Electronic Colloquium on Computational Complexity, Report No. 187 (2018)

Domain Reduction for Monotonicity Testing: A $o(d)$ Tester for Boolean Functions on Hypergrids

Hadley Black* Deeparnab Chakrabarty† C. Seshadhri‡

Abstract

Testing monotonicity of Boolean functions over the hypergrid, $f : [n]^d \to \{0, 1\}$, is a classic problem in property testing. When the range is real-valued, there are $\Theta(d \log n)$-query testers and this is tight. In contrast, the Boolean range *qualitatively* differs in two ways:

- *Independence of* n: There are testers with query complexity independent of n [Dodis et al. (RANDOM 1999); Berman et al. (STOC 2014)], with linear dependence on d.
- *Sublinear in* d: For the $n = 2$ hypercube case, there are testers with $o(d)$ query complexity [Chakrabarty, Seshadhri (STOC 2013); Khot et al. (FOCS 2015)].

It was open whether one could obtain both properties simultaneously. This paper answers this question in the affirmative. *We describe a* $\widetilde{O}(d^{5/6})$*-query monotonicity tester for* $f : [n]^d \to \{0, 1\}$.

Our main technical result is a *domain reduction theorem* for monotonicity. For any function f, let ε_f be its distance to monotonicity. Consider the restriction $\hat{f}$ of the function on a random $[k]^d$ sub-hypergrid of the original domain. We show that for $k = \mathrm{poly}(d/\varepsilon)$, the expected distance of the restriction $\mathbf{E}[\varepsilon_{\hat{f}}] = \Omega(\varepsilon_f)$. Therefore, for monotonicity testing in d dimensions, we can restrict to testing over $[n]^d$, where $n = \mathrm{poly}(d/\varepsilon)$. Our result follows by applying the $d^{5/6} \cdot \mathrm{poly}(1/\varepsilon, \log n, \log d)$-query hypergrid tester of Black-Chakrabarty-Seshadhri (SODA 2018).

*Department of Computer Science, University of California, Los Angeles. Email: hablack@cs.ucla.edu. Part of this work was done while the author was at University of California, Santa Cruz.

†Department of Computer Science, Dartmouth College. Email: deeparnab@dartmouth.edu. Supported by NSF CCF-1813165.

‡Department of Computer Science, University of California, Santa Cruz. Email: sesh@ucsc.edu. Supported by NSF TRIPODS CCF-1740850, CCF-1813165.

ISSN 1433-8092

1 Introduction

Monotonicity testing over hypergrid domains is a fundamental problem in property testing. Consider the hypergrid $[n]^d$, where $\prec$ denotes the coordinate-wise partial ordering. Let R be a total order. A function $f : [n]^d \to R$ is monotone if $f(x) \le f(y)$ for any $x \prec y$. The distance of a function f to monotonicity is the Hamming distance to the nearest monotone function, that is, $\varepsilon_f := \min_{g \in \mathcal{M}} d(f, g)$, where $d(f, g) = n^{-d} \cdot |\{x \in [n]^d : f(x) \neq g(x)\}|$, and $\mathcal{M}$ is the set of all monotone functions. A monotonicity tester is a randomized algorithm that makes queries to f and accepts with probability $\ge 2/3$ if the function is monotone, and rejects with probability $\ge 2/3$ if $\varepsilon_f \ge \varepsilon$, where $\varepsilon \in (0, 1)$ is an input parameter. The challenge is to determine the minimum query complexity of a monotonicity tester.

One of the earliest results in property testing is the $O(d/\varepsilon)$-query "edge-tester" due to Goldreich et al. [GGL+00] (see also [Ras99]) for testing monotonicity of Boolean functions over the hypercube, that is, $f : \{0, 1\}^d \to \{0, 1\}$. In the last few years, considerable work [CS14a, CST14, CDST15, KMS15, BB16, CWX17] has improved our understanding of Boolean monotonicity testing in the hypercube domain. In particular, Khot, Minzer, and Safra [KMS15] gave a $\widetilde{O}(\sqrt{d}/\varepsilon^2)$ query[1] tester, and Chen, Waingarten, and Xie [CWX17] show that any such tester must make $\widetilde{\Omega}(d^{1/3})$ queries. In contrast, for real-valued functions over the hypercube $f : \{0, 1\}^d \to \mathbb{R}$, the complexity is known to be $\Theta(d/\varepsilon)$ [BBM12, CS13, CS14b]. We give more details of previous work in §1.4.

In this paper, we investigate monotonicity testing for Boolean functions over the d-dimensional, n-hypergrid, $f : [n]^d \to \{0, 1\}$. Dodis et al. [DGL+99] (with improvements by Berman, Raskhodnikova, and Yaroslavtsev [BRY14a]) gave a $\widetilde{O}(d/\varepsilon)$-query tester. The important feature to note is the *independence* on n. Again, contrast this with the real-valued case; monotonicity testing of one-dimensional functions $f : [n] \to \mathbb{R}$ requires $\Omega(\log n)$ queries [EKK+00, Fis04]. Recently, Black, Chakrabarty, and Seshadhri [BCS18] gave a $\widetilde{O}(d^{5/6} \log^{4/3} n \, \varepsilon^{-4/3})$-query tester. Although the dependence on d was sublinear, there was a dependence on n. The following question remained open:

> *Is there a monotonicity tester for functions $f : [n]^d \to \{0, 1\}$, whose query complexity is independent of n and sublinear in d?*

The main outcome of this paper is an affirmative answer to this question.

Theorem 1.1 (*Informal*)**.** *There is a one-sided, non-adaptive $\widetilde{O}(d^{5/6}\varepsilon^{-7/3})$-query monotonicity tester for Boolean functions $f : [n]^d \to \{0, 1\}$.*

1.1 Domain Reduction

A natural approach, at least in hindsight, to tackle Boolean monotonicity testing over the hypergrid is to try reducing it to Boolean monotonicity testing over the hypercube. For a function f over $[n]^d$, consider the restriction $\widehat{f}$ to a *random* hypercube in this hypergrid. More precisely, for each dimension $i \in [d]$, sample two independent u.a.r values $a_i < b_i$ in $[n]$ and let $\widehat{f}$ be the restriction of f on the hypercube formed by the Cartesian product $\prod_{i=1}^{d} \{a_i, b_i\}$. If the expectation of $\varepsilon_{\widehat{f}}$ is $\Omega(\varepsilon_f)$, then we obtain a hypergrid tester by first *reducing our domain* to a random hypercube and then simply applying the best known monotonicity tester on the hypercube. However, we show that this *does not work*. In §5, we describe a function $f : [n]^d \to \{0, 1\}$ such that $\varepsilon_f = \Omega(1)$, but the restriction of f on a random hypercube is monotone with probability $1 - \Theta(1/d)$ (see Theorem 5.1).

[1] Throughout the paper $\widetilde{O}$ hides $\log(d/\varepsilon)$ factors.

Nonetheless, one can consider the question of reducing the domain to a $[k]^d$ hypergrid, for some parameter $k \ll n$. For each $i \in [d]$, consider sampling a subset $T_i \subset [n]$ by taking k i.i.d. uniform samples from $[n]$. Let $\boldsymbol{T} = \prod_{i=1}^d T_i$ and $f_{\boldsymbol{T}}$ be the restriction of f to $\boldsymbol{T}$. For k independent of n, can we lower bound $\mathbf{E}_{\boldsymbol{T}}[\varepsilon_{f_{\boldsymbol{T}}}]$? We refer to this as the problem of *domain reduction with k samples*. Our main technical result is a domain reduction theorem with $k = \mathrm{poly}(d/\varepsilon_f)$.

Theorem 1.2 (*Informal*)**.** *The expected distance to monotonicity of f restricted to a random $[k]^d$ hypergrid with $k = \Theta((d/\varepsilon_f)^7)$, is $\Omega(\varepsilon_f)$.*

The construction of §5 actually shows that such a theorem is impossible for $k = o(\sqrt{d})$, and thus, domain reduction requires k to be polynomial in d.

We can invoke [BCS18] to prove Theorem 1.1. Sample a random $[k]^d$ hypergrid denoted $\boldsymbol{T}$ and apply the tester of [BCS18] on $f_{\boldsymbol{T}}$. The final query complexity is $\widetilde{O}(d^{5/6}) \cdot \mathrm{poly}\log k$. Setting $k = \mathrm{poly}(d/\varepsilon)$, one gets a purely sublinear-in-d tester (see §1.5 for a formal proof).

An obvious question is whether the dependence on d can be brought down to $\sqrt{d}$ as in the hypercube case. Theorem 1.2 allows us to assume that $n = \mathrm{poly}(d)$. Therefore if one could design a $\sqrt{d} \cdot \mathrm{poly}\log n$ query monotonicity tester for the domain $[n]^d$, then Theorem 1.2 can be used as a black box to achieve an $\widetilde{O}(\sqrt{d})$ monotonicity tester.

Implication for Other Notions of Distance: Berman, Raskhodnikova, and Yaroslavtsev [BRY14a] introduce the notion of L_p testing, where $f : [n]^d \to [0,1]$ and the distance between functions is measured in terms of L_p-norms. They prove that monotonicity testing with L_p distances can be reduced to (non-adaptive, one-sided) Boolean monotonicity testing. Thus, Theorem 1.1 implies an L_p-tester for monotonicity over hypergrids, with the same query complexity.

Previous work has also considered distance under product distributions [HK08, CDJS15]. Let $\mathcal{D}$ be a product distribution over $[n]^d$, and define $d(f,g) = \mathbf{Pr}_{x\sim\mathcal{D}}[f(x) \neq g(x)]$. Chakrabarty et al. [CDJS15] observed that monotonicity testing of $f : [n]^d \to R$ over product distributions can be reduced to monotonicity testing of $f : [N]^d \to R$ over the uniform distribution (standard distance). Here, N is potentially much larger than n, and depends on the probabilities in $\mathcal{D}$. By this observation, for the Boolean range case $R = \{0,1\}$, the query complexity of Theorem 1.1 holds for distances measured according to any product distribution. Note that the independence of n is *necessary* in the reduction to get a query complexity independent of the distribution $\mathcal{D}$.

Domain Reduction for Variance: Recent works [CS14a, KMS15, BCS18] have shown that certain isoperimetric theorems for the undirected hypercube have *directed analogues* where the variance is replaced by the distance to monotonicity. Interestingly, for the case of domain reduction, the variance and distance to monotonicity behave differently. While domain reduction for the distance to monotonicity requires $k \geq \sqrt{d}$ (Theorem 5.1), we show that the expected variance of a restriction of f to a random hypercube ($k = 2$) is at least half the variance of f (see Theorem A.1). This statement may be of independent interest. We were unable to find a reference to such a statement and provide a proof in §A.

1.2 The Formal Result

Fix a function $f : [n]^d \to \{0,1\}$. We construct d random (multi-) sets $T_1, \ldots, T_d \subseteq [n]$, each formed by taking k i.i.d. uniform samples from $[n]$ with replacement. We define $\boldsymbol{T} := T_1 \times \cdots \times T_d$ and let $f_{\boldsymbol{T}}$ denote f restricted to $\boldsymbol{T}$. We treat duplicate elements of a multi-set as being distinct copies of that element, which are then treated as immediate neighbors in the total order. The function value of $f_{\boldsymbol{T}}$ is the same on these distinct copies. In our applications $k \ll n$ and so we can assume that T_i contains a duplicate with negligible probability. Nonetheless, sampling with replacement allows for no conditions on k.

Theorem 1.3 (Domain Reduction Theorem)**.** *Let $f : [n]^d \to \{0,1\}$ be any function and let $k \in \mathbb{Z}^+$. If $\boldsymbol{T} = T_1 \times \cdots \times T_d$ is a randomly chosen sub-grid, where for each $i \in [d]$, T_i is a (multi)-set formed by taking k i.i.d. samples from the uniform distribution on $[n]$, then*

$$\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f - \frac{C \cdot d}{k^{1/7}}$$

where $C > 0$ is a universal constant.

In particular, $k = O((d/\varepsilon_f)^7)$ samples in each dimension is sufficient to preserve the distance to monotonicity. The theorem is a direct corollary of the following lemma, applied to each dimension. This lemma is the heart of our result and we give an overview of its proof in §1.3.

Lemma 1.4 (Domain Reduction Lemma)**.** *Let $f : [n] \times \left(\prod_{i=2}^d [n_i]\right) \to \{0,1\}$ be any function over a rectangular hypergrid for some $n, n_2, \ldots, n_d \in \mathbb{Z}^+$ and let $k \in \mathbb{Z}^+$. Choose T to be a (multi-) set formed by taking k i.i.d. samples from the uniform distribution on $[n]$ and let f_T denote f restricted to $T \times \left(\prod_{i=2}^d [n_i]\right)$. Then $\mathbf{E}_T\left[\varepsilon_f - \varepsilon_{f_T}\right] \leq \frac{C}{k^{1/7}}$ where $C > 0$ is a universal constant.*

1.3 Proving the Domain Reduction Lemma : Overview

Let us start with the simple case of $d = 1$ (the line). Monotonicity testers for the line immediately imply domain reduction for $d = 1$ [DGL+99, BRY14a]. A u.a.r sample of $\widetilde{O}(1/\varepsilon_f)$ points in $[n]$ contains a monotonicity violation with high probability, and thus the restriction of f to the sample has distance $\Omega(\varepsilon_f)$. We note that these arguments typically get a lower bound of $\varepsilon_f/2$. Therefore, even if we could generalize this argument to the setting of Lemma 1.4, we would need to apply it d times to get the full domain reduction (Theorem 1.3). That would imply a final lower bound of $\varepsilon_f/2^d$, which has little value towards proving a sublinear-in-d bound.

Indeed, the first baby step towards Lemma 1.4 is to get a stronger domain reduction just for the line. We prove that if one samples $\Theta(d^2/\varepsilon_f^2)$ points, then the expected distance of the restricted function is at least $\varepsilon_f(1 - 1/d)$. Numerically speaking, this is encouraging news, since we could at least hope to iterate this argument d times. Of course, this result for the line alone is not enough to deal with the structure of general hypergrids, but forms a part of our final proof.

Let us go to the general case of Lemma 1.4. For brevity, we let $\mathcal{D} := [n] \times \left(\prod_{i=2}^d [n_i]\right)$ and $\mathcal{D}_T := T \times \left(\prod_{i=2}^d [n_i]\right)$ denote the original and reduced domains, respectively. Note that $|\mathcal{D}_T| = \frac{k}{n}|\mathcal{D}|$.

The violation graph of f has vertex set $\mathcal{D}$ and an edge (x, y) iff $x \prec y$ and $f(x) = 1, f(y) = 0$. A theorem of [FLN+02] states that any *maximum cardinality* matching in M in the violation graph satisfies $|M| = \varepsilon_f|\mathcal{D}|$. Fix such a matching M. For a fixed sample T, we let M_T denote a maximum cardinality matching in the violation graph of f_T. To argue about ε_{f_T}, we need to to lower bound the expected size $|M_T|$. To do so, we lower bound the expected number of *endpoints of* M that can still be matched in the violation graph of f_T.

We use the following standard notions of lines and slices in $\mathcal{D}$, with respect to the first dimension. Below, for $x \in \mathcal{D}$, the vector x_{-1} is used to denote $(x_2, x_3, \ldots, x_d)$.

- (Lines in $\mathcal{D}$) $\mathcal{L} := \left\{\ell_z : z \in \prod_{i=2}^d [n_i]\right\}$ where $\ell_z := \{x \in \mathcal{D} : x_{-1} = z\}$.
- (Slices in $\mathcal{D}$) $\mathcal{S} := \{S_i : i \in [n]\}$ where $S_i := \{x \in \mathcal{D} : x_1 = i\}$.

We partition M into a collection of "local" matchings for each line:

- (Line Decomposition of M) For each $\ell \in \mathcal{L}$: $M^{(\ell)} := \{(x,y) \in M : x \in \ell\}$.

We find a large matching in the violation graph of f_T by doing a line-by-line analysis. In particular, we define the following matching $M_T^{(\ell)}$.

- (The matching $M_T^{(\ell)}$) For each $\ell \in \mathcal{L}$, let $M_T^{(\ell)}$ be any maximum cardinality violation matching with respect to f_T on the set of vertices that (a) are matched by $M^{(\ell)}$, and (b) lie in some slice S_i where $i \in T$.

We stress that $M_T^{(\ell)}$ is *not* a subset of $M^{(\ell)}$; the endpoints of the pairs in $M_T^{(\ell)}$ are a subset of the endpoints of the pairs in $M^{(\ell)}$, but the actual pairs can be different. The above definition implies that the union of $M_T^{(\ell)}$ over all $\ell \in \mathcal{L}$ is a valid matching M_T in the violation graph of f_T and that $M_T^{(\ell)} \cap M_T^{(\ell')} = \emptyset$ for all $\ell \neq \ell' \in \mathcal{L}$. We will lower bound the size of this matching, M_T.

Fix some $\ell \in \mathcal{L}$ and notice that by definition, the lower-endpoints of $M^{(\ell)}$ all lie on ℓ, and thus are all comparable. Thus, let $M^{(\ell)} = \{(x_1,y_1),\ldots,(x_m,y_m)\}$ where $x_1 \prec \cdots \prec x_m$ and observe this implies that, for any $j \in [m]$, $x_1,\ldots,x_j \prec y_j,\ldots,y_m$. Combining this with the fact that the function is Boolean, we see that *any* $x \in \{x_1,\ldots,x_j\}$ forms a violation to monotonicity with *any* $y \in \{y_j,\ldots,y_m\}$, and therefore these vertices *can be matched* in $M_T^{(\ell)}$, if their 1-coordinates are sampled by T.

Since all the x_i's lie on the same line ℓ, their 1-coordinates are distinct. Suppose that the 1-coordinates of all the y_i's were also distinct and distinct from those of the x_i's too. Under this assumption we can project all the violations onto ℓ, and the analysis becomes identical to the one-dimensional case. As alluded to above, the one-dimensional case can be handled without much difficulty (Lemma 1.6). However, the assumption that the y_i's have distinct 1-coordinates is problematic. For instance, it could also be the case that all the y_i's have the same 1-coordinate. That is, they all lie in the same slice S_a, for some $a \in [n]$. In this case, with probability $(1-k/n)$ we would have the size of $M_T^{(\ell)}$ be 0 (if $a \notin T$), which in turn implies that $\mathbf{E}_T\left[|M_T^{(\ell)}|\right]$ could be as small as $(k/n)^2 \cdot |M^{(\ell)}|$. This would disprove the lemma if such a "collision of y's 1-coordinates" happened all the time. Unfortunately, there are examples of violation matchings where this happens. Consider Example 1, and the left part of Fig. 1. For the lowest line, all the corresponding y_i's in $M^{(\ell)}$ have the same coordinate. Indeed, this example is extremely pathological for our approach.

The only hope is to discover a *different* violation matching that does not have such a problem. Indeed, our main insight is that there *always* exists a violation matching M where the problem above does not arise (too often). This motivates the key definition of *stacks*; the stacks are what determine the "shape" of a matching. Formally, for any $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$, the (ℓ,S)-stack is the set of pairs $(x,y) \in M$, where $x \in \ell$ and $y \in S$.

- (Stacks) $M^{(\ell,S)} := \{(x,y) \in M^{(\ell)} : y \in S\} = \{(x,y) \in M : x \in \ell, y \in S\}$.

Often, we will use the notation "*size* of a stack (ℓ,S)" to denote $|M^{(\ell,S)}|$. To summarize the above discussion, small stacks are good news while big stacks are bad news. This is formalized in Lemma 1.6.

Intuitively, if there is a maximum cardinality matching M in the violation graph of f such that all stacks had size at most 1, then the one-dimensional argument can be directly applied (Lemma 1.6 would be enough to prove Lemma 1.4). Even if their sizes were at most a constant, this would suffice as well. Unfortunately, we do not know if this is possible. One reason for this difficulty may be that there can be various maximum cardinality matchings in the violation graph that have vastly different stack sizes (shape of the matching);

again consider Example 1. Nevertheless, we can show that for any $\lambda \geq 2$, there is a matching M where the total number of vertices participating in stacks of size at least λ is at most $|\mathcal{D}|/\mathrm{poly}(\lambda)$.

Lemma 1.5 (Stack Bound)**.** *For any integer $\lambda \geq 2$, there exists a maximum cardinality matching M in the violation graph of f, where $\sum_{(\ell,S):|M^{(\ell,S)}|\geq\lambda} |M^{(\ell,S)}| \leq \frac{6}{\sqrt{\lambda}} \cdot |\mathcal{D}|$.*

The main creativity to prove this lemma lies in the choice of M. Given a matching, we define the vector $\Lambda(M)$ which enumerates all the stack sizes in non-decreasing order. We show that the maximum cardinality matching M which has the lexicographically *largest* $\Lambda(M)$ suffices. That is, we choose M which maximizes the minimum stack size, and then subject to this maximizes the second minimum, and so on. It may seem a bit counter-intuitive that we want a matching with small stack sizes, and yet our potential function wishes to maximize the minimum. The explanation is that the sum of the stack sizes is $|M|$, which is fixed, and so in a sense maximizing the minimum also balances out the $\Lambda(M)$ vector. Of course, this is purely at an intuition level. The proof uses a matching rewiring argument to show that any large stack must be "adjacent" to many moderate size stacks. Essentially, if two stacks are appropriately aligned, one could change the matching to move points from one stack to the other. If a large stack was thus aligned with a small stack, one could rewire to get a lexicographically larger $\Lambda(\cdot)$ vector. Thus, large stacks can only potentially rewire with other large stacks. But since the function is Boolean one can show that there are many opportunities for rewiring the violation matching. One can then apply some technical charging arguments to bound the total number of points in large stacks. The full proof can be found in §3.

With the stack bound in hand, we need to generalize the one-dimensional argument to account for bounded stack sizes. This is precisely what the following lemma achieves. Thus, we can bound $|M_T^{(\ell)}|$ for all ℓ, and get the final lower bound on the distance ε_{f_T}.

Lemma 1.6 (Line Sampling)**.** *Suppose that M is a matching in the violation graph of f such that for $\lambda \in \mathbb{Z}^+$, $|M^{(\ell,S)}| \leq \lambda$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$. Then, for any $\ell \in \mathcal{L}$,*

$$\mathbf{E}_T\left[|M_T^{(\ell)}|\right] \geq \frac{k}{n} \cdot |M^{(\ell)}| - 3\lambda\sqrt{k \ln k}.$$

Note that in the one dimensional case, we have only one line ℓ and each slice is a singleton. Thus, any stack has at most one point and so we can set $\lambda = 1$. Therefore, setting $k = \omega\left(\frac{1}{\varepsilon^2}\ln(1/\varepsilon)\right)$ we get that the function restricted to this random set has expected distance to monotonicity $\varepsilon_f(1 - o(1))$. To see this, note that $\varepsilon_f = |M|/n = |M^{(\ell)}|/n$ and $\varepsilon_{f_T} \geq |M_T|/k = |M_T^{(\ell)}|/k$.

The proof technique is a careful generalization of the argument that we alluded to in the beginning of this section, which shows that $\widetilde{O}(1/\varepsilon)$ random points contain a violation with high probability. We show that one can control the size of the maximum matching $M_T^{(\ell)}$ by analyzing the discrepancy of a random subsequence of a sequence of 1s and 0s.

Example 1 (A Two Dimensional Example)**.** Consider the anti-majority function on two dimensions. More precisely, $f : [n]^2 \to \{0,1\}$ defined as $f(x,y) = 1$ if $x + y \leq n$, and $f(x,y) = 0$ otherwise. We describe two maximum cardinality matchings with vastly different stack sizes. The first matching R matches a point (x,y) with $x + y \leq n$ to the point $(n-y+1, n-x+1)$. For an illustration, see the left red matching in Fig. 1 for the case $n = 5$. Observe that whenever $x + y \leq n$, we have $(n-y+1)+(n-x+1) > n$. The second matching B matches a point (x,y) with $x+y \leq n$ to the point $(x+y, n-x+1)$. Again, observe that $(x+y)+(n-x+1) > n$. For an illustration, see the right blue matching in Fig. 1 for the case $n = 5$. Note that the stack sizes for the matching R are large; in particular, they are $n-1, n-2, \ldots, 2, 1$ for $n-1$ stacks and 0 for the rest. On the other hand, any stack in B is of size ≤ 1.

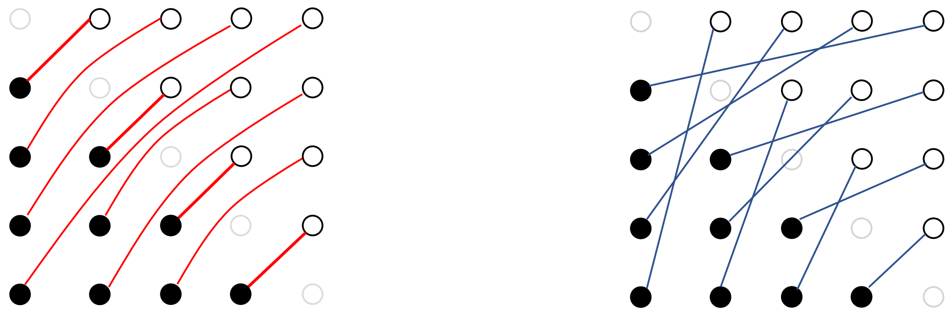

Figure 1: Accompanying illustration for Example 1 showing two different maximum cardinality violation matchings for the anti-majority function $f : [5]^2 \to \{0,1\}$ which have very different stack sizes. Black (white, resp.) circles represent vertices where $f = 1$ ($f = 0$, resp.) and connecting lines represent pairs of the matching. Observe that for the left matching, the bottom line and the right-most slice form a stack of size 4 while the right matching has stack sizes all ≤ 1.

1.4 Related Work

Monotonicity testing has been extensively studied in the past two decades [EKK+00, GGL+00, DGL+99, LR01, FLN+02, HK03, AC06, HK08, ACCL07, Fis04, SS06, Bha08, BCSM12, FR10, BBM12, RRSW11, BGJ+12, CS13, CS14a, CST14, BRY14a, BRY14b, CDST15, CDJS15, KMS15, BB16, CWX17].

We give a short summary of Boolean monotonicity testing over the hypercube. The problem was introduced by Goldreich et al [GGL+00] (refer to Raskhodnikova's thesis [Ras99]), with an $O(d/\varepsilon)$-query tester. The first improvement over that bound was the $\widetilde{O}(d^{7/8})$ tester of Chakrabarty and Seshadhri [CS14a], achieved via a directed analogue of Margulis' isoperimetric theorem. Chen-Servedio-Tan improved the analysis to get an $\widetilde{O}(d^{5/6})$ bound [CST14]. A breakthrough result of Khot-Minzer-Safra gave an $\widetilde{O}(\sqrt{d})$ tester [KMS15]. All these testers are non-adaptive and one-sided. Fischer et al. had proved a (nearly) matching lower bound of $\Omega(\sqrt{d})$ for this case [FLN+02]. The first polynomial two-sided lower bound was given by Chen-Servedio-Tan, subsequently improved to $\Omega(d^{1/2-\delta})$ by Chen et al. [CDST15]. The first polynomial lower bound of $\widetilde{\Omega}(d^{1/4})$ for adaptive testers was given recently by Belovs-Blais [BB16], and was improved to $\widetilde{\Omega}(d^{1/3})$ by Chen-Waingarten-Xie [CWX17].

For Boolean monotonicity testing over general hypergrids, Dodis et al. gave a non-adaptive, one-sided $O((d/\varepsilon)\log^2(d/\varepsilon))$-query tester [DGL+99]. This was improved to $O((d/\varepsilon)\log(d/\varepsilon))$ by Berman-Raskhodnikova-Yaroslavtsev [BRY14a]. They also prove an $\Omega(\log(1/\varepsilon))$ separation between adaptive and non-adaptive monotonicity testers for $f : [n]^2 \to \{0,1\}$. They show an $O(1/\varepsilon)$ adaptive tester (for any constant d), and an $\Omega(\log(1/\varepsilon)/\varepsilon)$ lower bound for non-adaptive monotonicity testers. Previous work by the authors give a monotonicity tester with query complexity $\widetilde{O}(d^{5/6}\log n)$ via directed isoperimetric inequalities for augmented hypergrids [BCS18].

1.5 The Monotonicity Tester: Proof of Theorem 1.1

We use the following theorem of [BCS18] on monotonicity testing for Boolean functions on $[n]^d$.

Theorem 1.7 (Theorem 1.1 of [BCS18]). *Given a function $f : [n]^d \to \{0,1\}$ and a parameter $\varepsilon \in (0,1)$, there is a randomized algorithm that makes $O\left(d^{5/6} \cdot \log^{3/2} d \cdot (\log n + \log d)^{4/3} \cdot \varepsilon^{-4/3}\right)$ non-adaptive* *queries and (a) returns YES with probability* 1 *if the function is monotone, and (b) returns NO with probability* $> 2/3$ *if the function is* ε*-far from being monotone.*

We refer to the tester of Theorem 1.7 as the `grid-path-tester`. Using this result along with our domain reduction theorem (Theorem 1.3), we design the following improved tester. Let C denote the universal constant from Theorem 1.3.

Algorithm 1 Improved Monotonicity Tester for $f : [n]^d \to \{0,1\}$ (f, ε, n)

1: **if** $n \leq (2C \cdot \frac{d}{\varepsilon})^7$: **return** `grid-path-tester`$(f, \varepsilon, n)$.
2: **else**:
3: **repeat** $256/\varepsilon$ times:
4: Sample $\boldsymbol{T} = T_1 \times \cdots \times T_d$ as in Theorem 1.3 with $k = (2C \cdot \frac{d}{\varepsilon})^7$.
5: **if** `grid-path-tester`$(f_{\boldsymbol{T}}, \varepsilon/4, k)$ returns NO, **then return** NO.
6: **return** YES.

Proof of Theorem 1.1: The result is a corollary of Theorem 1.3 and Theorem 1.7. If $n \leq \left(2C \cdot \frac{d}{\varepsilon}\right)^7$, then the tester of Theorem 1.7 (`grid-path-tester`(f, ε, n)) already achieves the stated guarantees. On the other hand if $n > \left(2C \cdot \frac{d}{\varepsilon}\right)^7$, then we set $k := \left(2C \cdot \frac{d}{\varepsilon}\right)^7$ and sample a sub-hypergrid $\boldsymbol{T} := T_1 \times \cdots \times T_d$, where each T_i is formed by taking k i.i.d. draws from the uniform distribution on $[n]$. By Theorem 1.3, $\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon - \frac{C \cdot d}{k^{1/7}} = \varepsilon/2$. Thus, by Markov's inequality, $\mathbf{Pr}_{\boldsymbol{T}}\left(\varepsilon_{f_{\boldsymbol{T}}} \geq \varepsilon/4\right) \geq \varepsilon/4$. Thus, *at least one* of the iterations of Step 4 in Alg. 1 returns $\boldsymbol{T}$ satisfying $\varepsilon_{f_{\boldsymbol{T}}} \geq \varepsilon/4$ with probability $\geq 1 - (1-\varepsilon/4)^{256/\varepsilon} = 1 - \left((1-\varepsilon/4)^{4/\varepsilon}\right)^4 \geq 1 - (1/e)^4 \geq 15/16$.

Thus, if f is ε-far from monotone, then Alg. 1 returns NO with probability $\geq \frac{15}{16} \cdot \frac{2}{3} = 5/8$. On the other hand, if f is monotone, then Alg. 1 clearly returns YES. For the query complexity, Alg. 1 runs `grid-path-tester` at most $256/\varepsilon$ times with parameters $\varepsilon/4$ and $k = \left(2C \cdot \frac{d}{\varepsilon}\right)^7$. Thus, substituting these values in place of ε and n in the query complexity of Theorem 1.7 and multiplying by $256/\varepsilon$ completes the proof of Theorem 1.1. ■

2 Domain Reduction: Proof of Lemma 1.4

In this section, we use Lemma 1.5 and Lemma 1.6 to prove Lemma 1.4. Recall that $\mathcal{D} := [n] \times \left(\prod_{i=2}^d [n_i]\right)$ and $\mathcal{D}_T := T \times \left(\prod_{i=2}^d [n_i]\right)$ denote the original and reduced domains, respectively. Note that $|\mathcal{D}_T| = \frac{k}{n}|\mathcal{D}|$. Let M be the matching given by Lemma 1.5 with $\lambda := \lceil 36k^{2/7} \rceil$.
By Lemma 1.5, we have $\left|\bigcup_{(\ell,S):|M^{(\ell,S)}| \geq 40k^{2/7}} M^{(\ell,S)}\right| \leq \frac{6}{\sqrt{\lambda}} \cdot |\mathcal{D}| < \frac{|\mathcal{D}|}{k^{1/7}}$. Let

$$\widehat{M} := M \setminus \left(\bigcup_{(\ell,S):|M^{(\ell,S)}| \geq 40k^{2/7}} M^{(\ell,S)} \right)$$

denote the set of pairs in M which do not belong to stacks larger than $40k^{2/7}$; we therefore have

$$\sum_{\ell \in \mathcal{L}} |\widehat{M}^{(\ell)}| = |\widehat{M}| \geq |M| - \frac{|\mathcal{D}|}{k^{1/7}}. \tag{1}$$

In this proof, our goal is to *construct* a matching M_T in the violation graph of f_T whose cardinality is sufficiently large. We measure $\mathbf{E}_T\left[|M_T|\right]$ by summing over all lines in $\mathcal{L}$ and applying Lemma 1.6 to each. Notice that $\widehat{M}$ is a matching in the violation graph of f which satisfies $|\widehat{M}^{(\ell,S)}| \leq 40k^{2/7}$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$. Thus by Lemma 1.6, for any $\ell \in \mathcal{L}$,

$$\mathbf{E}_T\left[|M_T^{(\ell)}|\right] \geq \frac{k}{n} \cdot |\widehat{M}^{(\ell)}| - 3 \cdot (40k^{2/7}) \cdot \sqrt{k \ln k} \geq \frac{k}{n} \cdot |\widehat{M}^{(\ell)}| - 120k^{5/6} \tag{2}$$

where we have used $\sqrt{\ln k} < k^{1/3-2/7}$. Now, using (1) and (2), we can calculate $\mathbf{E}_T\left[|M_T|\right]$. We use the fact that $\{\widehat{M}^{(\ell)}\}_{\ell \in \mathcal{L}}$ is a partition of $\widehat{M}$, apply linearity of expectation and use Lemma 1.6 to measure $\mathbf{E}_T[|M_T^{(\ell)}|]$ for each ℓ. Also note that the number of lines is $|\mathcal{L}| = |\mathcal{D}|/n$.

$$\begin{aligned}
\mathbf{E}_T\left[|M_T|\right] = \mathbf{E}_T\left[\sum_{\ell \in \mathcal{L}} |M_T^{(\ell)}|\right] &= \sum_{\ell \in \mathcal{L}} \mathbf{E}_T\left[|M_T^{(\ell)}|\right] \geq \sum_{\ell \in \mathcal{L}} \left(\frac{k}{n} \cdot |\widehat{M}^{(\ell)}| - 120k^{5/6}\right) \quad \text{(by (2))} \\
&= \left(\frac{k}{n} \cdot \sum_{\ell \in \mathcal{L}} |\widehat{M}^{(\ell)}|\right) - \left(120k^{5/6} \cdot \frac{|\mathcal{D}|}{n}\right) \geq \frac{k}{n} \cdot \left(|M| - \frac{|\mathcal{D}|}{k^{1/7}}\right) - \left(120k^{5/6} \cdot \frac{|\mathcal{D}|}{n}\right) \quad \text{(by (1))} \\
&= \frac{k}{n} \cdot \left(|M| - \frac{|\mathcal{D}|}{k^{1/7}} - \frac{120|\mathcal{D}|}{k^{1/6}}\right) \geq \frac{k}{n} \cdot \left(|M| - \frac{C \cdot |\mathcal{D}|}{k^{1/7}}\right)
\end{aligned} \tag{3}$$

for a constant $C > 0$, since $\frac{1}{k^{1/7}}$ dominates $\frac{1}{k^{1/6}}$. (3) gives the expected cardinality of our matching after sampling. To recover the distance to monotonicity we simply normalize by the size of the domain. Dividing by $|\mathcal{D}_T| = \frac{k}{n}|\mathcal{D}|$, we get $\mathbf{E}_T\left[\varepsilon_{f_T}\right] \geq \frac{|M|}{|\mathcal{D}|} - \frac{C}{k^{1/7}} = \varepsilon_f - \frac{C}{k^{1/7}}$. This completes the proof of Lemma 1.4. ■

3 Stack Bound: Proof of Lemma 1.5

We are given a positive integer $\lambda \geq 2$ and a Boolean function $f : \mathcal{D} \to \{0,1\}$ where $\mathcal{D} = [n] \times \left(\prod_{i=2}^{d} [n_i]\right)$ is a rectangular hypergrid for some $n, n_2, \ldots, n_d \in \mathbb{Z}^+$. Lemma 1.5 asserts there is a maximum cardinality matching M such that $\sum_{(\ell,S):|M^{(\ell,S)}| \geq \lambda} |M^{(\ell,S)}| \leq \frac{6}{\sqrt{\lambda}} \cdot |\mathcal{D}|$.

Given a matching M, we consider the vector (or technically, list) $\Lambda(M)$ indexed by stacks (ℓ, S) with $\Lambda_{\ell,S} := |M^{(\ell,S)}|$, and list these in *non-decreasing* order. Consider the maximum cardinality matching M in the violation graph of f which has the lexicographically largest $\Lambda(M)$. That is, the minimum entry of $\Lambda(M)$ is maximized, and subject to that the second-minimum is maximized and so on. We claim that this matching serves as the matching we want. To prove this, we henceforth fix this matching M and introduce the following notation.

- (Low Stacks) $L := \{(\ell, S) \in \mathcal{L} \times \mathcal{S} : |M^{(\ell,S)}| \leq \lambda - 2\}$.
- (High Stacks) $H := \{(\ell, S) \in \mathcal{L} \times \mathcal{S} : |M^{(\ell,S)}| \geq \lambda\}$.

Let $V(H)$ denote the set of vertices matched by $\bigcup_{(\ell,S) \in H} M^{(\ell,S)}$. Let B (for blue) be the set of points in $V(H)$ with function value 0, and R (for red) be the set of points in $V(H)$ with function value 1. M induces a perfect matching between B and R, and we wish to prove $|B| = |R| \leq \frac{6}{\sqrt{\lambda}} \cdot |\mathcal{D}|$. Indeed, define δ to be such that $|B| = \delta|\mathcal{D}|$. In the remainder of the proof, we will prove that $\delta < \frac{6}{\sqrt{\lambda}}$.

We make a simple observation that for any fixed line ℓ, the number of stacks (ℓ, S) which are *non-low* cannot be "too many".

Claim 3.1. *For any line ℓ, the number of non-low stacks ℓ participates in is at most $\frac{n}{\lambda-1}$.*

Proof. Fix any line ℓ and consider the set $\bigcup_{S:(\ell,S)\notin L}\left\{x_1 : \exists(x,y)\in M^{(\ell,S)}\right\}$. That is, the set of 1-coordinates that are used by some non-low stack involving ℓ. The size of this set can't be bigger than the length of ℓ, which is n. Furthermore, each non-low stack contributes at least $\lambda-1$ *unique* entries to this set. The uniqueness follows since the union $\bigcup_{S:(\ell,S)\notin L} M^{(\ell,S)}$ is a matching. ■

We show that if the number of blue points $|B|$ is large ($> 6|\mathcal{D}|/\sqrt{\lambda}$), then we will find a line participating in more than $n/(\lambda-1)$ non-low stacks. To do so, we need to "find" these non-low stacks. We need some more notation to proceed. For a vertex z, we let ℓ_z (S_z, resp.) denote the unique line (slice, resp.) containing z. For each blue point $y\in B$, we define the following interval

$$\mathcal{I}_y := \{z\in\ell_y : z_1\in[x_1,y_1]\} \quad \text{where} \quad (x,y)\in M$$

Armed with this notation, we can find our non-low stacks. Our next claim, which is the heart of the proof and uses the potential function, shows that for every high stack (ℓ,S), we get a bunch of other "non-low" stacks participating with the line ℓ.

Claim 3.2. *Given $y\in B$, let $x := M^{-1}(y)$ and suppose $(\ell,S)\in H$ is such that $(x,y)\in M^{(\ell,S)}$ (note that this stack, (ℓ,S), exists by definition of B). Then, for any $z\in\mathcal{I}_y\cap B$, $(\ell,S_z)\notin L$.*

Proof. The claim is obviously true if $z=y$, since this implies $S_z=S$ (since $y\in S$) and $(\ell,S)\in H$ by assumption. Therefore, we may assume $z\neq y$, and we also assume, for contradiction's sake, $(\ell,S_z)\in L$. Note that $x\in\ell$ and by definition of $\mathcal{I}_y$, we get $x\prec z\prec y$.

Since $z\in B$, it is matched to some $w\in R$. Note $w\prec z\prec y$. Furthermore, the stack $(\ell_w,S_z)\in H$ (by definition of B). By assumption of the claim, $(\ell,S)\in H$. In particular, $x,w,z,y\in V(H)$. Now consider the new matching N which deletes (x,y) and (w,z) and adds (x,z) and (w,y). Note that the cardinality of M remains the same.

We now show that $\Lambda(N)$ is lexicographically bigger than $\Lambda(M)$. To see this, consider the stacks whose sizes have changed from M to N. There are four of them (since we swap two pairs), namely the stacks $(\ell,S),(\ell_w,S_z),(\ell,S_z)$, and (ℓ_w,S). For brevity's sake, let us denote their sizes in M as $\lambda_1,\lambda_2,\lambda_3$, and λ_4, respectively. In N, their sizes become $\lambda_1-1,\lambda_2-1,\lambda_3+1$, and λ_4+1. Note that $\lambda_3\leq\lambda-2$ and both λ_1 and λ_2 are $\geq\lambda$. In particular, the "new" size of stack (ℓ,S_z) is still smaller than the "new" sizes of stacks (ℓ,S) and (ℓ_w,S_z). That is, the vector $\Lambda(N)$, even without the increase in λ_4, is lexicographically larger than $\Lambda(M)$. Since increasing the smallest coordinate (among some coordinates) increases the lexicographic order, we get a contradiction to the lexicographic maximality of $\Lambda(M)$. ■

The rest of the proof is a (slightly technical) averaging argument to prove that $|B|$ is small. We introduce some more notation to carry this through. For a blue point $y\in B$, let $\rho_y := \frac{|\mathcal{I}_y\cap B|}{|\mathcal{I}_y|}$ denote the fraction of blue points in $\mathcal{I}_y$. For $\alpha\in(0,1)$, we say that $y\in B$ is α-rich if $\rho_y\geq\alpha$. A point $x\in R$ is α-rich if its blue partner $y\in B$ (i.e. $(x,y)\in M$) is α-rich. We also call the pair (x,y) an α-rich pair.

Claim 3.3. *If $|B|=\delta|\mathcal{D}|$, then at least $\delta|\mathcal{D}|/2$ of these points are $\delta/4$-rich.*

Proof. Let $B^{(\mathsf{poor})}\subseteq B$ be the points with $\rho_y<\delta/4$. We show $|B^{(\mathsf{poor})}|\leq\delta|\mathcal{D}|/2$ which would prove the claim. To see this, first observe $B^{(\mathsf{poor})}\subseteq\bigcup_{y\in B^{(\mathsf{poor})}}(\mathcal{I}_y\cap B)$. Now consider the *minimal* subset $B^{(\mathsf{poor})}_{\mathsf{min}}\subseteq B^{(\mathsf{poor})}$ such that $\bigcup_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}}\mathcal{I}_y=\bigcup_{y\in B^{(\mathsf{poor})}}\mathcal{I}_y$. That is, given a collection of intervals we are picking the minimal subset covering the same points. Since these are intervals, we get that no point is contained in more than two intervals $\mathcal{I}_y$ among $y \in B^{(\mathsf{poor})}_{\mathsf{min}}$. In particular, this implies

$$\sum_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}} |\mathcal{I}_y| \le 2 \cdot \left| \bigcup_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}} \mathcal{I}_y \right| \tag{4}$$

Therefore,

$$\begin{aligned} \left|B^{(\mathsf{poor})}\right| &\le \left| \bigcup_{y\in B^{(\mathsf{poor})}} (\mathcal{I}_y \cap B) \right| = \left| \bigcup_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}} (\mathcal{I}_y \cap B) \right| \le \sum_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}} |\mathcal{I}_y \cap B| \\ &< \frac{\delta}{4} \sum_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}} |\mathcal{I}_y| \le \frac{\delta}{2} \cdot \left| \bigcup_{y\in B^{(\mathsf{poor})}_{\mathsf{min}}} \mathcal{I}_y \right| \le \frac{\delta}{2} \cdot |\mathcal{D}| \end{aligned}$$

The first equality follows from the definition of $B^{(\mathsf{poor})}_{\mathsf{min}}$ (taking intersection with B), and the third (strict) inequality follows from the fact that none of these points are $\delta/4$-rich. The fourth inequality is (4). This completes the proof. ■

A corollary of Claim 3.3 is that there are at least $\delta|\mathcal{D}|/2$ red points which are $\delta/4$-rich. In particular, there must exist some line ℓ that contains $\ge \delta n/2$ red points in it which are $\delta/4$-rich. Let this line be ℓ and let $R_\ell \subseteq \ell$ be the set of rich red points. Let B_ℓ be their partners in M. Let $\mathcal{S}^\ell = \left\{S \in \mathcal{S} : \exists z \in S \cap \left(\cup_{y\in B^\ell} \mathcal{I}_y \cap B\right)\right\}$ denote the set of slices containing blue points from the collection of rich intervals, $\{\mathcal{I}_y : y \in B^\ell\}$. By Claim 3.2, we know that all these stacks are non-low, that is, $(\ell, S) \notin L$ for all $S \in \mathcal{S}^\ell$. We now *lower bound* the cardinality of this set.

Consider the set of blue points in our union of rich intervals from B^ℓ, $\bigcup_{y\in B^\ell} \mathcal{I}_y \cap B$. There are precisely n slices in total, and for a vertex $z \in \mathcal{D}$, S_z is the slice indexed by the 1-coordinate of z. Thus, we have $|\mathcal{S}^\ell| = |\{z_1 : z \in \bigcup_{y\in B^\ell} \mathcal{I}_y \cap B\}|$. That is, $|\mathcal{S}^\ell|$ is exactly the number of unique 1-coordinates among vertices in $\bigcup_{y\in B^\ell} \mathcal{I}_y \cap B$.

Since we care about the number of unique 1-coordinates, we consider the "projections" of our sets of interest onto dimension 1. For a set $X \subseteq \mathcal{D}$, let $\mathsf{proj}_1(X) := \{x_1 : x \in X\}$ be the set of 1-coordinates used by points in X. In particular, note that for $y \in B$, $\mathsf{proj}_1(\mathcal{I}_y) := [x_1, y_1] \subset [n]$, where $x := M^{-1}(y)$ and observe that $|\mathcal{S}^\ell| = \left|\bigcup_{y\in B^\ell} \mathsf{proj}_1(\mathcal{I}_y \cap B)\right|$. Now, given that each interval from $\{\mathcal{I}_y\}_{y\in B^\ell}$ is a $\frac{\delta}{4}$-fraction blue, the following claim says that at least a $\frac{\delta}{8}$-fraction of the union of intervals consists of blue points with unique 1-coordinates.

Claim 3.4. $\left|\bigcup_{y\in B^\ell} \mathsf{proj}_1(\mathcal{I}_y \cap B)\right| \ge \frac{\delta}{8} \left|\bigcup_{y\in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)\right|$

Proof. As in the proof of Claim 3.2, let $B^\ell_{\mathsf{min}} \subseteq B^\ell$ be a minimal cardinality subset of B^ℓ such that $\bigcup_{y\in B^\ell_{\mathsf{min}}} \mathsf{proj}_1(\mathcal{I}_y) = \bigcup_{y\in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)$. For any $y \in B$, y belongs to *at most* two intervals from B^ℓ_{min}.

$$\left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y \cap B)\right| = \left|\bigcup_{y \in B^\ell_{\mathsf{min}}} \mathsf{proj}_1(\mathcal{I}_y \cap B)\right| \geq \frac{1}{2} \sum_{y \in B^\ell_{\mathsf{min}}} |\mathsf{proj}_1(\mathcal{I}_y \cap B)|$$
$$\geq \frac{\delta}{8} \sum_{y \in B^\ell_{\mathsf{min}}} |\mathsf{proj}_1(\mathcal{I}_y)| \geq \frac{\delta}{8} \left|\bigcup_{y \in B^\ell_{\mathsf{min}}} \mathsf{proj}_1(\mathcal{I}_y)\right| = \frac{\delta}{8} \left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)\right|. \quad \blacksquare$$

Now importantly, $|\mathsf{proj}_1(R^\ell)| = |R^\ell| \geq \frac{\delta}{2} \cdot n$ since the 1-coordinates of elements of R^ℓ are distinct (since R^ℓ is contained on a single line). Moreover, by definition of $\mathcal{I}_y$, $\mathsf{proj}_1(R^\ell) \subseteq \bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)$ and so $\left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)\right| \geq |\mathsf{proj}_1(R^\ell)| \geq \frac{\delta}{2} \cdot n$. Finally, combining this with Claim 3.4, we get

$$|\mathcal{S}^\ell| = \left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y \cap B)\right| \geq \frac{\delta}{8} \left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)\right| \geq \frac{\delta^2}{16} \cdot n$$

Therefore, ℓ participates in at least $\frac{\delta^2}{16} \cdot n$ non-low stacks. Thus, using Claim 3.1, if $\frac{\delta^2}{16} \cdot n > \frac{n}{\lambda - 1} \iff \delta > \frac{4}{\sqrt{\lambda - 1}}$, then we have a contradiction. Since $\lambda \geq 2$, we conclude that $\delta < 6/\sqrt{\lambda}$. This concludes the proof of Lemma 1.5. $\blacksquare$

4 Line Sampling: Proof of Lemma 1.6

We recall the lemma for ease of reading. Given a line $\ell \in \mathcal{L}$, we have defined $M^{(\ell)} := \{(x, y) \in M : x \in \ell\}$. Given a stack S, we have defined $M^{(\ell,S)} := \{(x, y) \in M^{(\ell)} : y \in S\}$. Given a multi-set $T \subseteq [n]$, recall $M_T^{(\ell)}$ is the maximum cardinality matching of violations (x, y) such that (a) x and y are both matched by $M^{(\ell)}$, and (b) x_1 and y_1 both lie in T. Given $\lambda \in \mathbb{Z}^+$ such that $|M^{(\ell,S)}| \leq \lambda$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$, the line sampling lemma (Lemma 1.6) states

$$\mathbf{E}_T\left[|M_T^{(\ell)}|\right] \geq \frac{k}{n} \cdot |M^{(\ell)}| - 3\lambda\sqrt{k \ln k}. \tag{5}$$

Consider an arbitrary, fixed line $\ell \in \mathcal{L}$. We use the matching $M^{(\ell)}$ to induce weights $w^+(i), w^-(i)$ on $[n]$ as follows. Initially $w^+(i), w^-(i) = 0$ for all $i \in [n]$. For each $(x, y) \in M^{(\ell)}$ if $x \in S_i$ then we increase $w^+(i)$ by 1, and if $y \in S_j$ then we increase $w^-(j)$ by 1. We let $V^+ := \{i : w^+(i) > 0\}$ and $V^- := \{j : w^-(j) > 0\}$.

Claim 4.1. *We make a few observations.*

1. *For any $i \in [n]$, $w^+(i) \leq 1$.*
2. *For any $i \in [n]$, $w^-(i) \leq \lambda$.*
3. *For any $t \in [n]$, $\sum_{s \leq t}(w^-(s) - w^+(s)) \leq 0$.*

Proof. The first observation follows since the lower endpoints of $M^{(\ell)}$ all lie on ℓ, and thus have distinct 1-coordinates. The second observation follows from the assumption that $|M^{(\ell,S)}| \leq \lambda$ for all $(\ell, S) \in \mathcal{L} \times \mathcal{S}$. The third observation follows by noting that whenever $w^-(j)$ is increased for some j, we also increase $w^+(i)$ for some $i < j$. $\blacksquare$

Given a multiset $T \subseteq [n]$, denote $V_T^+ := V^+ \cap T$ and $V_T^- := V^- \cap T$. Also, define the bipartite graph $G_T := (V_T^+, V_T^-, E_T)$ where $(i,j) \in E_T$ iff $i \leq j$. A w-matching A in G_T is a subset of edges of E_T such that every vertex $i \in V_T^+$ has at most $w^+(i)$ edges of A incident on it, and every vertex $j \in V_T^-$ has at most $w^-(j)$ edges of A incident on it. Let $\nu(G_T)$ denote the size of the largest w-matching in G_T.

Lemma 4.2. *For any multiset $T \subseteq [n]$ and any w-matching $A \subseteq E_T$ in G_T, we have $|M_T^{(\ell)}| \geq |A|$. In particular, $\mathbf{E}_T\left[|M_T^{(\ell)}|\right] \geq \mathbf{E}_T\left[\nu(G_T)\right]$.*

Proof. Consider any w-matching $A \subseteq E_T$. For any vertex $i \in V_T^+$, there are at most $w^+(i)$ edges in A incident on it. Each increase of $w^+(i)$ is due to an edge $(x,y) \in M^{(\ell)}$ where $x_1 = i$. Thus, we can charge each of these edges of A (arbitrarily, but uniquely) to $w^+(i)$ different $x \in \ell$. Similarly, for any vertex $j \in V_T^-$, there are at most $w^-(j)$ edges in A incident on it. Each increase of $w^-(j)$ is due to an edge $(x,y) \in M^{(\ell)}$ with $y_1 = j$. Thus, we can charge each of these edges of A (arbitrarily, but uniquely) to $w^-(j)$ different $y \in S_j$, the jth slice. Furthermore, any $z \in \ell$ with $z_1 \leq j$ satisfies $z \prec y$. In sum, each $(i,j) \in A$ can be uniquely charged to an $x \in \ell$ with $x_1 = i$ and $y \in S_j$ such that (a) (x,y) forms a violation, (b) x, y were matched in $M^{(\ell)}$, and (c) $x_1, y_1 \in T$. Therefore, $|M_T^{(\ell)}| \geq |A|$ since the LHS is the maximum cardinality matching. ∎

Lemma 4.3. *For any $T \subseteq [n]$, we have*

$$\nu(G_T) = \sum_{j \in T} w^-(j) - \max_{t \in T} \sum_{s \in T : s \leq t} \big(w^-(s) - w^+(s)\big)$$

Proof. By Hall's theorem, the maximum w-matching in G_T is given by the total weight on the V_T^- side, that is, $\sum_{j \in T} w^-(j)$, minus the total *deficit* $\delta(T) := \max_{S \subseteq V_T^-} \left(\sum_{s \in S} w^-(s) - \sum_{s \in \Gamma_T(S)} w^+(s)\right)$ where for $S \subseteq V_T^-$, $\Gamma_T(S) \subseteq V_T^+$ is the neighborhood of S in G_T. Consider such a maximizer S, and let t be the largest index present in S. Then note that $\sum_{s \in \Gamma_T(S)} w^+(s)$ is precisely $\sum_{s \in T : s \leq t} w^+(s)$. Furthermore note that adding any $s \leq t$ from V_T^- won't increase $|\Gamma_T(S)|$. Thus, given that the largest index present in S is t, we get that $\delta(T)$ is precisely the summation in the second term of the RHS. $\delta(T)$ is maximized by choosing the t which maximizes the summation. ∎

Next, we bound the expectation of the RHS in Lemma 4.3. Recall that $T := \{s_1, \ldots, s_k\}$ is a multiset where each s_i is u.a.r. picked from $[n]$. For the first term, we have

$$\mathbf{E}_T\left[\sum_{j \in T} w^-(j)\right] = \sum_{i=1}^{k}\sum_{j=1}^{n} \mathbf{Pr}[s_i = j] \cdot w^-(j) = \frac{k}{n} \cdot \sum_{j=1}^{n} w^-(j) = \frac{k}{n} \cdot |M^{(\ell)}|. \tag{6}$$

The second-last equality follows since s_i is u.a.r in $[n]$ and the last equality follows since $\sum_j w^-(j)$ increases by exactly one for each edge in $M^{(\ell)}$. Next we upper bound the expectation of the second term. For a fixed t, define

$$Z_t := \sum_{s \in T : s \leq t} (w^-(s) - w^+(s)) = \sum_{i=1}^{k} X_{i,t} \text{ where } X_{i,t} = \begin{cases} w^-(s_i) - w^+(s_i) & \text{if } s_i \leq t \\ 0 & \text{otherwise} \end{cases}$$

Note that the $X_{i,t}$'s are i.i.d random variables with $X_{i,t} \in [-1, \lambda]$ with probability 1. Thus, applying Hoeffding's inequality we get

$$\mathbf{Pr}\left[Z_t > \mathbf{E}[Z_t] + a\right] \le 2\exp\left(\frac{-a^2}{2k\lambda^2}\right)$$

Now we use Claim 4.1, part (3) to deduce that

$$\mathbf{E}[Z_t] = \sum_{i=1}^{k} \mathbf{E}[X_{i,t}] = \sum_{i=1}^{k}\sum_{s\le t}(w^-(s) - w^+(s)) \cdot \mathbf{Pr}[s_i = s] \le 0$$

since $\mathbf{Pr}[s_i = s] = 1/n$. Therefore, the RHS of the Hoeffding bound is an upper-bound on $\mathbf{Pr}[Z_t \ge a]$. In particular, invoking $a := 2\lambda\sqrt{k\ln k}$ and applying a union bound, we get

$$\mathbf{Pr}\left[\max_{t\in T} Z_t > 2\lambda\sqrt{k\ln k}\right] = \mathbf{Pr}\left[\exists t \in T: \quad Z_t > 2\lambda\sqrt{k\ln k}\right] \le k \cdot e^{-2\ln k} = 1/k \tag{7}$$

and since $\max_{t\in T} Z_t$ is trivially upper-bounded by λk, this implies that

$$\mathbf{E}_T\left[\max_{t\in T}\sum_{s\in T: s\le t} w^-(s) - w^+(s)\right] \le \lambda k \cdot \mathbf{Pr}\left[\max_{t\in T} Z_t > a\right] + a \le \lambda + a \le 3\lambda\sqrt{k\ln k}. \tag{8}$$

Lemma 1.6 follows from Lemma 4.2, Lemma 4.3, (6), and (8).

5 Lower Bound for Domain Reduction

In this section we prove the following lower bound for the number of samples needed for a domain reduction result to hold for distance to monotonicity. Recall the domain reduction experiment: given $f : [n]^d \to \{0,1\}$ and an integer $k \in \mathbb{Z}^+$, we choose $\boldsymbol{T} := T_1 \times \cdots \times T_d$ where for each $i \in [d]$, T_i is formed by taking k i.i.d. uniform draws from $[n]$. We then consider the restriction $f_{\boldsymbol{T}}$.

Theorem 5.1 (Lower Bound for Domain Reduction)**.** *There exists a function $f : [n]^d \to \{0,1\}$ with distance to monotonicity $\varepsilon_f = \Omega(1)$, for which $\mathbf{E}_{\boldsymbol{T}}[\varepsilon_{f_{\boldsymbol{T}}}] \le O(k^2/d)$.*

In particular, the above theorem implies that $k = \Omega(\sqrt{d})$ samples in each dimension is necessary to preserve distance to monotonicity.

5.1 Proof of Theorem 5.1

We define the function $\mathsf{Centrist} : [0,1]^d \to \{0,1\}$. The continuous domain is just a matter of convenience; any n that is a multiple of d would suffice. It is easiest to think of d individuals voting for an outcome, where the ith vote x_i is the "strength" of the vote. Based on their vote, an individual is labeled as follows.

- $x_i \in [0, 1 - 2/d]$: skeptic
- $x_i \in (1 - 2/d, 1 - 1/d]$: supporter
- $x_i \in (1 - 1/d, 1]$: fanatic

$\mathsf{Centrist}(x) = 1$ *iff* there exists some individual who is a supporter. The non-monotonicity is created by fanaticism. If a unique supporter increases her vote to become a fanatic, the function value can decrease.

Claim 5.2. *The distance to monotonicity of* $\mathsf{Centrist}$ *is* $\Omega(1)$.

Proof. It is convenient to talk in terms of probability over the uniform distribution in $[0,1]^d$. Define the following events, for $i \in [d]$.

- $\mathcal{S}_i$: The ith individual is a supporter, and all others are skeptics.
- $\mathcal{F}_i$: The ith individual is a fanatic, and all others are skeptics.

Observe that all these events are disjoint. Also, $\mathbf{Pr}[\mathcal{S}_i] = \mathbf{Pr}[\mathcal{F}_i] = (1/d)(1-2/d)^{d-1} = \Omega(1/d)$. Note that $\forall x \in \mathcal{S}_i$, $\mathsf{Centrist}(x) = 1$ and $\forall x \in \mathcal{F}_i$, $\mathsf{Centrist}(x) = 0$.

We construct a violation matching $M : \bigcup_i \mathcal{S}_i \to \bigcup_i \mathcal{F}_i$. For $x \in \mathcal{S}_i$, $M(x) = x + e_i/d$, where e_i is the unit vector in dimension i. For $x \in \mathcal{S}_i$, $x_i \in (1-2/d, 1-1/d]$, so $M(x)_i \in (1-1/d, 1]$, and $M(x) \in \mathcal{F}_i$. M is a bijection between $\mathcal{S}_i$ and $\mathcal{F}_i$, and all the $\mathcal{S}_i, \mathcal{F}_i$ sets are disjoint. Thus, M is a violation matching. Since $\mathbf{Pr}\left[\bigcup_i \mathcal{S}_i\right] = \Omega(d \cdot 1/d)$, the distance to monotonicity is $\Omega(1)$. ■

Lemma 5.3. *Let* $k \in \mathbb{Z}^+$ *be any positive integer. If* $\boldsymbol{T} := T_1 \times \cdots \times T_d$ *is a randomly chosen hypergrid, where for each* $i \in [d]$, T_i *is a set formed by taking* k *i.i.d. samples from the uniform distribution on* $[0,1]$*, then with probability* $> 1 - 4k^2/d$, $\mathsf{Centrist}_{\boldsymbol{T}}$ *is a monotone function.*

Proof. Each T_i consists of k u.a.r. elements in $[0,1]$. We can think of each as a sampling of the ith individual's vote. For a fixed i, let us upper bound the probability that T_i contains strictly more than one non-skeptic vote. This probability is

$$\begin{aligned} 1-(1-2/d)^k - k(1-2/d)^{k-1}(2/d) &= 1-(1-2/d)^{k-1}(1-2/d+2k/d) \\ &\leq 1 - \left(1 - \frac{2(k-1)}{d}\right)\left(1 + \frac{2(k-1)}{d}\right) \leq 4k^2/d^2 \end{aligned}$$

where we have used the bound $(1-x)^r \geq 1 - xr$, for any $x \in [0,1]$ and $r \geq 1$. By the union bound over all dimensions, with probability $> 1-4k^2/d$, all T_i's contain at most one non-skeptic vote. Consider $\mathsf{Centrist}_{\boldsymbol{T}}$, some $x \in \boldsymbol{T}$, and a dimension $i \in [d]$. If the ith individual increases her vote (from x), there are three possibilities.

- The vote does not change. Then the function value does not change.
- The vote goes from a skeptic to a supporter. The function value can possibly increase, but not decrease.
- The vote goes from a skeptic to a fanatic. If $\mathsf{Centrist}_{\boldsymbol{T}}(x) = 1$, there must exist some $j \neq i$ that is a supporter. Thus, the function value remains 1 regardless of i's vote.

In no case does the function value decrease. Thus, $\mathsf{Centrist}_{\boldsymbol{T}}$ is monotone. ■

Theorem 5.1 follows from Claim 5.2 and Lemma 5.3.

References

[AC06] Nir Ailon and Bernard Chazelle. Information theory in property testing and monotonicity testing in higher dimension. *Information and Computation*, 204(11):1704–1717, 2006.

[ACCL07] Nir Ailon, Bernard Chazelle, Seshadhri Comandur, and Ding Liu. Estimating the distance to a monotone function. *Random Structures Algorithms*, 31(3):371–383, 2007.

[BB16] Aleksandrs Belovs and Eric Blais. A polynomial lower bound for testing monotonicity. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2016.

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012.

[BCS18] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. A $o(d) \cdot \text{polylog } n$ monotonicity tester for boolean functions over the hypergrid $[n]^d$. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018.

[BCSM12] Jop Briët, Sourav Chakraborty, David García Soriano, and Ari Matsliah. Monotonicity testing and shortest-path routing on the cube. *Combinatorica*, 32(1):35–53, 2012.

[BGJ$^+$12] Arnab Bhattacharyya, Elena Grigorescu, Madhav Jha, Kyoming Jung, Sofya Raskhodnikova, and David Woodruff. Lower bounds for local monotonicity reconstruction from transitive-closure spanners. *SIAM Journal on Discrete Mathematics (SIDMA)*, 26(2):618–646, 2012.

[Bha08] Arnab Bhattacharyya. A note on the distance to monotonicity of boolean functions. Technical Report 012, Electronic Colloquium on Computational Complexity (ECCC), 2008.

[BRY14a] Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. $\mathrm{L_p}$-testing. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2014.

[BRY14b] Eric Blais, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Lower bounds for testing properties of functions over hypergrid domains. In *Proceedings, IEEE Conference on Computational Complexity (CCC)*, 2014.

[CDJS15] Deeparnab Chakrabarty, Kashyap Dixit, Madhav Jha, and C. Seshadhri. Property testing on product distributions: Optimal testers for bounded derivative properties. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2015.

[CDST15] Xi Chen, Anindya De, Rocco A. Servedio, and Li-Yang Tan. Boolean function monotonicity testing requires (almost) $O(n^{1/2})$ non-adaptive queries. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2015.

[CS13] Deeparnab Chakrabarty and C. Seshadhri. Optimal bounds for monotonicity and Lipschitz testing over hypercubes and hypergrids. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2013.

[CS14a] Deeparnab Chakrabarty and C. Seshadhri. An o(n) monotonicity tester for boolean functions over the hypercube. *SIAM Journal on Computing (SICOMP)*, 45(2):461–472, 2014.

[CS14b] Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory of Computing*, 10:453–464, 2014.

[CST14] Xi Chen, Rocco A. Servedio, and Li-Yang. Tan. New algorithms and lower bounds for monotonicity testing. In *Proceedings, IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.

[CWX17] Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond talagrand: New lower bounds for testing monotonicity and unateness. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2017.

[DGL+99] Yevgeny Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 1999.

[EKK+00] Funda Ergun, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. *J. Comput. System Sci.*, 60(3):717–751, 2000.

[Fis04] Eldar Fischer. On the strength of comparisons in property testing. *Information and Computation*, 189(1):107–116, 2004.

[FLN+02] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, and Ronitt Rubinfeld. Monotonicity testing over general poset domains. *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2002.

[FR10] Shahar Fattal and Dana Ron. Approximating the distance to monotonicity in high dimensions. *ACM Trans. on Algorithms (TALG)*, 6(3), 2010.

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samordinsky. Testing monotonicity. *Combinatorica*, 20:301–337, 2000.

[HK03] Shirley Halevy and Eyal Kushilevitz. Distribution-free property testing. *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2003.

[HK08] Shirley Halevy and Eyal Kushilevitz. Testing monotonicity over graph products. *Random Structures Algorithms*, 33(1):44–67, 2008.

[KMS15] Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric type theorems. In *Proceedings, IEEE Symposium on Foundations of Computer Science (FOCS)*, 2015.

[LR01] Eric Lehman and Dana Ron. On disjoint chains of subsets. *Journal of Combinatorial Theory, Series A*, 94(2):399–404, 2001.

[Ras99] Sofya Raskhodnikova. Monotonicity testing. *Masters Thesis, MIT*, 1999.

[RRSW11] Dana Ron, Ronitt Rubinfeld, Muli Safra, and Omri Weinstein. Approximating the Influence of Monotone Boolean Functions in $O(\sqrt{n})$ Query Complexity. In *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2011.

[SS06] Mike Saks and C. Seshadhri. Parallel monotonicity reconstruction. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2006.

A Domain Reduction for Variance

In this section, we prove that, given $f : [n]^d \to \{0, 1\}$, restricting f to a random hypercube (domain reduction with $k = 2$) suffices to preserve the *variance* of f.

Theorem A.1 (Domain Reduction for Variance)**.** *Let $f : [n]^d \to \{0,1\}$ be any function. If $\boldsymbol{T} := T_1 \times \cdots \times T_d$ is a randomly chosen sub-hypercube, where for each $i \in [d]$, T_i is a (multi)-set formed by taking 2 i.i.d. samples from the uniform distribution on $[n]$, then $\mathbf{E}_{\boldsymbol{T}}[var(f_{\boldsymbol{T}})] \geq var(f)/2$.*

Proof. We will interpret f as a Boolean function with $d \log n$ (Boolean) inputs, so $f : \{-1,1\}^{d \log n} \to \{-1,1\}$. We will index the inputs in $[d \log n]$, where the interval $I_i := [(i-1)\log n + 1, i \log n]$ (the ith block) corresponds to the ith input in the original representation. Henceforth, i will always index a block (and thereby, an input in the original representation). We use x_j to denote the jth input bit.

Let us think of the restriction in Boolean terms. Note that $f_{\boldsymbol{T}} : \{-1,1\}^d \to \{-1,1\}$, and we use y to denote an input to the restriction. In Boolean terms, T_i picks two u.a.r $\log n$ bit strings, and forces the ith block of inputs, I_i, to be one of these. The choice between these is decided by y_i. Let us think of T_i as follows. For every $j \in I_i$, it adds it to a set R_i with probability $1/2$. All the inputs in R_i will be fixed, while the inputs in $I_i \setminus R_i$ are alive (but correlated by y_i). Then, for every $j \in I_i$, it picks a u.a.r bit b_j. (Call this string B_i.) This is interpreted as follows. For every $j \in R_i$, x_j is fixed to b_j. For every $j \in I_i \setminus R_i$, x_j is set to $y_i b_j$. The randomness of T_i can therefore be represented as independently choosing R_i and B_i.

Consider some non-empty $S \subseteq I_i$.

$$\prod_{j \in S} x_j = \prod_{j \in S \cap R_i} b_j \prod_{j \in S \setminus R_i} b_j y_i = y_i^{|S \setminus R_i|} \prod_{j \in S} b_j \tag{9}$$

The expected value of the Fourier basis function is (as expected) zero. Recall that S is non-empty.

$$\mathbf{E}_{T_i}\left[\mathbf{E}_y\left[\prod_{j \in S} x_j\right]\right] = \mathbf{E}_{R_i,B_i}\left[\mathbf{E}_y\left[y_i^{|S\setminus R_i|}\prod_{j \in S} b_j\right]\right] = \mathbf{E}_{R_i}\left[\mathbf{E}_y\left[y_i^{|S\setminus R_i|}\right]\right] \cdot \mathbf{E}_{B_i}\left[\prod_{j \in S} b_j\right] = 0 \tag{10}$$

If $|S \setminus R_i|$ is even, then $\prod_{j \in S} x_j$ is *independent* of y. Then, $\mathbf{E}_y\left[\prod_{j\in S} x_j\right]^2 = 1$. If $|S \setminus R_i|$ is odd, then $\prod_{j \in S} x_j$ is linear in y_i and $\mathbf{E}_y\left[\prod_{j \in S} x_j\right] = 0$.

$$\mathbf{E}_{T_i}\left[\mathbf{E}_y\left[\prod_{j \in S} x_j\right]^2\right] = \mathbf{Pr}_{R_i}\left[|S \setminus R_i| \text{ is even}\right] = 1/2 \tag{11}$$

Let us write out the Fourier expansion of f.

$$f(x) = \sum_{S \in [d \log n]} \widehat{f}_S \cdot \chi_S(x) = \sum_{\substack{S = S_1 \cup \ldots \cup S_d \\ \forall i, S_i \subseteq I_i}} \widehat{f}(S) \prod_{i \in [d]} \prod_{x_j \in S_i} x_j$$

Let us write an expression for the square of the zeroth Fourier coefficient of the restriction.

$$\mathbf{E}_{\boldsymbol{T}}\left[\widehat{f_{\boldsymbol{T}}}(\emptyset)^2\right] = \mathbf{E}_{\boldsymbol{T}}\left[\left(\sum_{S \in [d \log n]} \widehat{f}_S \mathbf{E}_y[\chi_S(x)]\right)^2\right]$$

We stress that the choice of x inside the expectations depend on y (or y') in the manner described before

(10).

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{T}}\left[\widehat{f_{\boldsymbol{T}}}(\emptyset)^2\right] &= \mathbf{E}_{\boldsymbol{T}}\left[\sum_S \widehat{f}_S^2 \mathbf{E}_y\left[\chi_S(x)\right]^2 + \sum_{S,T:S\neq T} \widehat{f}_S \widehat{f}_T \mathbf{E}_y\left[\chi_S(x)\right]\mathbf{E}_y\left[\chi_T(x)\right]\right] && (12)\\
&= \sum_S \widehat{f}_S^2 \mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y[\chi_S(x)]^2\right] + \sum_{S,T:S\neq T} \widehat{f}_S \widehat{f}_T \mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y[\chi_S(x)]\mathbf{E}_y[\chi_T(x)]\right] && (13)
\end{aligned}$$

We will write $S = S_{i_1} \cup S_{i_2} \cdots \cup S_{i_k}$, where all S_{i_r}s are non-empty. We deal with the first term, using (11).

$$\mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y[\chi_S(x)]^2\right] = \mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y\left[\prod_{\ell\leq k}\prod_{j\in S_{i_\ell}} x_j\right]^2\right] = \prod_{\ell\leq k}\mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y\left[\prod_{j\in S_{i_\ell}} x_j\right]^2\right] = 1/2^k$$

The cross terms will be zero, using calculations analogous for (10) (which is not directly used). We write $S = S_1 \cup \cdots \cup S_d$, where some of these may be empty.

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{T}}[\mathbf{E}_y[\chi_S(x)]\mathbf{E}_y[\chi_T(x)]] &= \mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y\left[\prod_{i\in[d]}\prod_{j\in S_i} x_j\right]\mathbf{E}_y\left[\prod_{i\in[d]}\prod_{j\in T_i} x_j\right]\right]\\
&= \prod_{i\in[d]}\mathbf{E}_{R_i,B_i}\left[\mathbf{E}_{y_i}\left[y_i^{|S_i\setminus R_i|}\prod_{j\in S_i} b_j\right]\mathbf{E}_{y_i}\left[y_i^{|T_i\setminus R_i|}\prod_{j\in T_i} b_j\right]\right]\\
&= \prod_{i\in[d]}\mathbf{E}_{R_i}\left[\mathbf{E}_{y_i}\left[y_i^{|S_i\setminus R_i|}\right]\mathbf{E}_{y_i}\left[y_i^{|T_i\setminus R_i|}\right]\mathbf{E}_{B_i}\left[\prod_{j\in S_i\Delta T_i} b_j\right]\right]
\end{aligned}$$

There must exist some i such that $S_i\Delta T_i \neq \emptyset$. For that i, $\mathbf{E}_{B_i}\left[\prod_{j\in S_i\Delta T_i} b_j\right] = 0$, and thus for $S \neq T$, $\mathbf{E}_{\boldsymbol{T}}[\mathbf{E}_y[\chi_S(x)]\mathbf{E}_y[\chi_T(x)]] = 0$. Plugging these bounds in,

$$\mathbf{E}_{\boldsymbol{T}}\left[\widehat{f_{\boldsymbol{T}}}(\emptyset)^2\right] \leq \widehat{f}(\emptyset)^2 + \sum_{S\neq\emptyset}\widehat{f}(S)^2/2 = 1 - \mathrm{var}(f) + \mathrm{var}(f)/2 = 1 - \mathrm{var}(f)/2$$

We rearrange to get $\mathbf{E}_{\boldsymbol{T}}[\mathrm{var}(f)] = \mathbf{E}_{\boldsymbol{T}}\left[1 - \widehat{f_{\boldsymbol{T}}}(\emptyset)^2\right] \geq \mathrm{var}(f)/2$. ■

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il