Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 187 (2018)

# Domain Reduction for Monotonicity Testing: A $o(d)$ Tester for Boolean Functions in $d$-Dimensions

Hadley Black* Deeparnab Chakrabarty† C. Seshadhri‡

**Abstract**

We describe a $\widetilde{O}(d^{5/6})$-query monotonicity tester for Boolean functions $f : [n]^d \to \{0,1\}$ on the $n$-hypergrid. This is the first $o(d)$ monotonicity tester with query complexity independent of $n$. Motivated by this independence of $n$, we initiate the study of monotonicity testing of measurable Boolean functions $f : \mathbb{R}^d \to \{0,1\}$ over the continuous domain, where the distance is measured with respect to a product distribution over $\mathbb{R}^d$. We give a $\widetilde{O}(d^{5/6})$-query monotonicity tester for such functions.

Our main technical result is a *domain reduction theorem* for monotonicity. For any function $f$ : $[n]^d \to \{0,1\}$, let $\varepsilon_f$ 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 $k = \text{poly}(d/\varepsilon)$, the expected distance of the restriction is $\mathbf{E}[\varepsilon_{\hat{f}}] = \Omega(\varepsilon_f)$. Previously, such a result was only known for $d = 1$ (Berman-Raskhodnikova-Yaroslavtsev, STOC 2014). Our result for testing Boolean functions over $[n]^d$ then follows by applying the $d^{5/6} \cdot \text{poly}(1/\varepsilon, \log n, \log d)$-query hypergrid tester of Black-Chakrabarty-Seshadhri (SODA 2018).

To obtain the result for testing Boolean functions over $\mathbb{R}^d$, we use standard measure theoretic tools to reduce monotonicity testing of a measurable function $f$ to monotonicity testing of a discretized version of $f$ over a hypergrid domain $[N]^d$ for large, but finite, $N$ (that may depend on $f$). The independence of $N$ in the hypergrid tester is crucial to getting the final tester over $\mathbb{R}^d$.

*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 is a fundamental problem in property testing. Let $(D, \prec)$ be a partially ordered set (poset) and let $R$ be a total order. A function $f : D \to R$ is monotone if $f(x) \le f(y)$ whenever $x \prec y$. The hypercube, $\{0,1\}^d$ and the hypergrid $[n]^d$ have been the most studied posets in monotonicity testing, where $\prec$ denotes the coordinate-wise partial ordering. The distance between two functions $f$ and $g$ is $\texttt{dist}(f,g) := \mathbf{Pr}_{x\sim D}[f(x) \neq g(x)]$ where $x$ is drawn uniformly from the domain. The distance of $f$ to monotonicity, denoted as $\varepsilon_f$, is its distance to the nearest monotone function. That is, $\varepsilon_f := \min_{g\in\mathcal{M}} \texttt{dist}(f,g)$, where $\mathcal{M}$ 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] give 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], that is, linear in $d$.

The problem of monotonicity testing Boolean functions $f : [n]^d \to \{0,1\}$ over hypergrids is not as well understood. Dodis et al. [DGL+99] (with improvements by Berman, Raskhodnikova, and Yaroslavtsev [BRY14a], henceforth BRY) give a $\widetilde{O}(d/\varepsilon)$-query tester. The important feature to note is the *independence* of $n$. Contrast this, again, with the real-valued case; monotonicity testing of functions $f : [n] \to \mathbb{R}$ requires $\Omega(\log n)$ queries [EKK+00, Fis04]. Recently, the authors [BCS18] describe a $\widetilde{O}(d^{5/6} \log^{4/3} n\, \varepsilon^{-4/3})$-query tester. Although the dependence on $d$ is sublinear, there is a dependence on $n$. The following question has 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 $d$?* One of the main outcomes of this work is an affirmative answer to this question.

**Theorem 1.1.** *There is a randomized algorithm that, given a parameter $\varepsilon \in (0,1)$ and query access to any Boolean function $f : [n]^d \to \{0,1\}$ defined over the hypergrid, makes $\widetilde{O}(d^{5/6}\varepsilon^{-7/3})$ non-adaptive queries to $f$ and (a) always accepts if $f$ is monotone, and (b) rejects with probability $> 2/3$ if $\varepsilon_f > \varepsilon$.*

**Continuous Domains.** To the best of our knowledge, monotonicity testing has so far been restricted to discrete domains. What can one say about monotonicity testing when the domain is $\mathbb{R}^d$? Indeed, for functions whose range is $\mathbb{R}$, the aforementioned lower bound of $\Omega(\log n)$ precludes any such tester (with finite query complexity) even in one dimension. On the other hand, Theorem 1.1 (and indeed the results of Dodis et al. [DGL+99] and BRY [BRY14a]) suggest that for Boolean functions $f : \mathbb{R}^d \to \{0,1\}$, a monotonicity tester might be possible. In this work, we spell out the natural definitions for monotonicity testing over $\mathbb{R}^d$, and show that $o(d)$-testers do exist when the distance is with respect to any product measure.

**Theorem 1.2** (*Informal*, Formal version: Theorem 6.3)**.** *There is a one-sided, non-adaptive $\widetilde{O}(d^{5/6}\varepsilon^{-7/3})$-query monotonicity tester for* measurable *Boolean functions $f : \mathbb{R}^d \to \{0,1\}$ with respect to arbitrary product measures*[2] $\mu = \prod_i \mu_i$.

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

[2] Each $\mu_i$ is described by a non-negative Lebesgue integrable function over $\mathbb{R}$, whose integral over $\mathbb{R}$ is 1.

To gain perspective, the reader may restrict attention to functions defined over the continuous cube $[0,1]^d$, and assume the uniform measure $\mu$ on this cube. This is the natural generalization of property testing on the domains $\{0,1\}^d$ and $[n]^d$ as described above. The only restriction on the function we are testing is that the set of points where the function takes value 1 (or 0) must be (Lebesgue)-measurable. The distance between two functions $\texttt{dist}(f,g) := \mathbf{Pr}_{x\sim\mu}[f(x) \neq g(x)]$ is the measure of the points at which they differ. The distance to monotonicity of a function $f$ is $\inf_{g\in\mathcal{M}} \texttt{dist}(f,g)$ where $\mathcal{M}$ is the set of all monotone functions. (In general, we use any measure to define distance. For instance, we can test monotonicity of functions $f : \mathbb{R}^d \to \{0,1\}$ over the Gaussian measure.)

Note that the result of Theorem 1.2 holds for all measurable functions, with no dependence on surface area or "complexity" of $f$. This can be contrasted with the recent result of De, Mossel, and Neeman [DMN18], who showed that Junta testing of Boolean functions $f : \mathbb{R}^d \to \{0,1\}$ over the Gaussian measure requires some dependence on the surface area of $f$.

Given the proof techniques for Theorem 1.1, the proof of Theorem 1.2 follows from standard measure theoretic methods. Nonetheless, we believe that there is a useful conceptual message in Theorem 1.2. It gives the natural "limit" of monotonicity testing for hypergrids $[n]^d$, as $n \to \infty$. This result also underscores the significance of getting testers independent of $n$ (for hypergrids), since it leads to testers for all measurable functions.

## 1.1 Domain Reduction

**Discrete Hypergrid $[n]^d$.** A natural approach 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 $\hat{f}$ 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 $\hat{f}$ 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 $\varepsilon_{\hat{f}}$ 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 §8, 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 8.1).

Nonetheless, one can consider the question of reducing the domain to a $[k]^d$ hypergrid, for some parameter $k \ll n$, by sampling $k$ iid uniform elements of $[n]$ across each dimension. For $k$ *independent* of $n$, can we lower bound the expected distance of the function restricted to a random $[k]^d$ hypergrid? BRY studied this question for the $d = 1$ case (the line domain), and prove that this is indeed possible [BRY14a]. Our main technical result is a domain reduction theorem for all $d$, by setting $k = \mathrm{poly}(d/\varepsilon_f)$. That is, we show that *if $k = \Theta((d/\varepsilon_f)^7)$, then the expected distance to monotonicity of $f$ restricted to a random $[k]^d$ hypergrid is $\Omega(\varepsilon_f)$.*

For a precise statement, let us fix a function $f : [n]^d \to \{0,1\}$. Construct $d$ random (multi-) sets $T_1, \ldots, T_d \subseteq [n]$, each formed by taking $k$ iid uniform samples from $[n]$. 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.)

**Theorem 1.3** (Domain Reduction Theorem for Hypergrids)**.** *Let $f : [n]^d \to \{0,1\}$ be any function and let $k \in \mathbb{Z}^+$ be a positive integer. 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 $k$ iid 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, if $k \geq \left(\frac{2Cd}{\varepsilon_f}\right)^7$, then $\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f/2$.*

The construction in §8 shows that such a theorem is impossible for $k = o(\sqrt{d})$, and thus, domain reduction requires $k$ and $d$ to be polynomially related. We leave figuring out the best dependence on $k$ and $d$ as an open question. For the $d = 1$ case, BRY give a much better lower bound of $\varepsilon_f - 5\sqrt{\varepsilon_f/k}$ (Theorem 3.1 of [BRY14a]).

Given Theorem 1.3, one can sample a random $[k]^d$ hypergrid denoted $\boldsymbol{T}$ and apply the tester in [BCS18] on $f_{\boldsymbol{T}}$. The final query complexity is $\widetilde{O}(d^{5/6}) \cdot \operatorname{poly}\log k$. Setting $k = \operatorname{poly}(d/\varepsilon)$, one gets a purely sublinear-in-$d$ tester (see §7 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. If one could design a $\sqrt{d} \cdot \operatorname{poly}\log n$ query monotonicity tester for the domain $[n]^d$, then Theorem 1.3 can be used as a black box to achieve an $\widetilde{O}(\sqrt{d})$ monotonicity tester.

Note that because the dependence of [BCS18] is $\operatorname{poly}\log k$, and in light of the fact that $k = \operatorname{poly}(d)$ is needed for domain reduction to hold (Theorem 8.1), any improvement to Theorem 1.3 would only give a constant factor improvement to the query complexity of the overall tester.

**Continuous Domains.** The independence of $n$ in Theorem 1.3 suggests the possibility of a domain reduction result for Boolean functions defined over $\mathbb{R}^d$. We show that this is indeed true if $f : \mathbb{R}^d \to \{0,1\}$ is measurable (formal definitions in §6) and defined with respect to a (Lebesgue integrable) product distribution.

**Theorem 1.4** (Domain Reduction Theorem for $\mathbb{R}^d$)**.** *Let $f : \mathbb{R}^d \to \{0,1\}$ be any measurable function and let $k \in \mathbb{Z}^+$ be a positive integer. Let $\mu = \prod_{i=1}^d \mu_i$ be a (Lebesgue integrable) product distribution such that the distance to monotonicity of $f$ wrt $\mu$ is $\varepsilon_f$. If $\boldsymbol{T} = T_1 \times \cdots \times T_d$ is a randomly chosen hypergrid, where for each $i \in [d]$, $T_i \subset \mathbb{R}$ is formed by taking $k$ iid samples from $\mu_i$, 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, if $k \geq \left(\frac{2Cd}{\varepsilon_f}\right)^7$, then $\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f/2$.*

The above theorem essentially reduces the continuous domain to a discrete hypergrid $[k]^d$ where $k$ is at most some polynomial of the dimension $d$. At this point, our result from [BCS18] implies Theorem 1.2; a formal proof is given in §7.

The main ingredient in the proof of Theorem 1.4 is a discretization lemma (Lemma 6.6). Using standard measure theory, one can show that for any measurable Boolean function over $\mathbb{R}^d$ and any $\delta > 0$, there exists a large enough natural number $N = N(f, \delta)$ with the following property. The domain $\mathbb{R}^d$ can be divided into an $N^d$ sized $d$-dimensional grid, such that in at least a $(1-\delta)$-fraction of grid boxes, the function $f$ has the same value. (In some sense, this is what it means for $f$ to be measurable.) Ignoring the $\delta$-fraction of "mixed" boxes, the function $f$ can be thought of as a discrete function on $[N]^d$.

*The only guarantee on $N$ is that it is finite*; as it depends on $f$, $N$ could be extremely large compared to $d$. This is where Theorem 1.3 shows its power. The sampling parameter $k$ is independent of $N$, and this establishes Theorem 1.4. We give a detailed proof in §6.2.

We remark here that given Theorem 1.4, one can also apply the techniques of Dodis et al. [DGL+99] and BRY [BRY14a] to get a $\widetilde{O}(d/\varepsilon)$-query tester. However, as we mentioned before, we are unaware of an explicit study of monotonicity testing over the continuous domain.

### 1.2 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, SS08, Bha08, BCSM12, FR10, BBM12, RRSW11, BGJ+12, CS13, CS14a, CST14, BRY14a, BRY14b, CDST15, CDJS15, KMS15, BB16, CWX17, BCS18].

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

For Boolean monotonicity testing over general hypergrids, Dodis et al. [DGL+99] give a non-adaptive, one-sided $O((d/\varepsilon)\log^2(d/\varepsilon))$-query tester. This was improved to $O((d/\varepsilon)\log(d/\varepsilon))$ by Berman, Raskhodnikova and Yaroslavtsev [BRY14a]. This paper also proves an $\Omega(\log(1/\varepsilon))$ separation between adaptive and non-adaptive monotonicity testers for $f : [n]^2 \to \{0,1\}$ by demonstrating 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 [BCS18] gives a monotonicity tester with query complexity $\widetilde{O}(d^{5/6}\log^{4/3} n)$ via directed isoperimetric inequalities for augmented hypergrids.

### 1.3 Further Remarks

**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 [BRY14a]. They prove (Lemma 2.2 + Fact 1.1, [BRY14a]) that $L_p$-monotonicity testing can be reduced to (non-adaptive, one-sided) Boolean monotonicity testing. Thus, Theorem 1.1 implies an $L_p$-monotonicity tester for $f : [n]^d \to [0,1]$ functions which makes $o(d)$ queries (the dependence on $\varepsilon$ will depend on $p$). This improves upon Theorem 1.3 of [BRY14a].

We also believe our main theorem Theorem 1.1 can also be used to *estimate* the distance-to-monotonicity for functions $f : [n]^d \to \{0,1\}$ in time independent of $n$. The works of [BRY14a, PRR06] also relate distance estimation for Boolean functions and *tolerant* testing over $L_p$-distances, and our results should have implications for this. Finally, generalizing $L_p$-testing to the continuous domain should be possible. We leave all these interesting directions as future work and omit it from this extended abstract.

**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 $k \geq \Omega(\sqrt{d})$ (Theorem 8.1), 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 9.1). This statement may be of independent interest. We were unable to find a reference to such a statement and provide a proof in §9.

## 2 Proving the Domain Reduction Theorem 1.3: Overview

The theorem is a direct corollary of the following lemma, applied to each dimension.

**Lemma 2.1** (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 $k$ iid 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.*

This lemma is the heart of our results, and in this section we give an overview of its proof. 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 large probability ($> 9/10$, say), and thus the restriction of $f$ to this sample has distance $\Omega(\varepsilon_f)$. However, $\Omega(\varepsilon_f)$ is weak for what we need since, even if one could generalize this argument to the setting of Lemma 2.1, we would need to apply it $d$ times to get the full domain reduction (Theorem 1.3). This would imply a final lower bound of $\varepsilon_f/2^d$, which has little value towards proving a sublinear-in-$d$ query tester.

Fortunately, quantitatively stronger domain reduction exists for the line. BRY ([BRY14a], Theorem 3.1) proves that if one samples $\Theta(s^2/\varepsilon_f)$ points, then the expected distance of the restricted function is at least $\varepsilon_f(1 - 1/s)$. Numerically speaking, this is encouraging news, since we could try to set $s = \Theta(d)$ and iterate this argument $d$ times (over each dimension). Of course, this result for the line alone is not enough to deal with the structure of general hypergrids, but forms a good sanity check.

Consider the general case of Lemma 2.1. For brevity, we let $D := [n] \times \left(\prod_{i=2}^{d}[n_i]\right)$ and $D_T := T \times \left(\prod_{i=2}^{d}[n_i]\right)$ denote the original and reduced domains, respectively. Note that $|D_T| = \frac{k}{n}|D|$.

The standard handle on the distance to monotonicity is the violation graph of $f$, arguably first formalized by Fischer et al [FLN+02]. The graph has vertex set $D$ 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 $M$ in the violation graph satisfies $|M| = \varepsilon_f|D|$. 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 $\varepsilon_{f_T}$, we need 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 (simultaneously) in the violation graph of $f_T$.

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

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

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 $\ell \in \mathcal{L}$. By definition, the lower-endpoints of $M^{(\ell)}$ all lie on $\ell$, and thus are all comparable. Let $M^{(\ell)} = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ where $x_1 \prec \cdots \prec x_m$ and observe that, for any $j \in [m]$, $x_1, \ldots, x_j \prec y_j, \ldots, y_m$. Since the function is Boolean, every $x \in \{x_1, \ldots, x_j\}$ forms a violation to monotonicity with every $y \in \{y_j, \ldots, y_m\}$, 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 $\ell$, 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 $\ell$, and the analysis becomes identical to the one-dimensional case. We could thus apply Theorem 3.1 of [BRY14a] to each $\ell \in \mathcal{L}$ to prove Lemma 2.1. However, the assumption that the $y_i$'s have distinct 1-coordinates is far from the truth. As we explain below, there are examples where all the $y_i$'s have the same 1-coordinate, thereby lying 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 $a \notin T$), implying that $\mathbf{E}_T\left[|M_T^{(\ell)}|\right]$ could be as small as $(k/n)^2 \cdot |M^{(\ell)}|$. Thus, if there existed a function $f$ such that a "collision of $y$'s 1-coordinates" could not be avoided for a large number of lines, then this would preclude such a line-by-line approach to proving Lemma 2.1. Unfortunately, there are examples of violation matchings where this happens. Consider Example 1, and the left part of Fig. 1, shown at the end of this section. For the lowest line, all the corresponding $y$'s in $M^{(\ell)}$ have the same 1-coordinate.

Our main insight is that for any $f$, 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 $(\ell, S)$-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 $(\ell, S)$" 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 2.3.

If there is a maximum cardinality matching $M$ in the violation graph of $f$ such that all stacks have size at most 1, then the one-dimensional domain reduction can be directly applied. 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 (shapes); again consider Example 1. Nevertheless, we prove that for any $\lambda \geq 2$, there is a matching $M$ where the total number of vertices participating in stacks of size at least $\lambda$ is at most $|D|/\mathrm{poly}(\lambda)$.

**Lemma 2.2** (Stack Bound). *For any integer $\lambda \geq 2$, there exists a maximum cardinality matching $M$ in the violation graph of $f$, where $\sum_{(\ell,S):|M^{(\ell,S)}|\geq\lambda} |M^{(\ell,S)}| \leq \frac{6}{\sqrt{\lambda}} \cdot |D|$.*

The main creativity to prove this lemma lies in the choice of $M$. Given a matching, we define the vector $\Lambda(M)$ that enumerates all the stack sizes in non-decreasing order. We show that the maximum cardinality matching $M$ with the lexicographically largest $\Lambda(M)$ serves our purpose. That is, we choose $M$ that maximizes the minimum stack size, and then subject to this maximizes the second minimum, and so on. It may seem counter-intuitive that we want a matching with small stack sizes, and yet our potential function maximize the minimum. The intuitive 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. The proof uses a matching rewiring argument to show that any large stack must be "adjacent" to many moderate size stacks. If two stacks are appropriately "aligned", one could change the matching to move points from one stack to the other. Large stacks cannot be aligned with small stacks, since one could rewire the matching to increase the potential. But since the function is Boolean one can show that there are many opportunities for rewiring the violation matching. Thus, there isn't enough "room" for many large stacks. We then apply some technical charging arguments to bound the total number of points in large stacks. The full proof is given in §4.

With the stack bound in hand, we need to generalize the one-dimensional argument of BRY (Theorem 3.1 [BRY14a]) to account for bounded stack sizes. Then, we bound $|M_T^{(\ell)}|$ for all $\ell$, and get the final lower bound on the distance $\varepsilon_{f_T}$.

**Lemma 2.3** (Line Sampling)**.** *Suppose that $M$ is a matching in the violation graph of $f$, such that for some $\lambda \in \mathbb{Z}^+$, $|M^{(\ell,S)}| \leq \lambda$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$. Then, for any $\ell \in \mathcal{L}$,*

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

The proof is a fairly straightforward generalization of the arguments in [BRY14a] for the $\lambda = 1$ case. The idea is to control the size of the maximum cardinality matching $M_T^{(\ell)}$ by analyzing the discrepancy of a random subsequence of a sequence of 1s and 0s. For the sake of simplicity, we give a proof that achieves a weaker dependence on $\varepsilon_f$ than in [BRY14a]. Our proof of Lemma 2.3 is given in §5. We note that BRY give a stronger lower bound (without the $\sqrt{\ln k}$) and also bound the variance.

**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 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 $\leq 1$.

# 3 Domain Reduction: Proof of Lemma 2.1

In this section, we use Lemma 2.2 and Lemma 2.3 to prove Lemma 2.1. Recall that $D := [n] \times \left(\prod_{i=2}^d [n_i]\right)$ and $D_T := T \times \left(\prod_{i=2}^d [n_i]\right)$ denote the original and reduced domains, respectively. Note that $|D_T| = \frac{k}{n}|D|$. Let $M$ be the matching given by Lemma 2.2 with $\lambda := \lceil 36k^{2/7} \rceil$.

By Lemma 2.2, we have $\left|\bigcup_{(\ell,S):|M^{(\ell,S)}|\geq 40k^{2/7}} M^{(\ell,S)}\right| \leq \frac{6}{\sqrt{\lambda}} \cdot |D| \leq \frac{|D|}{k^{1/7}}$. Let

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

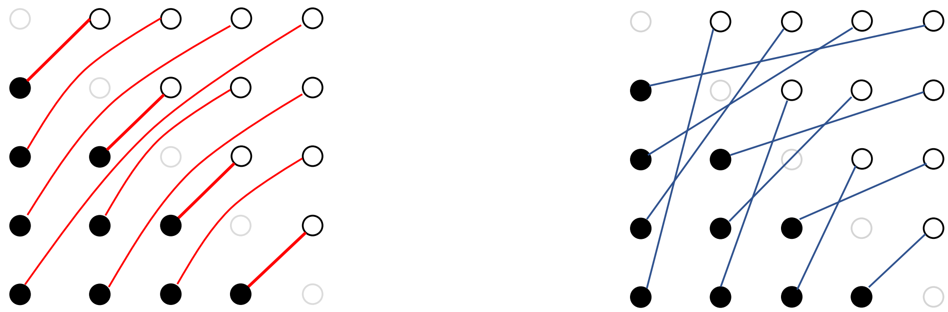

Figure 1: Accompanying illustration for Example 1 showing two different maximum cardinality violation matchings for the anti-majority function $f : [5]^2 \to \{0,1\}$ which have very different stack sizes. Black (white, resp.) circles represent vertices where $f = 1$ ($f = 0$, resp.) and connecting lines 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 $\leq 1$.

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

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

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 2.3 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 2.3, 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 2.3 to measure $\mathbf{E}_T[|M_T^{(\ell)}|]$ for each $\ell$. Also note that the number of lines is $|\mathcal{L}| = |D|/n$.

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

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 $|D_T| = \frac{k}{n}|D|$, we get $\mathbf{E}_T\left[\varepsilon_{f_T}\right] \geq \frac{|M|}{|D|} - \frac{C}{k^{1/7}} = \varepsilon_f - \frac{C}{k^{1/7}}$. This completes the proof of Lemma 2.1. ■

# 4 Stack Bound: Proof of Lemma 2.2

We are given a positive integer $\lambda \geq 2$ and a Boolean function $f : D \to \{0,1\}$ where $D = [n] \times \left(\prod_{i=2}^{d}[n_i]\right)$ is a rectangular hypergrid for some $n, n_2, \ldots, n_d \in \mathbb{Z}^+$. Lemma 2.2 asserts there is a maximum cardinality matching $M$ such that $\sum_{(\ell,S):|M^{(\ell,S)}|\geq\lambda} |M^{(\ell,S)}| \leq \frac{6}{\sqrt{\lambda}} \cdot |D|$.

Given a matching $M$, we consider the vector (or technically, list) $\Lambda(M)$ indexed by stacks $(\ell, S)$ 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 $|B| = |R| \leq \frac{6}{\sqrt{\lambda}} \cdot |D|$. Indeed, define $\delta$ to be such that $|B| = \delta|D|$. 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 $\ell$, the number of stacks $(\ell, S)$ which are *non-low* cannot be "too many".

**Claim 4.1.** *For any line $\ell$, the number of non-low stacks $\ell$ participates in is at most $\frac{n}{\lambda-1}$.*

*Proof.* Fix any line $\ell$ 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 $\ell$. The size of this set can't be bigger than the length of $\ell$, 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 ($> 6|D|/\sqrt{\lambda}$), 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 $\ell_z$ ($S_z$, 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 $(\ell, S)$, we get a bunch of other "non-low" stacks participating with the line $\ell$.

**Claim 4.2.** *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, $(\ell, S)$, 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 $(\ell_w, S)$. For brevity's sake, let us denote their sizes in $M$ as $\lambda_1, \lambda_2, \lambda_3$, and $\lambda_4$, respectively. In $N$, their sizes become $\lambda_1 - 1, \lambda_2 - 1, \lambda_3 + 1$, and $\lambda_4 + 1$. Note that $\lambda_3 \leq \lambda - 2$ and both $\lambda_1$ and $\lambda_2$ are $\geq \lambda$. In particular, the "new" size of stack $(\ell, S_z)$ is still smaller than the "new" sizes of stacks $(\ell, S)$ and $(\ell_w, S_z)$. That is, the vector $\Lambda(N)$, even without the increase in $\lambda_4$, 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 $\alpha$-rich if $\rho_y \geq \alpha$. A point $x \in R$ is $\alpha$-rich if its blue partner $y \in B$ (i.e. $(x, y) \in M$) is $\alpha$-rich. We also call the pair $(x, y)$ an $\alpha$-rich pair.

**Claim 4.3.** *If $|B| = \delta|D|$, then at least $\delta|D|/2$ of these points are $\delta/4$-rich.*

*Proof.* Let $B^{(\text{poor})} \subseteq B$ be the points with $\rho_y < \delta/4$. We show $|B^{(\text{poor})}| \leq \delta|D|/2$ which would prove the claim. To see this, first observe $B^{(\text{poor})} \subseteq \bigcup_{y \in B^{(\text{poor})}} (\mathcal{I}_y \cap B)$. Now consider the *minimal* subset $B^{(\text{poor})}_{\text{min}} \subseteq B^{(\text{poor})}$ such that $\bigcup_{y \in B^{(\text{poor})}_{\text{min}}} \mathcal{I}_y = \bigcup_{y \in B^{(\text{poor})}} \mathcal{I}_y$. 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 $y \in B^{(\text{poor})}_{\text{min}}$. In particular, this implies

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

Therefore,

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

The first equality follows from the definition of $B^{(\text{poor})}_{\text{min}}$ (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 4.3 is that there are at least $\delta|D|/2$ red points which are $\delta/4$-rich. In particular, there must exist some line $\ell$ that contains $\geq \delta n/2$ red points in it which are $\delta/4$-rich. Let this line be $\ell$ and let $R_\ell \subseteq$

$\ell$ be the set of rich red points. Let $B_\ell$ 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 4.2, 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 $B^\ell$, $\bigcup_{y \in B^\ell} \mathcal{I}_y \cap B$. There are precisely $n$ slices in total, and for a vertex $z \in D$, $S_z$ 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 $X \subseteq D$, 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 4.4.** $\left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y \cap B)\right| \geq \frac{\delta}{8} \left|\bigcup_{y \in B^\ell} \mathsf{proj}_1(\mathcal{I}_y)\right|$

*Proof.* As in the proof of Claim 4.2, let $B^\ell_{\mathsf{min}} \subseteq B^\ell$ be a minimal cardinality subset of $B^\ell$ 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 $B^\ell_{\mathsf{min}}$.

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

Now importantly, $|\mathsf{proj}_1(R^\ell)| = |R^\ell| \geq \frac{\delta}{2} \cdot n$ since the 1-coordinates of elements of $R^\ell$ are distinct (since $R^\ell$ 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 4.4, 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, $\ell$ participates in at least $\frac{\delta^2}{16} \cdot n$ non-low stacks. Thus, using Claim 4.1, 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 2.2. $\blacksquare$

## 5 Line Sampling: Proof of Lemma 2.3

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 2.3) states

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

We note that BRY (Theorem 3.1, [BRY14a]) prove a stronger theorem for the $\lambda = 1$ case (that gets an additive error of $\Theta(\sqrt{k})$). Our proof follows a similar approach.

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 5.1.** *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 $\ell$, 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 $i < j$. ∎

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 5.2.** *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 $j$th 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 5.3.** *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 $\sum_{s\in T: s\le t} w^+(s)$. Furthermore note that adding any $s \le t$ 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 5.3. 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

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

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 5.1, 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

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

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

$$\mathbf{E}_T\left[\max_{t\in T}\sum_{s\in T: s\le t} w^-(s) - w^+(s)\right] \le \lambda k\cdot\mathbf{Pr}\left[\max_{t\in T} Z_t > a\right] + a \le \lambda + a \le 3\lambda\sqrt{k\ln k}. \tag{7}$$

Lemma 2.3 follows from Lemma 5.2, Lemma 5.3, (6), and (7).

## 6 The Continuous Domain

We start with measure theory preliminaries. We refer the reader to Nelson [Nel15] and Stein-Shakarchi [SS05] for more background. Given two reals $a < b$, we use $(a, b)$ to denote the open interval, and $[a, b]$ to denote the closed interval. Given $d$ closed intervals $[a_i, b_i]$ for $1 \le i \le d$, we call their Cartesian product $\prod_{i\in[d]}[a_i, b_i]$ a *box*. Two intervals/boxes are *almost disjoint* if their interiors are disjoint (they can intersect only at their boundary). An *almost partition* of a set $S$ is a collection $\mathcal{P}$ of sets that are pairwise almost disjoint and $\bigcup_{P\in\mathcal{P}} P = S$. A set $U$ is *open* if for each point $x \in U$, there exists an $\varepsilon > 0$ such that the sphere centered at $x$ of radius $\varepsilon$ is contained in $U$.

We let $\mu = \prod_{i\in[d]} \mu_i$ be an arbitrary *product measure* over $\mathbb{R}^d$. That is, each $\mu_i$ is described by a non-negative Lebesgue integrable function over $\mathbb{R}$, whose total integral is 1 (this is the pdf). Abusing notation, we use $\mu_i([a_i, b_i]) = \mathbf{Pr}_{x\sim\mu_i}[a_i \le x \le b_i]$ to denote the integral of $\mu_i$ over this interval. Indeed, this is the probability measure of the interval. The volume of a box $B = \prod_{i\in[d]}[a_i, b_i]$ is denoted as $\mu(B) = \prod_{i\in[d]} \mu_i([a_i, b_i]) = \mathbf{Pr}_{x\sim\mu}[x \in B]$.

We use the definition of measurability of Chapter 1.1.3 of [SS05]. Technically, this is given with respect to the standard notion of volume in $\mathbb{R}^d$. Chapter 6, Lemma 1.4 and Chapter 6.3.1 show that the definition is valid for the notion of volume with respect to $\mu$, as we've defined above. The *exterior measure* $\mu_*$ of any set $E$ is the infimum of the sum of volumes of a collection of closed boxes that contain $E$.

**Definition 6.1.** *Given a product measure $\mu = \prod_i \mu_i$ over $\mathbb{R}^d$, we say $E \subseteq \mathbb{R}^d$ is Lebesgue-measurable with respect to $\mu$ if for any $\varepsilon > 0$, there exists an open set $U \supseteq E$ such that $\mu_*(U \setminus E) < \varepsilon$. If this holds, then the $\mu$-measure of $E$ is defined as $\mu(E) := \mu_*(E)$.*

Given a function $f : \mathbb{R}^d \to \{0, 1\}$, we will often slightly abuse notation by letting $f$ denote the *set* it indicates, i.e. the set in $\mathbb{R}^d$ where $f$ evaluates to 1. We say that $f$ is a *measurable function wrt* $\mu$ if this set is measurable wrt $\mu$. Similarly, we use $\overline{f}$ to denote the *set* where $f$ evaluates to 0.

We are now ready to define the notion of distance between two functions. In §6.3, we prove that all monotone Boolean functions are measurable (Theorem 6.7) with respect to $\mu$. Also, measurability is closed under basic set operations and thus the following notion of distance to monotonicity is well-defined.

**Definition 6.2** (Distance to Monotonicity)**.** *Fix a product measure $\mu$ on $\mathbb{R}^d$. We define the distance between two measurable functions $f, g : \mathbb{R}^d \to \{0, 1\}$ with respect to $\mu$, as*

$$\mathtt{dist}_\mu(f, g) := \mu\left(\left\{z \in \mathbb{R}^d : f(z) \neq g(z)\right\}\right) = \mu\left(f\Delta g\right) \tag{8}$$

*The distance to monotonicity of $f$ wrt $\mu$ is defined as*

$$\varepsilon_{f,\mu} := \inf_{g\in\mathcal{M}} \mathtt{dist}_\mu(f, g) = \inf_{g\in\mathcal{M}} \mu\left(f\Delta g\right) \tag{9}$$

*where $\mathcal{M}$ denotes the set of monotone Boolean functions over $\mathbb{R}^d$.*

We are now equipped to state the formal version of Theorem 1.2, for testing Boolean functions over $\mathbb{R}^d$.

**Theorem 6.3.** *Let $\mu = \prod_{i=1}^d \mu_i$ be a product measure for which we have the ability to take independent samples from each $\mu_i$. There is a randomized algorithm which, given a parameter $\varepsilon > 0$ and a measurable function $f : \mathbb{R}^d \to \{0, 1\}$ that can be queried at any $x \in \mathbb{R}^d$, makes $\widetilde{O}(d^{5/6}\varepsilon^{-7/3})$ non-adaptive queries to $f$, and (a) always accepts if $f$ is monotone, and (b) rejects with probability $> 2/3$ if $\varepsilon_{f,\mu} > \varepsilon$.*

We give a formal proof of Theorem 6.3 in §7. The proof requires some tools to discretize measurable sets, which we provide in the next two sections.

## 6.1 Approximating measurable sets by grids

We first start with a lemma about probability measures over $\mathbb{R}$.

**Lemma 6.4.** *Given any probability measure $\mu$ over $\mathbb{R}$, and any $N \in \mathbb{N}$, there exists an almost partition of $\mathbb{R}$ into $N$ intervals $\boldsymbol{I}_N = \{\mathcal{I}_1, \ldots, \mathcal{I}_N\}$ of equal $\mu$-measure. That is, for each $j \in [N]$, $\mathbf{Pr}_{x\sim\mu}[x \in \mathcal{I}_j] = \frac{1}{N}$. Furthermore, for any $k \in \mathbb{N}$, $\boldsymbol{I}_{kN}$ is a refinement of $\boldsymbol{I}_N$.*

*Proof.* $\mu$ is a probability measure, and thus is described by a non-negative Lebesgue-integrable function (it's pdf). Chapter 2, Prop 1.12 (ii) of [SS05] states that the Lebesgue integral is continuous and thus it's CDF, $F(t) := \mu(\{x \in \mathbb{R} : x \leq t\})$, is continuous. Moreover $F$ is non-decreasing with range $[0, 1]$. Therefore, for every $\theta \in (0, 1)$ there is at least one $t$ with $F(t) = \theta$. Thus, let's define $F^{-1}(\theta)$ to be the supremum over all $t$ satisfying $F(t) = \theta$. Let $F^{-1}(0) = -\infty$ and $F^{-1}(1) = +\infty$. The lemma is proved by the intervals $\mathcal{I}_j = [F^{-1}((j-1)/N), F^{-1}(j/N)]$ for $j \in \{1, \ldots, N\}$. The refinement is evident by the fact that any interval in $\boldsymbol{I}_N$ can be expressed as an almost partition of intervals from $\boldsymbol{I}_{kN}$ (for $k \in \mathbb{N}$). ∎

Thus, given a product distribution $\mu = \prod_{i=1}^d \mu_i$ and any $N \in \mathbb{N}$, we can apply the above lemma to each of the $d$ coordinates to obtain the set of $Nd$ intervals $\left\{\mathcal{I}_j^{(i)} : i \in [d] : j \in [N]\right\}$ for which $\mu_i\left(\mathcal{I}_j^{(i)}\right) = 1/N$ for every $i \in [d]$, $j \in [N]$. We define

$$\mathbf{G}_N := \left\{\prod_{i=1}^{d} \mathcal{I}_{z_i}^{(i)} : z \in [N]^d\right\}$$

and observe that (a) $\mathbf{G}_N$ is an almost partition of $\mathbb{R}^d$ and (b) $\mathbf{G}_{kN}$ is a refinement of $\mathbf{G}_N$ for any $k \in \mathbb{N}$. (Since $d$ is fixed, we will not carry the dependence on $d$.) We informally refer to $\mathbf{G}_N$ as a *grid*. Since $\mathbf{G}_N$ is an almost partition, we can define the function $\mathsf{box}_N : \mathbb{R}^d \to [N]^d$ as follows. For $x \in \mathbb{R}^d$, we define $\mathsf{box}_N(x)$ to be the lexicographically least $z \in [N]^d$ such that the box $\prod_{i=1}^d \mathcal{I}_{z_i}^{(i)}$, of $\mathbf{G}_N$, contains $x$. (Note that for all but a measure zero set, points in $\mathbb{R}^d$ are contained in a unique box of $\mathbf{G}_N$.)

In the following lemma, we show that any measurable set can be approximated by a sufficiently fine grid. In some sense, this *is* the definition of measurability.

**Lemma 6.5.** *For any measurable set $E$ and any $\alpha > 0$, there exists $N = N(E, \alpha) \in \mathbb{N}$ such that there is a collection $\mathbf{B} \subseteq \mathbf{G}_N$ satisfying $\mu(E \,\Delta\, \bigcup_{B\in\mathbf{B}} B) \leq \alpha$.*

*Proof.* Chapter 1, Theorem 3.4 (iv) of [SS05] states that for any measurable set $E$ and any $\epsilon > 0$, there exists a finite union $\bigcup_{r=1}^m B_r$ of closed boxes such that $\mu(E\Delta \bigcup_{r=1}^m B_r) \leq \epsilon$. We invoke this theorem with $\epsilon = \alpha/2$ to get the collection of boxes $B_1, \ldots, B_m$. Note that these boxes may intersect, and might not form a grid. We build a grid by setting $N = \lceil 2md/\alpha \rceil$ and considering $\mathbf{G}_N$. The desired collection $\mathbf{B} \subseteq \mathbf{G}_N$ is the set of boxes in $\mathbf{G}_N$ contained in $\bigcup_{r=1}^m B_r$. Observe that

$$\mu\left(E\Delta \bigcup_{B\in\mathbf{B}} B\right) \leq \mu\left(E\Delta \bigcup_{r=1}^{m} B_r\right) + \mu\left(\bigcup_{r=1}^{m} B_r \setminus \bigcup_{B\in\mathbf{B}} B\right) \leq \alpha/2 + \sum_{r=1}^{m} \mu\left(B_r \setminus \bigcup_{B\in\mathbf{B}} B\right) \tag{10}$$

by subadditivity of measure. We complete the proof by bounding $\mu(B_r \setminus \bigcup_{B\in\mathbf{B}} B)$ for an arbitrary $r \in [m]$.

Let $B_r := \prod_{i=1}^d [a_i, b_i]$ denote an arbitrary box from $\{B_1, \ldots, B_m\}$ and let $\delta_i := \mu_i([a_i, b_i])$. Observe that the interval $[a_i, b_i]$ contains exactly $\lfloor \delta_i N \rfloor$ contiguous intervals from the almost partition $\{\mathcal{I}_j^{(i)} : j \in$

$[N]\}$ of $\mathbb{R}$. Let $\boldsymbol{I}_i$ denote the set of such intervals. Thus, $\mu_i([a_i, b_i] \setminus \bigcup_{I \in \boldsymbol{I}_i} I) \leq \delta_i - (1/N)(\lfloor \delta_i N \rfloor) \leq \delta_i - (1/N)(\delta_i N - 1) = 1/N$. Thus, the total measure of $B_r$ we discard is $\mu(B_r \setminus \bigcup_{B \in \mathbf{B}} B) \leq \prod_i \delta_i - \prod_i (\delta_i - 1/N)$. This quantity is maximized when the $\delta_i$'s are maximized; since $\delta_i \leq 1$ (each $\mu_i$ is a probability measure), we get that $\mu(B_r \setminus \bigcup_{B \in \mathbf{B}} B) \leq 1 - (1 - 1/N)^d \leq \frac{d}{N}$.

Finally, plugging this into (10), we get $\mu(E \Delta \bigcup_{B \in \mathbf{B}} B) \leq \alpha/2 + m \cdot \frac{d}{N} \leq \alpha$, since $N \geq 2md/\alpha$. ∎

We are now ready to prove our main tool, the discretization lemma.

**Lemma 6.6** (Discretization Lemma). *Given a measurable function $f : \mathbb{R}^d \to \{0, 1\}$ and $\delta > 0$, there exists $N := N(f, \delta) \in \mathbb{N}$, and a function $f^{\mathsf{disc}} : [N]^d \to \{0, 1\}$, such that $\mathbf{Pr}_{x \sim \mu}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] \leq \delta$.*

*Proof.* By assumption, $f$ and $\overline{f}$ are measurable sets. By Lemma 6.5, there exists some $N_1$ and a collection of boxes $\boldsymbol{Z}_1 \subseteq \mathbf{G}_{N_1}$ such that $\mu(f \Delta \bigcup_{B \in \boldsymbol{Z}_1} B) \leq \delta/6$. (An analogous statement holds for $\overline{f}$, with some $N_0$ and a collection $\boldsymbol{Z}_0$.) Since Lemma 6.5 also holds for any refinement of the relevant grid, let us set $N = N_0 N_1$. Abusing notation, we have two collections $\boldsymbol{Z}_0, \boldsymbol{Z}_1 \subseteq \mathbf{G}_N$ such that $\mu(f \Delta \bigcup_{B \in \boldsymbol{Z}_1} B) \leq \delta/6$ and $\mu(\overline{f} \Delta \bigcup_{B \in \boldsymbol{Z}_0} B) \leq \delta/6$.

For convenience, let us treat the boxes in $\boldsymbol{Z}_0 \cup \boldsymbol{Z}_1$ as open, so that all boxes in the collection are disjoint. Define $h : \mathbb{R}^d \to \{0, 1\}$ as follows:

$$h(x) = \begin{cases} 1 & \text{if } x \in \bigcup_{B \in \boldsymbol{Z}_1 \setminus \boldsymbol{Z}_0} B \\ 0 & \text{if } x \in \bigcup_{B \in \boldsymbol{Z}_0 \setminus \boldsymbol{Z}_1} B \\ 0 & \text{if } x \in \bigcup_{B \notin \boldsymbol{Z}_0 \Delta \boldsymbol{Z}_1} B \end{cases}$$

Since $f$ and $\overline{f}$ partition $\mathbb{R}^d$, $\mu(\bigcup_{B \in \boldsymbol{Z}_0 \cap \boldsymbol{Z}_1} B)$ and $\mu(\bigcup_{B \notin \boldsymbol{Z}_0 \cup \boldsymbol{Z}_1} B)$ are both at most $\mu(f \Delta \bigcup_{B \in \boldsymbol{Z}_1} B) + \mu(\overline{f} \Delta \bigcup_{B \in \boldsymbol{Z}_0} B) \leq \delta/3$. Combining these bounds, we have $\mu(\bigcup_{B \notin \boldsymbol{Z}_0 \Delta \boldsymbol{Z}_1} B) \leq 2\delta/3$. Thus

$$\begin{aligned} \mathtt{dist}_\mu(f, h) = \mathbf{Pr}_{x \sim \mu}[f(x) \neq h(x)] &\leq \mu\left(\bigcup_{B \in \boldsymbol{Z}_1 \setminus \boldsymbol{Z}_0} B \cap \overline{f}\right) + \mu\left(\bigcup_{B \in \boldsymbol{Z}_0 \setminus \boldsymbol{Z}_1} B \cap f\right) + \mu\left(\bigcup_{B \notin \boldsymbol{Z}_0 \Delta \boldsymbol{Z}_1} B\right) \\ &\leq \delta/6 + \delta/6 + 2\delta/3 = \delta. \end{aligned}$$

By construction, $h$ is constant in (the interior of) every grid box. Any $z \in [N]^d$ indexes a (unique) box in $\mathbf{G}_N$ (recall the map $\mathsf{box}_N : \mathbb{R}^d \to [N]^d$). Formally, we can define a function $f^{\mathsf{disc}} : [N]^d \to \{0, 1\}$ so that $\forall x \in \mathbb{R}^n$, $f^{\mathsf{disc}}(\mathsf{box}_N(x)) = h(x)$. Thus, $\mathbf{Pr}_{x \sim \mu}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] = \mathtt{dist}_\mu(f, h) \leq \delta$. ∎

### 6.2 Proof of Theorem 1.4

*Proof.* Recall that $\boldsymbol{T} = T_1 \times \cdots \times T_d$ is a randomly chosen hypergrid, where for each $i \in [d]$, $T_i \subset \mathbb{R}$ is formed by taking $k$ iid samples from $\mu_i$. We need to show that

$$\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f - \frac{C' \cdot d}{k^{1/7}}$$

for some universal constant $C' > 0$.

Set $\delta \leq k^{-d} \cdot \frac{C \cdot d}{k^{1/7}}$, where $C$ is the universal constant in Theorem 1.3. Applying Lemma 6.6 to $f$ with this $\delta$, we know there exists $N > 0$ and $f^{\mathsf{disc}} : [N]^d \to \{0, 1\}$, such that $\mathbf{Pr}_{x \sim \mu}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] \leq \delta$.

Given a random $\boldsymbol{T}$ sampled as described above, define $\widehat{\boldsymbol{T}} := \{\mathsf{box}_N(x) \in [N]^d : x \in \boldsymbol{T}\}$. Observe that (a) $\widehat{\boldsymbol{T}}$ is a $[k]^d$ sub-hypergrid in $[N]^d$ which (b) can be equivalently defined as $\widehat{\boldsymbol{T}} = \widehat{T}_1 \times \cdots \times \widehat{T}_d$ where each $\widehat{T}_i$ is formed by taking $k$ iid uniform samples from $[N]$. This is by construction of the partition $\{\mathsf{box}_z : z \in [N]^d\}$ and by definition of $\mathsf{box}_N(x)$. Theorem 1.3 and the observations above imply

$$\mathbf{E}_{\widehat{\boldsymbol{T}}}\left[\varepsilon_{f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}}\right] \geq \varepsilon_{f^{\mathsf{disc}}} - \frac{C \cdot d}{k^{1/7}} \tag{11}$$

where $C$ is some universal constant. Next, we relate $\varepsilon_{f^{\mathsf{disc}}}$ and $\varepsilon_f$. Observe that there is a bijection between $\boldsymbol{T}$ and $\widehat{\boldsymbol{T}}$ (namely, $\mathsf{box}_N$ restricted to $\boldsymbol{T}$). We say $f_{\boldsymbol{T}} = f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}$ if for all $x \in \boldsymbol{T}$, $f(x) = f^{\mathsf{disc}}(\mathsf{box}_N(x))$.

By a union bound over the $k^d$ samples,

$$\mathbf{Pr}_{\boldsymbol{T}}\left[f_{\boldsymbol{T}} \neq f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}\right] = \mathbf{Pr}_{\boldsymbol{T}}\left[\exists x \in \boldsymbol{T} : f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))\right] \leq \delta \cdot k^d \leq \frac{C \cdot d}{k^{1/7}} =: \delta'$$

since each $x \in \boldsymbol{T}$ has the same distribution as $x \sim \mu$, and $\mathbf{Pr}_{x\sim\mu}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] \leq \delta$. Thus, we get $\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq (1-\delta')\mathbf{E}_{\widehat{\boldsymbol{T}}}\left[\varepsilon_{f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}}\right] - \delta'$, since in the case $f_{\boldsymbol{T}} \neq f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}$, the difference in their distance to monotonicity is at most 1. Substituting in (11), we get

$$\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq (1-\delta') \cdot \left(\varepsilon_{f^{\mathsf{disc}}} - \frac{C \cdot d}{k^{1/7}}\right) - \delta' \geq \varepsilon_{f^{\mathsf{disc}}} - \frac{3C \cdot d}{k^{1/7}} \tag{12}$$

by definition of $\delta'$.

Now, let $g : [N]^d \to \{0,1\}$ be any monotone function satisfying $d(f^{\mathsf{disc}}, g) = \varepsilon_{f^{\mathsf{disc}}}$. Define the monotone function $\hat{f}(x) = g(\mathsf{box}_N(x))$ for all $x \in \mathbb{R}^d$. Note that $\varepsilon_f \leq \mathtt{dist}(f, \hat{f}) \leq \mathbf{Pr}_{x\sim\mu}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] + \mathtt{dist}(f^{\mathsf{disc}}, g) \leq \delta + \varepsilon_{f^{\mathsf{disc}}}$. This, in turn, implies $\varepsilon_{f^{\mathsf{disc}}} \geq \varepsilon_f - \delta \geq \varepsilon_f - \frac{C \cdot d}{k^{1/7}}$. Substituting in (12), we get

$$\mathbf{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f - \frac{4C \cdot d}{k^{1/7}}$$

which proves the theorem. ∎

### 6.3 Measurability of Monotone Functions

**Theorem 6.7.** *Monotone functions $f : \mathbb{R}^d \to \{0,1\}$ are measurable wrt product measures $\mu = \prod_{i=1}^d \mu_i$.*

*Proof.* The proof is by induction over the number of dimensions, $d$. For $d = 1$, the set $f$ is either $[z, \infty)$ or $(z, \infty)$ for some $z \in \mathbb{R}$, since $f$ is a monotone function. Any open or closed set is measurable.

Now for the induction. Choose any $\varepsilon > 0$. We will construct an open set $\mathcal{O}$ such that $\mu_*(\mathcal{O} \setminus f) \leq 8\varepsilon$. Consider the first dimension, and the corresponding measure $\mu_1$. We use $\mu_{-1}$ for the $(d-1)$-dimensional product measure in the remaining dimensions. (We use $\mu_{-1,*}$ for the $(d-1)$-dimensional exterior measure.) As shown in Lemma 6.4, there is an almost partition of $\mathbb{R}$ into $N = \lceil 1/\varepsilon^2 \rceil$ closed intervals such that each interval has $\mu_1$-measure at most $\varepsilon^2$. Let these intervals be $I_1, I_2, I_3, \ldots, I_N$. We will consider the set of intervals $\boldsymbol{I} = \{I_1 \cup I_2, I_2 \cup I_3, \ldots, I_{N-1} \cup I_N\}$ (let us treat these as open intervals). Observe that $\cup_{I \in \boldsymbol{I}} I = \mathbb{R}$, and $\mu_1(I) \leq 2\varepsilon^2$ for all $I \in \boldsymbol{I}$.

For any $x \in \mathbb{R}$, let $S_x$ be the subset of $f$ with first coordinate $x$. We will treat $S_x$ as a subset of $\mathbb{R}^{d-1}$ and use $\{x\} \times S_x$ to denote the corresponding subset of $\mathbb{R}^d$. By monotonicity, $\forall x < y$, $S_x \subseteq S_y$. By induction, each set $S_x$ is measurable in $\mathbb{R}^{d-1}$ and thus there exists an open set $\mathcal{O}_x \subseteq \mathbb{R}^{d-1}$ such that

$\mu_{-1,*}(\mathcal{O}_x \setminus S_x) \leq \varepsilon$. Define the function $h : \mathbb{R} \to [0, 1]$ such that $h(x)$ is the measure of $S_x$ (in $\mathbb{R}^{d-1}$). Crucially, $h$ is monotone because $f$ is monotone.

Call an interval $(x, y)$ *jumpy* if $h(y) > h(x) + \varepsilon$ and let $\boldsymbol{J} \subseteq \boldsymbol{I}$ be the set of jumpy intervals in $\boldsymbol{I}$. For a non-jumpy interval $I = (x, y) \in \boldsymbol{I} \setminus \boldsymbol{J}$, define $\mathcal{O}_I := I \times \mathcal{O}_y$. Note that $\mathcal{O}_I$ is open and by monotonicity, $\mathcal{O}_I \supseteq \bigcup_{z \in I}(\{z\} \times S_z) = \{z \in f : z_1 \in I\}$.

The open set $\mathcal{O} := (\bigcup_{J \in \boldsymbol{J}} J \times \mathbb{R}^{d-1}) \cup (\bigcup_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mathcal{O}_I)$ contains (the set) $f$. It remains to bound

$$\begin{aligned}\mu_*(\mathcal{O} \setminus f) &\leq \mu_*\left(\bigcup_{J \in \boldsymbol{J}} J \times \mathbb{R}^{d-1}\right) + \mu_*\left(\bigcup_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mathcal{O}_I \setminus f\right) \\ &\leq \sum_{J \in \boldsymbol{J}} \mu_1(J) + \sum_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mu_*(\mathcal{O}_I \setminus f) \leq 2\varepsilon^2 |\boldsymbol{J}| + \sum_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mu_*(\mathcal{O}_I \setminus f). \end{aligned} \tag{13}$$

To handle the first term, note that there are at least $|\boldsymbol{J}|/2$ disjoint intervals in $\boldsymbol{J}$ and each such interval represents a jump of at least $\varepsilon$ in the value of $h$. Thus, $|\boldsymbol{J}|/2 \leq 1/\varepsilon$ and so $|\boldsymbol{J}| \leq 2/\varepsilon$.

Now, consider $I = (x, y) \in \boldsymbol{I} \setminus \boldsymbol{J}$. We have $\mathcal{O}_I = I \times \mathcal{O}_y$. By monotonicity $\mathcal{O}_I \setminus f \subseteq \mathcal{O}_I \setminus (I \times S_x) = (I \times \mathcal{O}_y) \setminus (I \times S_x) = I \times (\mathcal{O}_y \setminus S_x)$. Since $S_y \supseteq S_x$, $\mathcal{O}_y \setminus S_x = (\mathcal{O}_y \setminus S_y) \cup (S_y \setminus S_x)$. By sub-additivity of exterior measure, $\mu_{-1,*}(\mathcal{O}_y \setminus S_x) \leq \mu_{-1,*}(\mathcal{O}_y \setminus S_y) + \mu_{-1,*}(S_y \setminus S_x)$. The former term is at most $\varepsilon$, by the choice of $\mathcal{O}_y$. Because $I$ is not jumpy, the latter term is $h(y) - h(x) \leq \varepsilon$. Thus,

$$\sum_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mu_*(\mathcal{O}_I \setminus f) \leq \sum_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mu_1(I) \cdot (\mu_{-1,*}(\mathcal{O}_y \setminus S_y) + \mu_{-1,*}(S_y \setminus S_x)) \leq 2\varepsilon \sum_{I \in \boldsymbol{I} \setminus \boldsymbol{J}} \mu_1(I) \leq 4\varepsilon.$$

All in all, we can upper bound the expression in (13) by $2\varepsilon^2(2/\varepsilon) + 4\varepsilon = 8\varepsilon$. ■

# 7 The Monotonicity Tester

In this section we prove our main monotonicity testing results, Theorem 1.1 and Theorem 1.2 (recall the formal statement, Theorem 6.3). We use the following theorem of [BCS18] on monotonicity testing for Boolean functions over $[n]^d$.

**Theorem 7.1** (Theorem 1.1 of [BCS18])**.** *There is a randomized algorithm which, given a parameter $\varepsilon \in (0, 1)$ and a function $f : [n]^d \to \{0, 1\}$, makes $O(d^{5/6} \cdot \log^{3/2} d \cdot (\log n + \log d)^{4/3} \cdot \varepsilon^{-4/3})$ non-adaptive queries to $f$ and (a) always accepts if $f$ is monotone, and (b) rejects with probability $> 2/3$ if $\varepsilon_f > \varepsilon$.*

We refer to the tester of Theorem 7.1 as the `grid-path-tester`. Using this result along with our domain reduction theorems Theorem 1.3 and Theorem 1.4, we design testers for Boolean-valued functions over $[n]^d$ and $\mathbb{R}^d$ (refer to Alg. 1). We restrict to the $\mathbb{R}^d$ case and prove Theorem 1.2 (that is, Theorem 6.3); the proof of Theorem 1.1 is analogous (and the corresponding tester is analogous to Alg. 1). Let $C$ denote the universal constant from Theorem 1.4.

**Proof of Theorem 6.3:** We set $k := (2C \cdot \frac{d}{\varepsilon})^7$ and sample a hypergrid $\boldsymbol{T} = \prod_{i=1}^{d} T_i$, where each $T_i$ is formed by $k$ iid draws from $\mu_i$. By Theorem 1.4, $\mathbf{E}_{\boldsymbol{T}}[\varepsilon_{f_{\boldsymbol{T}}}] \geq \varepsilon - \frac{C \cdot d}{k^{1/7}} = \varepsilon/2$. By Markov's inequality, $\mathbf{Pr}_{\boldsymbol{T}}[\varepsilon_{f_{\boldsymbol{T}}} \geq \varepsilon/4] \geq \varepsilon/4$ and so *at least one* of the iterations of Step 2 returns $\boldsymbol{T}$ satisfying $\varepsilon_{f_{\boldsymbol{T}}} \geq \varepsilon/4$ with

**Algorithm 1** Monotonicity Tester for $f : \mathbb{R}^d \to \{0,1\}$. Inputs: $f$ and $\varepsilon \in (0,1)$.

```
repeat 16/ε times:
    Sample T = T_1 × ··· × T_d as in Theorem 1.4 with k = (2C · d/ε)^7.
    if grid-path-tester(f_T, ε/4, k) returns REJECT, then return REJECT.
return ACCEPT.
```

probability $\geq 1-(1-\varepsilon/4)^{16/\varepsilon} = 1-((1-\varepsilon/4)^{4/\varepsilon})^4 \geq 1-(1/e)^4 \geq 15/16$. Thus, if $\varepsilon_f > \varepsilon$, then Alg. 1 rejects with probability $> \frac{15}{16}\cdot\frac{2}{3} = 5/8$. On the other hand, if $f$ is monotone, then $f_{\boldsymbol{T}}$ is always monotone and so Alg. 1 accepts. For the query complexity, Alg. 1 runs `grid-path-tester` at most $16/\varepsilon$ times with parameters $\varepsilon/4$ and $k = (2C\cdot\frac{d}{\varepsilon})^7$. Thus, substituting these values in place of $\varepsilon$ and $n$ in the query complexity of Theorem 7.1 and multiplying by $16/\varepsilon$ completes the proof. ■

# 8 Lower Bound for Domain Reduction

In this section we prove the following lower bound for the number of uniform samples needed for a domain reduction result to hold for distance to monotonicity. Recall the domain reduction experiment for the hypergrid: 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 each $T_i$ is formed by taking $k$ iid uniform draws from $[n]$ with replacement. We then consider the restriction $f_{\boldsymbol{T}}$.

**Theorem 8.1** (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 $\mathbf{E}_{\boldsymbol{T}}[\varepsilon_{f_{\boldsymbol{T}}}] \leq O(k^2/d)$. In particular, $k = \Omega(\sqrt{d})$ samples in each dimension is necessary to preserve distance to monotonicity.*

## 8.1 Proof of Theorem 8.1

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 $i$th 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 8.2.** *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 $i$th individual is a supporter, and all others are skeptics.
- $\mathcal{F}_i$: The $i$th 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 8.3.** *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 $k$ iid 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 $i$th 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 $i$th 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 8.1 follows from Claim 8.2 and Lemma 8.3.

# 9 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 9.1** (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 iid samples from the uniform distribution on $[n]$, then $\mathbf{E}_{\boldsymbol{T}}[\mathsf{var}(f_{\boldsymbol{T}})] \geq \mathsf{var}(f)/2$.*

*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 $i$th block) corresponds to the $i$th 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 $j$th 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 $i$th 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 $S \subseteq I_i$.

$$\prod_{j \in S} x_j = \prod_{j \in S \cap R_i} b_j \prod_{j \in S \setminus R_i} b_j y_i = y_i^{|S \setminus R_i|} \prod_{j \in S} b_j \tag{14}$$

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

$$\mathbf{E}_{T_i}\left[\mathbf{E}_y\left[\prod_{j \in S} x_j\right]\right] = \mathbf{E}_{R_i, B_i}\left[\mathbf{E}_y\left[y_i^{|S \setminus R_i|} \prod_{j \in S} b_j\right]\right] = \mathbf{E}_{R_i}\left[\mathbf{E}_y\left[y_i^{|S \setminus R_i|}\right]\right] \cdot \mathbf{E}_{B_i}\left[\prod_{j \in S} b_j\right] = 0 \tag{15}$$

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 $\mathbf{E}_y\left[\prod_{j \in S} x_j\right] = 0$.

$$\mathbf{E}_{T_i}\left[\mathbf{E}_y\left[\prod_{j \in S} x_j\right]^2\right] = \mathbf{Pr}_{R_i}\left[|S \setminus R_i| \text{ is even}\right] = 1/2 \tag{16}$$

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 (15).

$$\begin{aligned}
\mathbf{E}_{\boldsymbol{T}}\left[\widehat{f_{\boldsymbol{T}}}(\emptyset)^2\right] &= \mathbf{E}_{\boldsymbol{T}}\left[\sum_S \widehat{f}_S^2 \mathbf{E}_y\left[\chi_S(x)\right]^2 + \sum_{S,T: S \neq T} \widehat{f}_S \widehat{f}_T \mathbf{E}_y\left[\chi_S(x)\right] \mathbf{E}_y\left[\chi_T(x)\right]\right] && (17) \\
&= \sum_S \widehat{f}_S^2 \mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y[\chi_S(x)]^2\right] + \sum_{S,T: S \neq T} \widehat{f}_S \widehat{f}_T \mathbf{E}_{\boldsymbol{T}}\left[\mathbf{E}_y[\chi_S(x)] \mathbf{E}_y[\chi_T(x)]\right] && (18)
\end{aligned}$$

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 (16).

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

The cross terms will be zero, using calculations analogous for (15) (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,

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

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


### Acknowledgements

We would like to thank the anonymous reviewers who have given constructive comments and pointed us to relevant material.


## 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. $L_p$-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.

[DMN18] Anindya De, Elchanan Mossel, and Joe Neeman. Is your data low-dimensional? *arXiv preprint arXiv:1806.10057*, 2018.

[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.

[Nel15] Gail S. Nelson. *A user-friendly introduction to Lebesgue measure and integration*, volume 78. American Mathematical Soc., 2015.

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 6(72):1012–1042, 2006.

[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.

[SS05] Elias Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2005.

[SS08] Michael E. Saks and C. Seshadhri. Parallel monotonicity reconstruction. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008.



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