In case both are empty (all violated edges incident to y have an end-point in U), the value of y may remain unchanged.

[3]While stated for a totally ordered set Σ, the result and the same proof hold for any partially ordered set S.

We note that the size of the minimum vertex cover actually equals $\epsilon_{\rm M}(f) \cdot |\Sigma|$. Consider any set U such that by modifying the value of f only on strings in U we can obtain a monotone function g. Then U must be a vertex cover of G', as otherwise there remain violated edges with respect to g. ■

DISTRIBUTION p_0: This distribution is uniform over pairs $(k, k+1)$. That is, $p_0(k, k+1) = 1/(d-1)$, for $k = 1, \ldots, d-1$.

Proposition 8 *For any Ξ and $f : \Sigma \mapsto \Xi$,* $\textsc{Detect}(f, p_0) \geq \frac{2}{d-1} \cdot \epsilon_{\rm M}(f)$.

The lower bound can be shown to be tight even for $\Xi = \{0, 1\}$ (by considering the function f defined by $f(x) = 1$ if $x < d/2$ and $f(x) = 0$ otherwise).

Proof: If $\epsilon_{\rm M}(f) > 0$, then there exists some $k \in \{1, \ldots, d-1\}$ so that $f(k) > f(k+1)$. If there are at least two such k's, then we reject with probability at least $2/(d-1) \geq 2\epsilon_{\rm M}(f)/(d-1)$ as $\epsilon_{\rm M}(f) \leq 1$. Otherwise, there is a unique k that causes us to reject. In this case $\epsilon_{\rm M}(f) \leq 1/2$ since we can change either all $f(i)$ to $f(k+1)$ for $i \leq k$, or all $f(i)$ to $f(k)$ for $i > k$ in order to make f monotone. Thus, we reject with probability $1/(d-1) \geq 2\epsilon_{\rm M}(f)/(d-1)$ in this case as well. ■

We see that the above test is too "short-sighted" since it only looks at the neighboring pairs of vertices. We now describe a distribution that spots the violated edges much better. As noted before, an alternative distribution which meets the same bound was previously suggested and analyzed in [6].

DISTRIBUTION p_1: This distribution is uniform on a set $P \subset \Sigma \times \Sigma$ which is defined as follows. The set P consists of pairs (k, ℓ), where $0 < \ell - k \leq 2^t$ and 2^t is the largest power of 2 which divides either k or ℓ. That is, let $\text{power}_2(i) \in \{0, 1..., \log_2 i\}$ denote the largest power of 2 which divides i. Then,

$$P \stackrel{\text{def}}{=} \{(k, \ell) \in \Sigma \times \Sigma : 0 < \ell - k \leq 2^{\max(\text{power}_2(k), \text{power}_2(\ell))}\} \tag{7}$$

and $p_1(k, \ell) = \frac{1}{|P|}$ for every $(k, \ell) \in P$, and is 0 otherwise.

Proposition 9 *For any Ξ and $f : \Sigma \mapsto \Xi$,* $\textsc{Detect}(f, p_1) \geq \frac{1}{O(\log d)} \cdot \epsilon_{\rm M}(f)$.

Proof: We first show that $|P| = O(d \log d)$. This can be shown by charging each pair $(k, \ell) \in P$ to the element divisible by the larger power of 2 (i.e., to k if $\text{power}_2(k) > \text{power}_2(\ell)$ and to ℓ otherwise), and noting that the charge incurred on each i is at most $2 \cdot 2^{\text{power}_2(i)}$. It follows that the total charge is at most $\sum_{i=1}^{d} 2^{\text{power}_2(i)+1} = \sum_{j=0}^{\log_2 d} \frac{d}{2^j} \cdot 2^{j+1} = O(d \log d)$.

Since p_1 is uniform over P, the value of $\textsc{Detect}(f, p_1)$ is the ratio between the number of violated edges of f in P and the size of P. Thus, it remains to show that the former is $\Omega(\epsilon_{\rm M}(f) \cdot d)$. In the following argument it will be convenient to view the indices $1, \ldots, d$ as vertices of a graph and the pairs $(k, \ell) \in P$ as directed edges. We refer to this graph as G_P, and note that it is a subgraph if G^1_Σ.

Claim 9.1: For every two vertices k and ℓ in G_P with $k < \ell$, there is a directed path of length at most 2 from k to ℓ in G_P.

Proof of Claim: Let $r = \lceil \log d \rceil$, and consider the binary strings of length r representing k and ℓ. Let $k = (x_{r-1}, \ldots, x_0)$ and $\ell = (y_{r-1}, \ldots, y_0)$. Let j be the highest index such that $x_j = 0$ and $y_j = 1$. Note that $x_i = y_i$ for $j < i < r$. We claim that the vertex $s = (x_{r-1}, \ldots, x_{t+1}, 1, 0, \ldots 0)$ is

on a path of length 2 from k to ℓ. This follows from the definition of P, since s is divided by 2^j, while both $s - k = 2^j - \sum_{i=0}^{j-1} x_i 2^i \leq 2^j$ and $\ell - s = \sum_{i=0}^{j-1} y_i 2^i < 2^j$. $\square$

We now apply Lemma 7 to obtain a matching M of size $m \geq (\epsilon_{\mathrm{M}}(f) \cdot d/2)$ consisting of violated edges of f. By the above claim, there is path of length at most 2 in G_P between every matched pair. Each edge e of G_P belongs to at most 2 such paths: on at most one path it is the first edge, and on at most one it is the second edge (or otherwise M is not a matching). Since for every $(x, y) \in M$ we have $f(x) > f(y)$ (while $x \prec y$), the length-2 path between x and y must contain a violated edge. Thus, we obtain at least $m/2 \geq (\epsilon_{\mathrm{M}}(f) \cdot d/4)$ violated edges in G_P, and the proposition follows. ■

On the Optimality of Distribution p_1. We show that the result of Proposition 9 is optimal (up to a constant factor), even for $\Xi = \{0, 1\}$. The following argument is due to Michael Krivelevich.

Proposition 10 *For any distribution $p : \Sigma \times \Sigma \mapsto [0, 1]$, with support only on pairs (k, ℓ) such that $k < \ell$, there exists a non-monotone $f : \Sigma \mapsto \{0, 1\}$ so that*

$$\textsc{Detect}(f, p) \;\leq\; \frac{2}{\log_2 d} \cdot \epsilon_{\mathrm{M}}(f)$$

Proof: Let p be a distribution on pairs as above. We define

$$\rho \stackrel{\mathrm{def}}{=} \max_{f:\Sigma\mapsto\{0,1\} \text{ s.t. } \epsilon_{\mathrm{M}}(f)>0} \left\{ \frac{\textsc{Detect}(f, p)}{\epsilon_{\mathrm{M}}(f)} \right\}$$

Our aim is to show that $\rho \leq 2/\log_2 d$. The key observation is that for any consecutive $2a$ indices, p has to assign a probability mass of at least $\rho \cdot a/d$ to pairs (k, ℓ) where k is among the lowest a indices and ℓ is among the higher a such indices. This observation is proven as follows. Let L, H be the low and high parts of the interval in question; that is, $L = \{s + 1, ..., s + a\}$ and $H = \{s+a+1, ..., s+2s\}$, for some $s \in \{0, ..., d-2a\}$. Consider the function f defined by $f(i) = 1$ if $i \in L \cup \{s + 2a + 1, ..., d\}$ and $f(i) = 0$ otherwise. Then $\epsilon_{\mathrm{M}}(f) = a/d$. On the other hand, the only pairs (k, ℓ) with $f(k) > f(\ell)$, are those satisfying $k \in L$ and $\ell \in H$. Thus, by definition of ρ, it must hold that $\rho \leq \Pr_{(k,\ell)\sim p}[k \in L \;\&\; \ell \in H]/(a/d)$, and the observation follows.

The rest of the argument is quite straightforward: Consider $\log_2 d$ partitions of the interval $[1, d]$, so that the i^{th} partition is into consecutive segments of length 2^i. For each segment in the i^{th} partition, probability p assign a probability mass of at least $2^{i-1}\rho/d$ to pairs where one element is in the low part of the segment and the other element is in the high part. Since these segments are disjoint and their number is $d/2^i$, it follows that p assigns a probability mass of at least $\rho/2$ to pairs among halves of segments in the i^{th} partition. These pairs are disjoint from pairs considered in the other partitions and so we conclude that $(\log_2 d) \cdot \frac{\rho}{2} \leq 1$. The proposition follows. ■

We now describe a distribution that works well for the boolean range only.

Distribution p_2: This distribution is uniform over all pairs (k, ℓ) such that $k < \ell$. That is, $p_2(k, \ell) = 2/((d-1)d)$ for $1 \leq k < \ell \leq d$.

Proposition 11 *For any $f : \Sigma \mapsto \{0, 1\}$, $\textsc{Detect}(f, p_2) \geq \epsilon_{\mathrm{M}}(f)^2/2$.*

A slightly more careful analysis (which we omit) can extend the bound to $\epsilon_{\mathrm{M}}(f)^2$, which is tight. For any integer $e < d/2$, consider the function $f(x) = 0$ if $x \in \{2, 4, 6 \ldots 2e\}$ and $f(x) = 1$ otherwise.

Then $\epsilon_{\mathrm{M}}(f) = e/d$ and $|\textsc{Viol}(f)| = 1 + \ldots + e \approx e^2/2$. Thus, $\textsc{Detect}(f, p_2) \approx (e^2/2)/(d^2/2) = (e/d)^2 = \epsilon_{\mathrm{M}}(f)^2$.

Proof: Let z be the number of zeroes in f and let $2e$ be the number of mismatches between f and its sorted form. Then $\epsilon_{\mathrm{M}}(f) \leq 2e/d$ as by swapping the $2e$ mismatches we make f monotone. On the other hand, considering the e 1-entries in the z-prefix of f and the e 0-entries in the $(d-z)$-suffix, we lower bound the rejection probability by $e^2/((d-1)d/2) > 2(e/d)^2$. Combining the two, we conclude that $\textsc{Detect}(f, p_2) \geq 2 \cdot (\epsilon_{\mathrm{M}}(f)/2)^2$. ■

We remark that the restriction to boolean range in Proposition 11 is important. For any integer $e \leq d/2$, define $f : \Sigma \mapsto \Sigma$ by $f(2i) = 2i - 1$, $f(2i-1) = 2i$ for $i = 1 \ldots e$, and $f(i) = i$ for $i > 2e$. Clearly, $\epsilon_{\mathrm{M}}(f) = e/d$, while f has only e violated edges: $(2i-1, 2i)$, $i = 1 \ldots e$. Thus, $\textsc{Detect}(f, p_2) = e/(d(d-1)/2) = 2\epsilon_{\mathrm{M}}(f)/(d-1)$, which is much less than $\epsilon_{\mathrm{M}}(f)^2$ if e is large.

3.4 Proof of Theorem 2

In this subsection we combine Lemma 6 with the results for the case $n = 1$ provided in Subsection 3.3, and derive Theorem 2

Combining Lemma 6 and Proposition 9 (applied only to $\Xi = \{0, 1\}$), we have

$$\begin{array}{rcll} \textsc{Detect}(f, D^n_{p_1})) & \geq & \mathbf{E}_{i,\alpha,\beta}(\textsc{Detect}(f_{i,\alpha,\beta},\ p_1)) & \text{[By Part 2 of the lemma]} \\ & \geq & \mathbf{E}_{i,\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})/O(\log d)) & \text{[By the proposition]} \\ & \geq & \frac{\epsilon_{\mathrm{M}}(f)}{2n \cdot O(\log d)} = \Omega(\frac{\epsilon_{\mathrm{M}}(f)}{n \log d}) & \text{[By Part 1 of the lemma]} \end{array}$$

which establishes the the first item in the theorem.

Combining Lemma 6 and Proposition 11, we have

$$\begin{array}{rcll} \textsc{Detect}(f, D^n_{p_2})) & \geq & \mathbf{E}_{i,\alpha,\beta}(\textsc{Detect}(f_{i,\alpha,\beta},\ p_2)) & \text{[By Part 2 of the lemma]} \\ & \geq & \mathbf{E}_{i,\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})^2/2) & \text{[By the proposition]} \\ & \geq & [\mathbf{E}_{i,\alpha,\beta}(\epsilon_{\mathrm{M}}(f_{i,\alpha,\beta}))]^2/2 & \text{[as } \mathbf{E}(X^2) \geq [\mathbf{E}(X)]^2\text{]} \\ & \geq & (\epsilon_{\mathrm{M}}(f)/2n)^2/2 = \Omega(\epsilon_{\mathrm{M}}(f)^2/n^2) & \text{[By Part 1 of the lemma]} \end{array}$$

which establishes the second item in the theorem.

4 Testing Monotonicity over General Ranges

We now reduce the problem of testing arbitrary-range functions to the simpler problem of testing boolean functions, which was considered in the preceding section. This reduction works not only for functions with domain Σ^n, but more generally when the domain is any partially ordered set S. The reduction is characterized by Theorem 3, which states that a certain type of monotonicity test for functions of the form $f : S \mapsto \{0, 1\}$ also works well for functions of the form $f : S \mapsto \Xi$. Here Ξ is a finite totally ordered set of size r, which we can regard as the integers in the interval $[0, r-1]$. Furthermore, for simplicity, we assume that $r = 2^s$ for some integer s. All references to "edges" are references to edges of the partial order graph, whose vertices are strings in the domain S and directed edges correspond to ordered comparable pairs (i.e. (x, y) is an edge iff $x \prec y$).

To ensure that a function far from monotone can be readily detected by our test, we lower bound $\textsc{Detect}(f, D)$ in terms of $\epsilon_{\mathrm{M}}(f)$. Equivalently, we are looking for a good upper bound on $\epsilon_{\mathrm{M}}(f)$ in terms of $\textsc{Detect}(f, D)$. We reduce the task of obtaining an upper bound for functions with an arbitrary range to that of obtaining such an upper bound for functions with binary range.

The general idea of the reduction is to incrementally transform a function f into a monotone function, while ensuring that for each repaired violated edge, the value of the function is changed at only a few points. This transformation allows us to find a monotone function close to f and to upper bound $\epsilon_{\mathrm{M}}(f)$ by the distance from f to that function. The transformation produces the following chain of functions: $f \mapsto f_1 \mapsto f_2 \mapsto f_3$, where f_3 is monotone. The distance between any two consecutive functions in the chain is equal to the distance to monotone of some auxiliary function with a smaller range. Thus, we obtain an upper bound on $\epsilon_{\mathrm{M}}(f)$ in terms of the distance to monotone of smaller-range functions. In addition, edges violated by the auxiliary functions are also violated by f, and we can obtain a lower bound on $\textsc{Detect}(f, D)$ in terms of the corresponding probability for the smaller-range auxiliary functions. Using the inductive assumption for smaller-range functions and the two claims above, we finally obtain the needed upper bound on $\epsilon_{\mathrm{M}}(f)$ in terms of $\textsc{Detect}(f, D)$.

Subsection 4.1 describes and analyzes operators SQUASH, MON, and CLEAR later used to define functions f_1, f_2, and f_3 described above. Subsection 4.2 proves the range reduction lemma which upper bounds $\epsilon_{\mathrm{M}}(f)$ and lower bounds $\textsc{Detect}(f, D)$ by the corresponding quantities for smaller range functions. This section is concluded by the proof of Theorem 3 in Subsection 4.3.

4.1 Operators SQUASH, MON, and CLEAR

First, we introduce operators, later used for obtaining functions f_1, f_2, and f_3 related to f.

Definition 3 *The operators* SQUASH, MON, *and* CLEAR *each map a function $f : S \mapsto [0, r-1]$ to a related function with the same domain and the same or smaller range. In particular,* MON$[f]$ *is some arbitrary monotone function at distance $\epsilon_{\mathrm{M}}(f)$ from the function f. The operators* SQUASH *and* CLEAR *are defined below; in these definitions a and b are elements of $[0, r-1]$ and $a < b$.*

$$\textsc{squash}[f,a,b](x) = \begin{cases} a & \text{if } f(x) \le a \\ b & \text{if } f(x) \ge b \\ f(x) & \text{otherwise} \end{cases}$$

$$\textsc{clear}[f,a,b](x) = \begin{cases} \textsc{mon}[\textsc{squash}[f,a,b]](x) & \\ \qquad \text{if } \textsc{mon}[\textsc{squash}[f,a,b]](x) \neq \textsc{squash}[f,a,b](x) & \\ f(x) & \text{otherwise} \end{cases}$$

SQUASH operator simply "squashes" the range of f to $[a, b]$. Notice that if an edge is not violated by f, it is not violated by $\textsc{squash}[f, a, b]$.

Claim 12 *For all $f : S \mapsto [0, r-1]$ and all $a, b \in [0, r-1]$ such that $a < b$, the operator* SQUASH *does not introduce any new violated edges, i.e.* $\textsc{Viol}(\textsc{squash}[f, a, b]) \subseteq \textsc{Viol}(f)$. ■

CLEAR operator first "squashes" the range to $[a, b]$, then alters the resulting smaller-range function at some points to obtain the closest monotone function, and finally "unsquashes" the function at unaltered points to the original values. This leads to the following simple claim:

Claim 13 *For all $f : S \mapsto [0, r-1]$ and all $a, b \in [0, r-1]$ such that $a < b$,*

$$\mathrm{dist}(f, \textsc{clear}[f, a, b]) = \epsilon_{\mathrm{M}}(\textsc{squash}[f, a, b]).$$

Proof: By definitions of the CLEAR and MON operators:

$$\mathrm{dist}(f, \textsc{clear}[f, a, b]) = \mathrm{dist}(\textsc{mon}[\textsc{squash}[f, a, b]], \textsc{squash}[f, a, b]) = \epsilon_{\mathrm{M}}(\textsc{squash}[f, a, b]).$$

■

Define the *interval of a violated edge* (x, y) *with respect to function* f to be the interval $[f(y), f(x)]$ (since the edge is violated by f, $f(x) > f(y)$). We say that two intervals *cross* if they intersect in more than one point. Intuitively, if we consider $f(x) - f(y)$ as a measure of how violated an edge (x, y) is, then we can say that CLEAR$[f, a, b]$ partially repairs violated edges of f whose intervals cross $[a, b]$ without worsening other violated edges. The following lemma formalizes important properties of CLEAR.

Lemma 14 *The function* CLEAR$[f, a, b]$ *has the following properties for all* $f : S \mapsto [0, r-1]$ *and all* $a, b \in [0, r-1]$ *such that* $a < b$*:*

1. VIOL(CLEAR$[f, a, b]$) $\subseteq$ VIOL(f)*, i.e.* CLEAR *does not introduce any new violated edges.*
2. CLEAR$[f, a, b]$ *has no violated edges whose intervals cross* $[a, b]$*.*
3. *The interval of a violated edge with respect to* CLEAR$[f, a, b]$ *is contained in the interval of this edge with respect to* f*.*

Proof: For brevity, define $g =$ MON[SQUASH$[f, a, b]$] and $h =$ CLEAR$[f, a, b]$. Let (x, y) be an edge violated by h; that is, $h(x) > h(y)$. By its definition, g is monotone and takes values in $[a, b]$. Also notice that $h(x) = f(x)$ if $h(x) \notin [a, b]$, and $h(x) = g(x)$ if $h(x) \in [a, b]$. We consider four cases where each of $h(x)$ and $h(y)$ is either inside or outside the interval $[a, b]$.

- Case 1: $h(x), h(y) \in [a, b]$. This case cannot occur: (x, y) cannot be violated by h because $h(x) = g(x), h(y) = g(y)$ and g is monotone.
- Case 2: $h(x), h(y) \notin [a, b]$. Since f and h agree on both x and y, it follows that (x, y) is violated by f and $[h(y), h(x)] = [f(y), f(x)]$. This proves parts 1 and 3. To show that $[h(y), h(x)]$ does not cross $[a, b]$, it remains to prove that the case when $h(x) > b$ and $h(y) < a$ cannot happen. But in such a case we must have $g(x) = b$ and $g(y) = a$ and that contradicts the monotonicity of g.
- Case 3: $h(x) \notin [a, b]$, $h(y) \in [a, b]$. Since (x, y) is violated, $h(x) > b$. Consequently, $f(x) = h(x) > b$ and, thus, $g(x) = b$. Since g is monotone, $g(y) \geq g(x) = b$, and hence $h(y) = g(y) = b$. This proves that $[h(y), h(x)]$ intersects $[a, b]$ in at most one point (b), establishing part 2. If $f(y) = h(y) = b$, then f and h agree on both x and y, and parts 1 and (3) follow. If not, then $b = g(y) \neq$ SQUASH$[f, a, b](y)$. Thus, SQUASH$[f, a, b](y) < b$, and hence $f(y) < b = h(y)$. Since $f(x) = h(x) > b$, parts 1 and 3 follow.
- Case 4: $h(x) \in [a, b], h(y) \notin [a, b]$. This case is symmetrical to Case 3.

■

4.2 Range Reduction

We are now ready to define functions in the chain $f \mapsto f_1 \mapsto f_2 \mapsto f_3$, as well as auxiliary smaller-range functions f'_1, f'_2, and f'_3. Lemma 15 defines these functions and summarizes their properties. The transition from f to f_1 transforms violated edges with one endpoint in the lower half of the range and the other endpoint in the upper half into edges with both endpoints in the same half of the range. Then we repair violated edges with both endpoints in the lower half of the range to obtain f_2 and finally, upper half of the range to obtain f_3.

Lemma 15 (Range Reduction) *Given $f : S \mapsto [0, r-1]$, define*

$$f' = \text{SQUASH}[f, \tfrac{r}{2}-1, \tfrac{r}{2}], \qquad f'_1 = \text{SQUASH}[f_1, 0, \tfrac{r}{2}-1], \qquad f'_2 = \text{SQUASH}[f_2, \tfrac{r}{2}, r-1],$$
$$f_1 = \text{CLEAR}[f, \tfrac{r}{2}-1, \tfrac{r}{2}], \qquad f_2 = \text{CLEAR}[f_1, 0, \tfrac{r}{2}-1], \qquad f_3 = \text{CLEAR}[f_2, \tfrac{r}{2}, r-1].$$

These functions have the following properties, for any probability distribution D.

1. $\text{DETECT}(f, D) \geq \text{DETECT}(f', D)$
2. $\text{DETECT}(f, D) \geq \text{DETECT}(f'_1, D) + \text{DETECT}(f'_2, D)$
3. $\epsilon_{\text{M}}(f) \leq \epsilon_{\text{M}}(f') + \epsilon_{\text{M}}(f'_1) + \epsilon_{\text{M}}(f'_2)$

Proof: All references to "parts" are references to parts of Lemma 14.

(*1*) The SQUASH operator never adds new violated edges by Claim 12. Therefore, $\text{VIOL}(f') \subseteq \text{VIOL}(f)$, and the claim follows.

(*2*) It is enough to show that $\text{VIOL}(f'_1)$ and $\text{VIOL}(f'_2)$ are *disjoint* subsets of $\text{VIOL}(f)$. First, note that $\text{VIOL}(f'_1)$ and $\text{VIOL}(f'_2)$ are subsets of $\text{VIOL}(f)$ because f'_1 and f'_2 are constructed from f using a sequence of CLEAR and SQUASH operators, which never add new violated edges by Claim 12 and part 1.

It remains to prove that $\text{VIOL}(f'_1)$ and $\text{VIOL}(f'_2)$ are disjoint. By part 2, there is no edge violated by f_1 whose interval crosses $[\frac{r}{2}-1, \frac{r}{2}]$. Hence, the edges violated by f_1 are partitioned into two disjoint subsets: "low" edges with intervals contained in $[0, \frac{r}{2}-1]$ and "high" edges with intervals contained in $[\frac{r}{2}, r-1]$. The edges violated by f'_1 are a subset of the low edges, since the SQUASH operator repairs all high violated edges and adds no new violated edges by Claim 12. The edges violated by f'_2 are a subset of the high edges, since the CLEAR operator used to form f_2 repairs all low violated edges by parts 2 and 3, and no new violated edges are added by Claim 12 and part 1.

(*3*) First, we show that f_3 is monotone. Since the function f_3 is constructed from f using a sequence of three CLEAR operators, parts 2 and 3 imply that there is no edge violated by f_3 whose interval crosses any of the intervals $[\frac{r}{2}-1, \frac{r}{2}]$, $[0, \frac{r}{2}-1]$, or $[\frac{r}{2}, r-1]$. Therefore, f_3 violates no edges at all and is monotone.

Now the distance from f to the set of monotone functions is at most the distance from f to the particular monotone function f_3, and we get:

$$\epsilon_{\text{M}}(f) \leq \text{dist}(f, f_3) \leq \text{dist}(f, f_1) + \text{dist}(f_1, f_2) + \text{dist}(f_2, f_3) = \epsilon_{\text{M}}(f') + \epsilon_{\text{M}}(f'_1) + \epsilon_{\text{M}}(f'_2).$$

The last step uses Claim 13. ■

4.3 Proof of Theorem 3

In this subsection we use the results of the preceding lemma to prove Theorem 3. The proof is by induction on s with the inductive hypothesis that for every function $f : S \mapsto \Xi$ where $|\Xi| = 2^s$,

$$\epsilon_{\text{M}}(f) \leq C \cdot \text{DETECT}(f, D) \cdot s.$$

In the base case where $s = 1$, the hypothesis holds by the assumption stated in the theorem. Now assume that the hypothesis holds for $s-1$ to prove that it holds for s. We can reason as

follows:

$$\begin{aligned}\epsilon_{\mathrm{M}}(f) &\leq \epsilon_{\mathrm{M}}(f') + \epsilon_{\mathrm{M}}(f'_1) + \epsilon_{\mathrm{M}}(f'_2) \\ &\leq C \cdot \textsc{Detect}(f', D) + C \cdot \textsc{Detect}(f'_1, D) \cdot (s-1) + C \cdot \textsc{Detect}(f'_2, D) \cdot (s-1) \\ &\leq C \cdot (\textsc{Detect}(f, D) + \textsc{Detect}(f, D)(s-1)) \\ &= C \cdot \textsc{Detect}(f, D) \cdot s\end{aligned}$$

The first inequality was proved in part 3 of Lemma 15. The second inequality uses the induction hypothesis; recall that the range of f' has size 2^1, and the ranges of f'_1 and f'_2 have size $r/2 = 2^{s-1}$. The third step uses parts 1 and 2 of Lemma 15, and the final step is simplification. This completes the proof.

5 An Alternative Algorithmic Schema

The algorithmic schema described in Section 2 encompasses all the algorithms we have dealt with so far. In this section we present another monotonicity testing algorithmic schema for boolean functions. We describe an algorithm which adheres to this schema and has query complexity of $O\left((n/\epsilon)\log^2(n/\epsilon)\right)$ and time complexity of $O\left((n/\epsilon)\log^3(n/\epsilon)\right)$ (that is, independent of $|\Sigma|$), proving Theorem 4.

The new schema uses an arbitrary probability distribution $q_1 : \Sigma \mapsto [0,1]$ on the points[4] of Σ, and an arbitrary probability distribution $q_2 : \mathbb{N} \mapsto [0,1]$ on the integers. The functions t and s, referred to below, depend on q_1 and q_2.

Algorithmic schema: Given parameters ϵ, n, and oracle access to an arbitrary boolean function f, set $Q = 0$ (Q will count the number of queries performed) and repeat the following steps up to $t(\epsilon, n)$ times or until $Q \geq s(\epsilon, n)$.

1. Uniformly select dimension $i \in [n]$, prefix $\alpha \in \Sigma^{i-1}$, and suffix $\beta \in \Sigma^{n-i}$.
2. Select an integer k according to q_2. Increase Q by k.
3. Select k points $j_1, ..., j_k$ from Σ according to q_1.
4. Sort the points selected (according to their index j_a). For each pair of consecutive indices $j_a < j_b$, set $x = \alpha\, j_a\, \beta$ and $y = \alpha\, j_b\, \beta$. If $f(x) > f(y)$ (i.e., (x, y) witnesses that f is not monotone), then reject.

If all iterations were completed without rejecting then accept.

The analysis of this schema will be similar to the analysis presented in Section 2. An algorithm that belongs to this schema is fully specified by the distributions q_1, q_2 and the functions $t(\epsilon, n)$ and $s(\epsilon, n)$. Note that if among the points selected in Step 3 there is *some* pair $j_a < j_b$ such that $f(\alpha\, j_a\, \beta) > f(\alpha\, j_b\, \beta)$, then there must exist such a *consequtive* pair.

An iteration of the above test defines a probability distribution $P(q_1, q_2)$ on subsets of Σ^n. We are looking for distributions q_1 and q_2 for which the probability $\textsc{Detect}(f, q_1, q_2)$ of detecting a violating pair of points in a random subset $A \sim P(q_1, q_2)$ is well lower bounded as a function of $\epsilon_{\mathrm{M}}(f)$ and for which the expected size of A is not too large.

[4] As opposed to the preceding schema where elementary events in the probability space were the *pairs* in $\Sigma \times \Sigma$.

Let us now specify a "good" pair of distributions q_1,q_2. Let $\ell = \lceil \log(2n/\epsilon) \rceil$.

DISTRIBUTION q_1: This is the uniform distribution on Σ.

DISTRIBUTION q_2: This distribution is supported on the ℓ integer points $2, 4, \ldots, 2^{\ell-1}, 2^{\ell}$. It assigns probability $2^{-\ell+1}$ to 2^{ℓ}, and for $k = 1, \ldots \ell - 1$, it assigns a probability of $2^{-(\ell-k)}$ to $2^{\ell-k}$.

As before, we start the analysis from the one-dimensional case.

Lemma 16 *Let $g : \Sigma \mapsto \{0,1\}$ with $\epsilon_{\mathrm{M}}(g) = \epsilon$. Then, choosing uniformly $k \geq 2$ points $j_1, ..., j_k$ in Σ we will observe a violation of monotonicity with probability $p(k) \geq (1 - e^{-\epsilon k/2})^2$.*

Proof: We repeat almost verbatim an argument used in the proof of Lemma 6. Since Σ is a totally ordered set of size d, we may view g as a vector in $\{0,1\}^d$, and let $S(g)$ denote its sorted version. Then $S(g) = 0^z 1^{d-z}$, where z denotes the number of zeros in g. Thus, for some $r \geq 0$, the vector g has r 1-entries within its z-prefix and r 0-entries in its $(d-z)$-suffix. Clearly, $r \geq \epsilon_{\mathrm{M}}(g) \cdot d$, since otherwise changing the r 1-entries in the prefix to zero we obtain a monotone vector.

It follows that there are two subsets $A, B \subseteq \Sigma$, with $|A|, |B| \geq \epsilon_{\mathrm{M}}(g)|\Sigma|$, such that g restricted to A is the all–1 vector, g restricted to B is the all–0 vector, and for all $a \in A$, $b \in B$ holds $a < b$.

Therefore, to detect a violation we just have to choose a point in A and a point in B. The probability that this event happens is at least the probability that there is a point in A among the first $k/2$ choices and a point in B among the second $k/2$ choices. These two events are independent, and therefore:

$$p(k) \geq (1 - (1-\epsilon)^{k/2})^2 \geq (1 - e^{-\epsilon k/2})^2.$$

■

The following lemma is a key step in the prove of Theorem 4. It uses Lemma 6 to perform the dimension reduction in this setting.

Lemma 17 *For any $f : \Sigma^n \mapsto \{0,1\}$ with $\epsilon_{\mathrm{M}}(f) \geq \epsilon$ holds:*

$$\textsc{Detect}(f, q_1, q_2) \geq c_0 \cdot \frac{1}{\ell \cdot 2^{\ell}},$$

where c_0 is an absolute constant ($c_0 > 1/100$).

Proof: Let Ω denote the probability space of all triples $i \in [n]$, $\alpha \in \Sigma^{i-1}$, $\beta \in \Sigma^{n-i}$ endowed with uniform distribution. Let $\mathbb{N}$ be endowed with distribution q_2. We consider two random variables X and Y, where X is defined on Ω and Y on the product probability space $\Omega \times \mathbb{N}$. For $\omega = (i, \alpha, \beta) \in \Omega$, let $X(\omega) \stackrel{\mathrm{def}}{=} \epsilon_{\mathrm{M}}(f_{i,\alpha,\beta})$. Also, for $\omega = (i, \alpha, \beta) \in \Omega$ and $k \in \mathbb{N}$, let $Y(\omega, k)$ be the probability that chosing k points in Σ uniformly and independently, and checking all the obtained ordered pairs, a violation is detected for $f_{i,\alpha,\beta}$. Item 2 of Lemma 6 tells us that $\mathbf{E}_{\Omega}(X) \geq \frac{\epsilon}{2n} \geq \frac{1}{2^{\ell}}$. We have to prove that

$$\textsc{Detect}(f, q_1, q_2) = \mathbf{E}_{\Omega \times \mathbb{N}} Y(\omega, k) \geq c_0 \cdot \frac{1}{\ell \cdot 2^{\ell}}.$$

We define two families of sets: $\{A_j\} \subseteq \Omega$ and $\{B_j\} \subseteq \mathbb{N}$. For any $0 \leq j \leq \ell - 1$, let

$$A_j = \left\{ \omega \in \Omega \ : \ 2^{j-1} \mathbf{E}(X(\omega)) \leq X \leq 2^j \mathbf{E}(X) \right\}$$

and

$$B_j = \left\{ k \in \mathbb{N} \ : \ k \geq 2^{\ell - j} \right\}.$$

Next, we show that the probabilities of (at least some) of the sets A_j, B_j are not too small. Since by definition of the sets A_j we have $X(\omega) < \frac{1}{2}\mathbf{E}(X)$ on any ω in the complement of $\cup_j A_j$, we obtain

$$\sum_{j=0}^{\ell-1} \Pr[A_j] \cdot 2^j \mathbf{E}(X) \geq \frac{1}{2}\mathbf{E}(X),$$

implying the existence of an index $0 \leq j_0 \leq \ell - 1$ with $\Pr[A_{j_0}] \geq \frac{1}{\ell \cdot 2^{j_0+1}}$. As for the sets B_j, note that (this is the crucial property of the distribution q_2) $\Pr[B_j] \geq 2^{j-\ell}$ for all $0 \leq j \leq \ell - 1$.

The preceding lemma states that for any $0 \leq j \leq \ell - 1$ and for any pair $(\omega, k) \in A_j \times B_j$, it holds that $Y(\omega, k) \geq (1 - e^{-X(\omega)\cdot k/2})^2 \geq (1 - e^{-1/4})^2$. Therefore:

$$\textsc{Detect}(f, q_1, q_2) = \mathbf{E}(Y) \geq \mathbf{E}(Y | A_{j_0} \times B_{j_0}) \cdot \Pr[A_{j_0} \times B_{j_0}] \geq (1 - e^{-1/4})^2 \cdot \Pr[A_{j_0}] \cdot \Pr[B_{j_0}]$$

$$\geq (1 - e^{-1/4})^2 \cdot \frac{1}{\ell \cdot 2^{j_0+1}} \cdot 2^{j_0-\ell} = \frac{(1 - e^{-1/4})^2}{2} \cdot \frac{1}{\ell \cdot 2^\ell},$$

and we are done. ■

5.1 Proof of Theorem 4

Consider a monotonicity testing algorithm $\mathcal{A}$ consistent with the algorithmic schema described in this section, with the distributions q_1 and q_2 we have just defined. Set $t(\epsilon, n) = c \cdot \ell 2^\ell$ and $s(\epsilon, n) = 4c \cdot \ell^2 2^\ell$ where c is a large constant. Clearly this algorithm always accepts a monotone function and its query complexity is bounded by $s(\epsilon, n) = O((n/\epsilon)\log^2(n/\epsilon))$. As the time complexity is dominated by sorting the points selected in each iteration of the algorithm, it is at most $O(s(\epsilon, n)\log(s(\epsilon, n)) = O((n/\epsilon)\log^3(n/\epsilon))$. We have to prove that, for sufficiently large constant c, the algorithm rejects with probability $\geq \frac{2}{3}$ any $f : \Sigma^n \mapsto \{0, 1\}$ with $\epsilon_{\mathrm{M}}(f) \geq \epsilon$.

Let $\mathcal{A}'$ be an algorithm identical to $\mathcal{A}$, except for the fact that there is no upper limit on the number of queries Q it performs. Let B be the event $Q \geq s(\epsilon, n)$. Then $\{\mathcal{A} \text{ accepts }\} = \{\mathcal{A}' \text{ accepts }\} \bigcup B$, and therefore $\Pr[\mathcal{A} \text{ rejects }] \geq \Pr[\mathcal{A}' \text{ rejects }] - \Pr[B]$.

By Lemma 17, the probability that one iteration of $\mathcal{A}'$ detects a violation is at least $\Omega(\frac{1}{\ell \cdot 2^\ell})$. Therefore, if c is large enough, $t(\epsilon, n) = c\ell 2^\ell$ iterations of $\mathcal{A}'$ will detect a violation with probability $\geq \frac{11}{12}$, say. To conclude the proof that $\mathcal{A}$ is a tester of monotonicity, it suffices to show that for sufficiently large c, $\Pr[B] \leq \frac{1}{4}$.

Let k_i for $i = 1...t(\epsilon, n)$ count the number of queries $\mathcal{A}'$ performs in the i'th iteration. Then $k_1, ..., k_{t(\epsilon,n)}$ are $t := t(\epsilon, n)$ independent random variables with distribution q_2, and $Q = \sum_{i=1}^t k_i$. Therefore $\mathbf{E}(Q) = t \cdot \mathbf{E}(k) = t \cdot \ell$, and $\mathrm{Var}(Q) = t \cdot \mathrm{Var}(k_i) \leq t \cdot 2^\ell$. Chebyshev's inequality applied to Q readily gives

$$Pr(Q \geq 4c \cdot \ell^2 2^\ell) \leq Pr(|Q - \mathbf{E}(Q)| \geq c \cdot \ell^2 2^\ell) \leq \frac{\mathrm{Var}(Q)}{c^2 \cdot \ell^4 2^{2\ell}} \leq \frac{1}{c^2 \cdot \ell^3}$$

and we are done.

Acknowledgments

We would like to thank Noga Alon, Michael Krivelevich, Michael Sipser and Madhu Sudan for helpful discussions. In particular, Proposition 10 is due to Michael Krivelevich.

References

[1] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and intractability of approximation problems. *JACM*, 45(3):501–555, 1998.

[2] S. Arora and M. Sudan. Improved low degree testing and its applications. In *Proceedings of STOC97*, pages 485–495, 1997.

[3] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing in characteristic two. In *Proceedings of FOCS95*, pages 432–441, 1995.

[4] M. Bellare, S. Goldwasser, C. Lund, and A. Russell. Efficient probabilistically checkable proofs and applications to approximation. In *Proceedings of STOC93*, pages 294–304, 1993.

[5] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *JACM*, 47:549–595, 1993.

[6] F. Ergun, S. Kannan, S. R. Kumar, R. Rubinfeld, and M. Viswanathan. Spot-checkers. In *Proceedings of STOC98*, pages 259–268, 1998.

[7] P. Gemmell, R. Lipton, R. Rubinfeld, M. Sudan, and A. Wigderson. Self-testing/correcting for polynomials and for approximate functions. In *Proceedings of STOC91*, pages 32–42, 1991.

[8] O. Goldreich, S. Goldwasser, E. Lehman, and D. Ron. Testing monotonicity. In *Proceedings of FOCS98*, 1998.

[9] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *JACM*, 45(4):653–750, 1998. An extended abstract appeared in the proceedings of FOCS96.

[10] O. Goldreich and D. Ron. Property testing in bounded degree graphs. In *Proceedings of STOC97*, pages 406–415, 1997.

[11] O. Goldreich and D. Ron. A sublinear bipartite tester for bounded degree graphs. In *Proceedings of STOC98*, pages 289–298, 1998. To appear in *Combinatorica*, 1999.

[12] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of STOC97*, pages 475–484, 1997.

[13] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

A Other Stuff of Interest (Informal)

Let S be any partial order, e.g. $S = \Sigma^n$.

A.1 $\epsilon_{\mathrm{M}}(f)$ and its properties

One might ask the question of whether given (violation) oracle access to $f : S \mapsto \Xi$, we can compute $\epsilon_{\mathrm{M}}(f)$ *exactly* in time poly($|S|$). The answer is *yes*. We saw in Lemma 7 (generalized to any S) that $\epsilon_{\mathrm{M}}(f)|S|$ is the size of vertex cover of $\textsc{Viol}(f)$, hence, $(1-\epsilon_{\mathrm{M}}(f))|S|$ is the size of its maximal independent set (MIS). In general, computing MIS of a graph is a very hard problem. However, $\textsc{Viol}(f)$ is a *transitive* graph (a graph $G = (V, E)$ is transitive if $(x, y), (y, z) \in E \implies (x, z) \in E$), since it consists of violated edges of f. It is well know that transitive graphs are *perfect*. And for perfect graph we can compute its MIS in polynomial time. Thus,

Lemma 18 *One can compute $\epsilon_{\mathrm{M}}(f)$ in time* poly($|S|$).

We now give yet alternative characterization of $\epsilon_{\mathrm{M}}(f)$. A *chain* in S is a sequence of points $x_1, \ldots, x_k$ s.t. $x_1 \prec x_2 \prec \ldots \prec x_k$. Given a function f and a chain $x_1, \ldots, x_k$, we say that the chain is *decreasing* if $f(x_1) > f(x_2) > \ldots > f(x_k)$. Clearly, for any chain decreasing chain C of f, we must change at least $|C| - 1$ values of f along this chain in order to make f monotone, i.e. to leave at most 1 value of f unchanged. Given a collection of k decreasing chains of f that cover the entire vertex set S, we get an upper bound of k on the number of points where we can leave the value of f *unchanged*, so the number of chains in the smallest *decreasing chain cover* of f upper bounds the number of points we can leave intact. Remarkably, this bound is *tight*. In order to make f monotone by changing the minimal number ($\epsilon_{\mathrm{M}}(f)|S|$) of values of f, we need to partition the entire S into the *smallest* possible number of disjoint decreasing chains. Then we can in fact leave one point in each chain intact and make f monotone.

Theorem 19 (decreasing chain theorem) *For any partial order S, and any $f : S \mapsto \Xi$, the MAXIMUM number of POINTS WE CAN LEAVE and still make f monotone (i.e. $(1-\epsilon_{\mathrm{M}}(f))|S|$) = MINIMUM number of disjoint DECREASING CHAINS of f that partition points of S.*

The theorem turns out to follow immediately from the famous Dilworth theorem:

Theorem 20 (Dilworth) *In any partial order P (which will be the one induced by* $\textsc{Viol}(f)$ *above), the LARGEST ANTICHAIN (number of points we leave intact, as no two such points can have a violated edge in between) equals to the MINIMUM NUMBER OF CHAINS (which in* $\textsc{Viol}(f)$ *are the decreasing chains of f) needed to partition all points of P. Shortly, MAX antichain = MIN chain cover.*

The statement above already contains the proof, we apply the theorem to the partial order induced by the violated edges of f.

A.2 Other Observations

We can substantially generalize Proposition 9.

Definition 4 *Given a directed graph $G = (V, E)$, a 2-spanner of G is a subset $E' \subseteq E$ s.t. for all $(x, z) \in E$, x and z are of distance at most 2 in $G' = (V, E')$. That is, either $(x, z) \in E'$, or there is $y \in V$ s.t. $(x, y), (y, z) \in E'$.*

Given G, the hope is to construct a 2-spanner of G with the minimal number of edges. We saw that the set P in the definition of p_1 is a 2-spanner of a directed complete graph on d vertices,

which the graph of our total ordering on d vertices. It is not hard to see that any 2-spanner of the directed complete graph must have $\Omega(d \log d)$ edges, like G_P does.

Essentially analogous to the proof of Proposition 9 (i.e. we show that there is a matching of violated edges of size $\Omega(\epsilon_{\mathrm{M}}(f)|S|)$, and then P has at least $\Omega(\epsilon_{\mathrm{M}}(f)|S|)$ violated edges as well), we can show:

Lemma 21 *Let S be any partial order and let P be the smallest set of edges that is a 2-spanner of G. Let D be uniform distribution over P. Then for any $f : S \mapsto \Xi$,*

$$\textsc{Detect}(f, D) = \Omega\left(\frac{|S|}{|P|} \cdot \epsilon_{\mathrm{M}}(f)\right)$$

Thus, if a partial order has a very sparse 2-spanner, we get an efficient test. Unfortunately, the best spanner we can construct for Σ^n (which might be the best possible in fact) is just the direct product of the best spanner for Σ, so it has size $O(d^n \log^n d)$, giving a test of complexity $O(\log^n d/\epsilon)$, inferior to what we have using range reduction.

A.3 Some Questions

What is the best 2-spanner for (the transitive closure) of Σ^n?

Range reduction holds for any S, so we can concentrate on smooth testers for the boolean range only. *Is $\log r$ slowdown in the range reduction tight?*

Finally, couple of very related conjectures for general partial orders.

Conjecture 1: For any S, there is a tester of complexity $\mathrm{poly}(\log|S|, 1/\epsilon)$. In fact, $O(1/\epsilon^2)$ seems possible. Try $\Xi = \{0, 1\}$ as a start.

Open, might be true especially for $\Xi = \{0, 1\}$.

Conjecture 2: Same as above except can use value oracle rather than violation oracle.

Even more likely to hold.

Conjecture 3: Same as conjecture 1, but *smooth* testers, i.e. $\mathrm{poly}(\log|S|)/\epsilon$. By range reduction, can concentrate on $\Xi = \{0, 1\}$.

Seems false. The following is a bad partial order: $S = \{a_1, \ldots a_d, b_1, \ldots, b_d\}$, where $a_i \prec b_j$ for all $i, j \in [d]$. Using Yao minimax principle can show that need $\Omega(1/\epsilon^2)$ probes (which is easily seen to be possible by sampling an edge uniformly), which can be as bad as $\Omega(d/\epsilon)$.

Conjecture 4: Same as conjecture 3, i.e. $\mathrm{poly}(\log|S|)/\epsilon$ but with the value oracle (Note, can't use range reduction for the value oracle). Start maybe from $\Xi = \{0, 1\}$.

Might be true. Note, above example can be easily done in $O(1/\epsilon)$ probes. Btw, can we do the sorted list with $O(1/\epsilon)$, or Rubinfeld, etc. lower bound holds for value access as well?