Directed Isoperimetric Theorems for Boolean Functions on the Hypergrid and an $\widetilde{O}(n\sqrt{d})$ Monotonicity Tester

Hadley Black
University of California, Los Angeles
hablack@ucla.edu

Deeparnab Chakrabarty*
Dartmouth
deeparnab@dartmouth.edu

C. Seshadhri†
University of California, Santa Cruz
sesh@ucsc.edu

Abstract

The problem of testing monotonicity for Boolean functions on the hypergrid, $f : [n]^d \to \{0, 1\}$ is a classic topic in property testing. When $n = 2$, the domain is the hypercube. For the hypercube case, a breakthrough result of Khot-Minzer-Safra (FOCS 2015) gave a non-adaptive, one-sided tester making $\widetilde{O}(\varepsilon^{-2}\sqrt{d})$ queries. Up to polylog d and ε factors, this bound matches the $\widetilde{\Omega}(\sqrt{d})$-query non-adaptive lower bound (Chen-De-Servedio-Tan (STOC 2015), Chen-Waingarten-Xie (STOC 2017)). For any $n > 2$, the optimal non-adaptive complexity was unknown. A previous result of the authors achieves a $\widetilde{O}(d^{5/6})$-query upper bound (SODA 2020), quite far from the $\sqrt{d}$ bound for the hypercube.

In this paper, we resolve the non-adaptive complexity of monotonicity testing for all constant n, up to $\mathrm{poly}(\varepsilon^{-1} \log d)$ factors. Specifically, we give a non-adaptive, one-sided monotonicity tester making $\widetilde{O}(\varepsilon^{-2}n\sqrt{d})$ queries. From a technical standpoint, we prove new directed isoperimetric theorems over the hypergrid $[n]^d$. These results generalize the celebrated directed Talagrand inequalities that were only known for the hypercube.

*Supported by NSF-CAREER award 2041920

†Supported by NSF DMS-2023495, CCF-1740850, 1839317, 1813165, 1908384, 1909790, and ARO Award W911NF1910294.

1 Introduction

Monotonicity testing, especially over hypergrid domains, is one of the most well studied problems in property testing. We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. The set $[n]^d$ is the d-dimensional hypergrid where $\mathbf{x} \in [n]^d$ is a d-dimensional vector with $\mathbf{x}_i \in [n]$. The hypergrid is equipped with the natural partial order $\mathbf{x} \preceq \mathbf{y}$ iff $\mathbf{x}_i \leq \mathbf{y}_i$ for all $i \in [d]$. Note that when $n = 2$, the hypergrid $[n]^d$ is isomorphic to the hypercube $\{0,1\}^d$.

Let $f : [n]^d \to \{0,1\}$ be a Boolean function defined on the hypergrid. The function f is monotone if $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\mathbf{x} \preceq \mathbf{y}$. The Hamming distance between two Boolean functions f and g, denoted as $\Delta(f, g)$, is the fraction of points where they differ. The *distance to monotonicity* of a function $f : [n]^d \to \{0,1\}$ is defined as $\varepsilon_f := \min_{g \text{ monotone}} \Delta(f, g)$. The Boolean monotonicity testing problem on the hypergrid takes parameter ε and oracle access to $f : [n]^d \to \{0,1\}$. The objective is to design a randomized algorithm, called the tester, that accepts a monotone function with probability $\geq 2/3$ and rejects a function f with $\varepsilon_f \geq \varepsilon$ with probability $\geq 2/3$. A tester is one-sided if it accepts a monotone function with probability 1. A tester is non-adaptive if all its queries are made in one round before seeing any responses.

There has been a rich history of results on monotonicity testing over hypergrids, with a significant focus on hypercubes [GGL+00, DGL+99, CS13, CS14a, BRY14a, CST14, CDJS15, CDST15, KMS15, BB16, BCS18, CWX17, BCS20, BKR20, HY22]. We discuss the history more in Section 1.4, but for now, we give the state of the art. For hypercubes, after a long line of work, the breakthrough result [KMS15] of Khot, Minzer, and Safra gave an $\widetilde{O}_\varepsilon(\sqrt{d})$-query non-adaptive, one-sided tester. This result is tight due to a nearly matching $\widetilde{O}(\sqrt{d})$ query-lower bound for non-adaptive testers due to Chen, Waingarten, and Xie [CWX17]. For general hypergrids, the best upper bound is the $\widetilde{O}_\varepsilon(d^{5/6})$-query tester of the authors [BCS18, BCS20].

This $\widetilde{\Omega}(\sqrt{d})$ vs $\widetilde{O}(d^{5/6})$ gap for non-adaptive testers is a tantalizing and important open question in property testing. Even for the domain $[3]^d$, the optimal non-adaptive monotonicity testing bound is unknown. One of the main questions driving our work is:

Are there $\widetilde{O}_\varepsilon(\sqrt{d})$-query monotonicity testers for domains beyond the hypercube?

Directed isoperimetric theorems. The initial seminal work on monotonicity testing, by Goldreich, Goldwasser, Lehman, Ron, and Samorodnitsky [GGL+00] and Dodis, Goldreich, Lehman, Ron, Raskhodnikova and Samorodnitsky [DGL+99] prove the existence of $\widetilde{O}_\varepsilon(d)$-query testers. For almost a decade, it was not clear whether $o(d)$-query testers were possible. In [CS14a], the last two authors gave the first such tester via an exciting connection with *robust directed isoperimetric theorems*. Indeed, all $o(d)$-query testers are achieved through such theorems.

Think of a Boolean function f as the indicator for a subset of the domain. The variance of f, $\mathrm{var}(f)$, is a measure of the volume of the indicated subset. An isoperimetric theorem for a Boolean function relates the variance of f to the "boundary" of the function which correspond to the sensitive edges or/and their endpoints. The deep insight of these theorems comes from sophisticated ways of measuring boundary size, involving both the vertex and edge boundary. A *directed* isoperimetric theorem is an analog where we only measure "up-boundary" formed by monotonicity violations. Rather surprisingly, in the directed case, one can replace the variance as a measure of volume by the distance to monotonicity.

In Table 1, we list some classic isoperimetric results and their directed analogues for the hypercube. For a point $\mathbf{x}$, $I_f(\mathbf{x})$ is the number of sensitive edges incident to $\mathbf{x}$. We use I_f to denote $\mathbf{E}_{\mathbf{x}}[I_f(\mathbf{x})]$, the total influence of f, which the number of sensitive edges in f divided by the domain size 2^d. The quantity Γ_f is the vertex boundary size divided by 2^d. The directed analogues of these, $I^-_f, \Gamma^-_f, I^-_f(\mathbf{x})$, only consider sensitive edges that violate monotonicity.

Undirected Isoperimetry	Directed Isoperimetry
$I_f \geq \Omega(\mathrm{var}(f))$ (*Poincaré inequality, Folklore*)	$I_f^- \geq \Omega(\varepsilon_f)$ (*Goldreich et al.[GGL+00]*)
$I_f \cdot \Gamma_f \geq \Omega(\mathrm{var}(f)^2)$ (*Margulis [Mar74]*)	$I_f^- \cdot \Gamma_f^- \geq \Omega(\varepsilon_f^2)$ (*Chakrabarty, Seshadhri [CS14a]*)
$\mathbf{E}_{\mathbf{x}}\left[\sqrt{I_f(\mathbf{x})}\right] \geq \Omega(\mathrm{var}(f))$ (*Talagrand [Tal93]*)	$\mathbf{E}_{\mathbf{x}}\left[\sqrt{I_f^-(\mathbf{x})}\right] = \Omega(\frac{\varepsilon_f}{\log d})$ (*Khot, Minzer, Safra [KMS15]*)

Table 1: *Boolean hypercube isoperimetry results and their directed analogues. Pallavoor, Raskhodnikova, and Waingarten [PRW22] removed the* $\log d$*-dependence in the directed Talagrand inequality.*

Observe the remarkable parallel between the standard isoperimetric results and their directed versions. The Talagrand inequality is the strongest statement, and implies all other bounds. The directed versions imply the undirected versions, using standard inequalities regarding monotone functions. The [KMS15] $\widetilde{O}_\varepsilon(\sqrt{d})$-query tester is based on the directed Talagrand inequality.

The story for hypergrids is much more complicated. From an isoperimetric perspective, a common approach is to consider the *augmented hypergrid*, wherein we add edges between pairs in the same line. The dimension reduction technique in [DGL+99] used to prove the $\widetilde{O}_\varepsilon(d)$ testers can be thought of as establishing a directed Poincaré inequality . In previous work [BCS18], the authors proved a directed Margulis inequality, which led to the $\widetilde{O}_\varepsilon(d^{5/6})$ query tester. Another motivating question for our work is:

Can the directed Talagrand inequality be generalized beyond the hypercube?

1.1 Main results

We answer both questions mentioned above in the affirmative. To state our results more formally, we begin with some notations. For any $i \in [d]$, we use $\mathbf{e}_i$ to denote the d-dimensional vector which has 1 on the ith coordinate and zero everywhere else. For a dimension i, a pair $(\mathbf{x}, \mathbf{y})$ is called *i-aligned* if $\mathbf{x}$ and $\mathbf{y}$ only differ on their i-coordinate. An *i-line* is a 1D line of n points obtained by fixing all but the ith coordinate.

We define a notion of directed influence of Boolean functions on hypergrids, which generalizes the notion for Boolean functions on hypercubes. In plain English, for a point $\mathbf{x}$ we count the number of *dimensions* in which $\mathbf{x}$ takes part in a violation. We call this the *thresholded negative influence* of $\mathbf{x}$. Note that $\mathbf{x}$ could participate in multiple violations along the same dimension. Throughout this paper, we will be only talking about negative influences of functions on the hypergrid, and thus will often refer to the above as just thresholded influence, and for brevity's sake we also don't use the superscript "$-$" in the notation below to denote the negative aspect.

Definition 1.1 (Thresholded Influence)**.** *Fix $f : [n]^d \to \{0,1\}$ and a dimension $i \in [d]$. Fix a point $\mathbf{x} \in [n]^d$. The thresholded influence of $\mathbf{x}$ along coordinate i is denoted $\Phi_f(\mathbf{x}; i)$, and has value 1 if there exists an i-aligned violation $(\mathbf{x}, \mathbf{y})$. The thresholded influence of $\mathbf{x}$ is $\Phi_f(\mathbf{x}) = \sum_{i=1}^d \Phi_f(\mathbf{x}; i)$.*

Note that the thresholded influence coincides with the hypercube directed influence when $n = 2$. Also note that for any $\mathbf{x}$, $\Phi_f(\mathbf{x}) \in \{0, 1, \ldots, d\}$ and is independent of n. We prove the following theorem, a directed Talagrand theorem for hypergrids, which generalizes the [KMS15] result up to $\log n$ factors.

Theorem 1.2. *Let $f : [n]^d \to \{0,1\}$ be ε-far from monotone.*

$$\mathbf{E}_{\mathbf{x} \in [n]^d}\left[\sqrt{\Phi_f(\mathbf{x})}\right] = \Omega\left(\frac{\varepsilon}{\log n}\right)$$

Robust isoperimetric theorems and monotonicity testing. For the application to monotonicity testing, as [KMS15] showed, a significant strengthening of Theorem 1.2 is required. The weakness of Theorem 1.2, as stated, is that the same violation/influence is "double-counted" at both its endpoints. The LHS can significantly vary depending on whether we choose to only "count" influences at zero-valued or one-valued points, and this is true even on the hypercube. As a simple illustration, consider the function f that is 1 at the all zeros point and 0 everywhere else. Suppose we only count influences at one-valued points. Then the only vertex with any $I_f(\mathbf{x})$ is the all 0's point, and this value is d. Therefore, the Talagrand objective is $\frac{\sqrt{d}}{2^d}$. On the other hand, if we count influences at zero-valued points, then $I_f(\mathbf{x}) = 1$ for the d points $\mathbf{e}_1$ to $\mathbf{e}_d$, and 0 everywhere else. The Talagrand objective counted from zero-valued points is now much larger: $\frac{d}{2^d}$. Therefore, depending on how we count, one can potentially reduce the Talagrand objective, $\mathbf{E}_{\mathbf{x}}[\sqrt{I_f(\mathbf{x})}]$.

[KMS15] define a general way of deciding which endpoint "pays" for a violated edge. Consider a *coloring*[1] $\chi : E \to \{0,1\}$ of every edge $(\mathbf{x},\mathbf{y}) \in E$ of the hypercube to either 0 or 1. Now, given a violated edge $(\mathbf{x},\mathbf{y})$, we use this coloring to decide whose influence this edge contributes towards. More precisely, given this coloring χ, the *colored* directed influence $I^-_{f,\chi}(\mathbf{x})$ of $\mathbf{x}$ is defined as the number of violated edges $(\mathbf{x},\mathbf{y})$ incident on $\mathbf{x}$ which have the same color as $f(\mathbf{x})$. Given a coloring, the *colorful* Talagrand objective equals the expected root colored directed influence. What [KMS15] prove is that no matter what coloring χ one chooses, the Talagrand objective is still large, and in particular $\mathbf{E}_{\mathbf{x}}\left[\sqrt{I^-_{f,\chi}(\mathbf{x})}\right] = \Omega(\frac{\varepsilon_f}{\log d})$.

We define the robust/colorful generalizations of the thresholded negative influence on hypergrids. Consider the *fully augmented hypergrid*, where we put the edge $(\mathbf{x},\mathbf{y})$ if $\mathbf{x}$ and $\mathbf{y}$ differ on only one coordinate. Let E be the set of edges in the fully augmented hypergrid.

Definition 1.3 (Colorful Thresholded Influence)**.** *Fix $f : [n]^d \to \{0,1\}$ and $\chi : E \to \{0,1\}$. Fix a dimension $i \in [d]$ and a point $\mathbf{x} \in [n]^d$. The colorful thresholded negative influence of $\mathbf{x}$ along coordinate i is denoted $\Phi_{f,\chi}(\mathbf{x};i)$, and has value 1 if there exists an i-aligned violation $(\mathbf{x},\mathbf{y})$ such that $\chi(\mathbf{x},\mathbf{y}) = f(\mathbf{x})$, and has value 0 otherwise. The colorful thresholded negative influence of $\mathbf{x}$ is $\Phi_{f,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Phi_{f,\chi}(\mathbf{x};i)$.*

The main result of our paper is a robust directed Talagrand isoperimetry theorem for Boolean functions on the hypergrid. It is a strict generalization of the KMS Talagrand theorem for hypercubes.

Theorem 1.4. *Let $f : [n]^d \to \{0,1\}$ be ε-far from monotone, and let $\chi : E \to \{0,1\}$ be an arbitrary coloring of the edges of the augmented hypergrid.*

$$\mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right] = \Omega\left(\frac{\varepsilon}{\log n}\right)$$

As a consequence of this theorem, we can (up to log factors) resolve the question of non-adaptive monotonicity testing on hypergrids with constant n. We note that the best bound for any $n > 2$ was $\widetilde{O}(d^{5/6})$. Even for the simplest non-hypercube case of $n = 3$, it was open whether the optimal non-adaptive complexity of monotonicity testing is $\sqrt{d}$.

Theorem 1.5. *Consider Boolean functions over the hypergrid, $f : [n]^d \to \{0,1\}$. There is a one-sided, non-adaptive tester for monotonicity that makes $O(\varepsilon^{-2} n\sqrt{d}\log^5(nd))$ queries.*

[1] [KMS15] considered the colorings to be red/blue, but we find the $0,1$-coloring more natural.

The importance of being robust. We briefly explain why the robust Talagrand version is central to the monotonicity testing application. All testers that have a $o(d)$-query complexity are versions of a *path tester*, which can be thought of as querying endpoints of a directed random walk in the hypercube. Consider a function f as the indicator for a set $\mathbf{1}_f$, where the violating edges form the "up-boundary" between $\mathbf{1}_f$ and its complement. To analyze the random walk, we would like to lower bound the probability that a random walk starts in $\mathbf{1}_f$, crosses over the boundary, and stays in $\overline{\mathbf{1}_f}$, that is, the set of 0's. To analyze this, one needs some structural properties in the graph induced by the boundary edges, which [KMS15] express via their notion of a "good subgraph". In particular, one needs that there be a large number of edges, but also that they are regularly spread out among the vertices. It doesn't seem that the "uncolored" Talagrand versions (like Theorem 1.2) are strong enough to prove this regularity, but the robust version can "weed out" high-degree vertices via a definition of a suitable coloring function χ. In short, the robust version of the Talagrand-style isoperimetric theorem is much more expressive. Indeed, these style of robust results have found other applications in distribution testing [CCK+21] as well.

The dependence on n. Given Theorem 1.5, it is natural to ask whether the dependence on n is necessary. Previous *domain reduction* theorems have shown that one can reduce n to $\mathrm{poly}(d)$ in a black box manner [BCS20, HY22]. The monotonicity tester based on the directed Margulis inequality for hypergrids has a logarithmic dependence on n [BCS18]. Combining with domain reduction, we get a $\widetilde{O}(\mathrm{poly}(\varepsilon^{-1})d^{5/6})$-query tester. It is an outstanding open problem to remove the dependence on n from Theorem 1.5. In Section 8, we outline an approach to do so using the directed Talagrand inequality of Theorem 1.4.

1.2 Challenges

We explain the challenges faced in proving Theorem 1.4 and Theorem 1.5. The KMS proof of the directed Talagrand inequality for the hypercube is a tour-de-force [KMS15], and there are many parts of their proof that do not generalize for $n > 2$. We begin by giving an overview of the KMS proof for the hypercube case.

For the time being, let us focus on the uncolored case. For convenience, let $T(f) = \mathbf{E}_{\mathbf{x}}[\sqrt{I_f^-(\mathbf{x})}]$ denote the hypercube directed Talagrand objective for a $f : \{0,1\}^d \to \{0,1\}$. To lower bound $T(f)$, [KMS15] transform the function f to a function g using a sequence of what they call *split* operators. The ith split operator applied to f replaces the ith coordinate/dimension by two new coordinates $(i,+)$ and $(i,-)$. One way to think of the split operator is that takes the $((0,\mathbf{x}_{-i}),(1,\mathbf{x}_{-i}))$ edge and converts it into a square. (Here, $\mathbf{x}_{-i}$ denotes the collection of coordinates in $\mathbf{x}$ skipping $\mathbf{x}_i$.) The "bottom" and "top" corners of the square store the original values of the edge, while the "diagonal" corners store the min and max values (of the edge). The definition of this remarkably ingenious operator ensures that the split function is monotone in $(i,+)$ and anti-monotone in $(i,-)$. The final function $g : \{0,1\}^{2d} \to \{0,1\}$ obtained by splitting on all coordinates has the property that it is either monotone or anti-monotone on all coordinates. That is, g is unate (or pure, as [KMS15] call them), and for such functions the directed Talagrand inequality can be proved via a short reduction to the undirected case.

The utility of the split operator comes from the main technical contribution of [KMS15] (Section 3.4), where it is shown that splitting cannot increase the directed Talagrand objective. This is a "roll-your-sleeve-and-calculate" argument that follows a case-by-case analysis. So, we can lower bound $T(f) \geq T(g)$. Since g is unate, one can prove $T(g) \geq \varepsilon_g$ (the distance of g to monotonicity). But how does one handle ε_g, or g more generally? This is done by relating splitting to classic *switch operator* in monotonicity testing, introduced in [GGL+00]. The switch operator for the ith coordinate can be thought of as modifying the edges along the i-dimension: for any i-edge violation $(\mathbf{x},\mathbf{y})$, this operator switches the values, thereby

fixing the violation. The switching operator has the remarkable property of never increasing monotonicity violations in other dimensions; hence, switching in all dimensions leads to a monotone function. [KMS15] observe that the function g basically "embeds" disjoint variations of f, wherein each variation is obtained by performing a distinct sequence of switches on f. The function g contains all possible such variations of f, stored cleverly so that g is unate. One can then use properties of the switch operators to relate ε_g to ε_f. (The truth is more complicated; we will come back to this point later.)

Challenge #1, splitting on hypergrids? The biggest challenge in trying to generalize the [KMS15] argument is to generalize the split operator. One natural starting point would be to consider the *sort* operator, defined in [DGL+99], which generalizes the switch operator: the sort operator in the ith coordinate sorts the function along all i-lines. But it is not at all clear how to split the ith coordinate into a set of coordinates that contains the information about the sort operator thereby leading to a pure/unate function. In short, sorting is a much more complicated operation than switching, and it is not clear how to succinctly encode this information using a single operator.

We address this challenge by a reorientation of the KMS proof. Instead of looking at operators on dimensions to understand effects of switching/sorting, we do these via what we call "tracker functions" which are n^d different Boolean functions tracking the changes in f. We discuss this more in Section 1.3.

Challenge #2, the case analysis for decreasing Talagrand objective. As mentioned earlier, the central calculation of KMS is in showing that splitting does not increase the directed Talagrand objective. This is related (not quite, but close enough) to showing that the switch operator does not increase the Talagrand objective. A statement like this is proven in KMS by case analysis; there are 4 cases, for the possible values a Boolean function takes on an edge. One immediately sees that such an approach cannot scale for general n, since the number of possible Boolean functions on a line is 2^n. Even with our new idea of new idea of tracking functions, we cannot escape this complexity of arguing how the Talagrand-style objective decreases upon a sorting operation, and a case-by-case analysis depending on the values of the function is infeasible.

We address this challenge by a connection to the theory of majorization. We show that the sort operator is (roughly) a majorizing operator on the vector of influences. The concavity of square root function implies that sorting along lines cannot increase the Talagrand objective. More details are given in the next section.

Challenge #3, the colorings. Even if we circumvented the above issues, the robust colored Talagrand objective brings a new set of issues. Roughly speaking, colorings decide which points "pay" for violations of the Talagrand objective, the switching/sorting operator move points around by changing values, and the high-level argument to prove $T(f)$ drops is showing that these violations "pay" for the moves. In the hypercube, a switch either changes the values on all the points of the edge or none of the points, and this binary nature makes the handling of colors in the KMS proof fairly easy, merely introducing a few extra cases in their argument. Sorting, on the other hand, can change an arbitrary set of points, and in particular, even in the case of $n = 3$, a point participating in a violation may not change value in a sort.

To address this challenge, as we apply the sort operators to obtain a handle on our function, we also need to *recolor* the edges such that we obtain the drop in the T-objective. Once again, the theory of majorization is the guide. This part of the proof is perhaps the most technical portion of our paper.

Other minor challenges: the telescoping argument and tester analysis: The issues detailed here are not really conceptual challenges, but they do require some work to handle the richer hypergrid domain.

Recall that the KMS analysis proves the chain of inequalities, $T(f) \geq T(g) = \Omega(\varepsilon_g)$. Unfortunately, it can happen that $\varepsilon_g \ll \varepsilon_f$. In this case, KMS observe that one could redo the entire argument on random restrictions of f to half the coordinates. If the corresponding ε_g is still too small, then one restricts on one-fourth of the coordinates, so on and so forth. One can prove that somewhere along these $\log d$ restrictions, one must have $\varepsilon_g = \Omega(\varepsilon_f)$. Pallavoor, Raskhodnikova, and Waingarten [PRW22] improve this analysis to

remove a $\log d$ loss from the final bound. We face the same problems in our analysis, and have to adapt the analysis to our setting.

Finally, the tester analysis of KMS for the hypercube can be ported to the hypergrid path tester, with some suitable adaptations of their argument. It is convenient to think of the *fully augmented hypergrid*, where all pairs that lie along a line are connected by an edge. We can essentially view the hypergrid tester as sampling a random hypercube from the fully augmented hypergrid, and then performing a directed random walk on this hypercube. We can then piggyback on various tools from KMS for the hypercube tester, to bound the rejection probability of the path tester for hypergrids.

1.3 Main Ideas

We sketch some key ideas needed to prove Theorem 1.4 and address the challenges detailed earlier. We begin with a key conceptual contribution of this paper. Given a function $f : [n]^d \to \{0,1\}$, we define a collection of Boolean functions on the hypercube called *tracker functions*. We will lower bound the directed Talagrand objective on the hypergrid by the undirected Talagrand objective on these tracker functions. Indeed, the inspiration of these tracker functions arose out of understanding the analysis in [KMS15], in particular, the intermediate "g" function in their Section 4. As an homage, we also denote our tracker functions with the same Roman letter, even though it is different from their function.

1.3.1 Tracker functions $g_{\mathbf{x}}$ for all $\mathbf{x} \in [n]^d$

Let us begin with the sort operator discussed earlier. Without loss of generality, fix the ordering of coordinates in $[d]$ to be $(1, 2, \ldots, d)$. The operator $\texttt{sort}_i$ for $i \in [d]$ sorts the function on every i-line. Given a subset $S \subseteq [d]$ of coordinates, the function $(S \circ f)$ is obtained by sorting the coordinates in S in that order.

Sorting along any dimension cannot increase the number of violations along any other dimension, and therefore upon sorting on all dimensions, the result is a monotone function [DGL+99]. Suppose f is ε-far from monotone. Clearly, the total number of points changed by sorting along all dimensions must be at least εn^d. While this is not obvious here, it will be useful to to *track* how the function value changes when we sort along a certain subset S of coordinates. The intuitive idea is: if the function value changes for most such partial sortings, then perhaps the function is far from being monotone. To this end, for every point $\mathbf{x} \in [n]^d$, we define a Boolean function $g_{\mathbf{x}} : 2^{[d]} \to \{0,1\}$ that tracks how the function value f changes as we apply the sort operator a subset S of the coordinates. It is best to think of the domain of $g_{\mathbf{x}}$ as subsets $S \subseteq [d]$.

Definition 1.6 (Tracker Functions $g_{\mathbf{x}}$)**.** *Fix an $\mathbf{x} \in [n]^d$. The tracker function $g_{\mathbf{x}} : \{0,1\}^d \to \{0,1\}$ is defined as*

$$\forall S \subseteq [d], \quad g_{\mathbf{x}}(S) := (S \circ f)(\mathbf{x})$$

We provide an illustration of this definition in Figure 1.

Note that when f is a monotone function, all the functions $g_{\mathbf{x}}$ are constants. Sorting does not change any values, so $g_{\mathbf{x}}(S)$ is always $f(\mathbf{x})$. On the other hand, if f is not monotone along dimension i, then there are points such that $g_{\mathbf{x}}(\{i\}) \neq f(\mathbf{x})$. Indeed, one would expect the typical variance of these $g_{\mathbf{x}}$ functions to be related to the distance to monotonicity of f (technically not true, but we come to this point later).

The tracker functions help us lower bound the (colorful) Talagrand objective for thresholded influence, in particular, the LHS in Theorem 1.4. Recall that the Talagrand objective is the expected square root of the colorful thresholded influences on the hypergrid function f. We lower bound this quantity by the expected Talagrand objective on the *undirected* (colorful, however) influence of the various $g_{\mathbf{x}}$ functions. Note that $g_{\mathbf{x}}$ functions are defined on hypercubes. So we reduce the robust directed Talagrand inequality on hypergrids

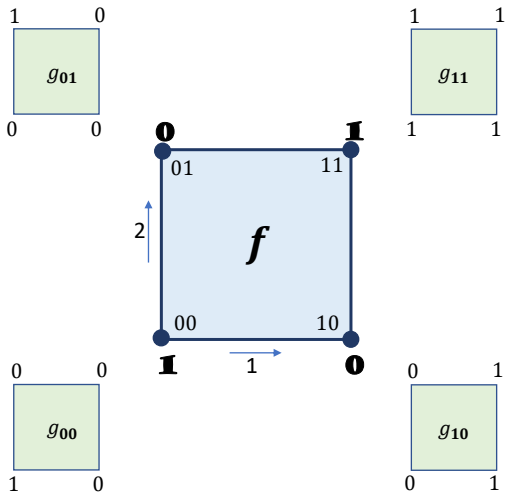


Figure 1: *The blue function* $f : [n]^d \to \{0,1\}$ *is defined in the middle using bold, gothic characters. We have* $d = 2$ *and* $n = 2$*. For each of the* 4 *points of this square, we have four different* $g_{\mathbf{x}} : \{0,1\}^2 \to \{0,1\}$ *and they are described in the four green squares. For any* $S \subseteq \{1,2\}$*, if we focus on the corresponding corners of the four squares, then we get the function* $(S \circ f)$*. For instance, if* $S = \{2\}$*, then if we focus on the top left corners, then starting from* g_{00} *and moving clockwise we get* $(0,1,1,0)$*. These will precisely the function* f *(read clockwise from* 00*) after we sort along dimension* 2*.*

to robust undirected Talagrand inequalities on hypercubes. This is the main technical contribution of our paper. Let us define the (colored) influences of these $g_{\mathbf{x}}$ functions.

Definition 1.7 (Influence of the Tracking Functions)**.** *Fix a* $\mathbf{x} \in [n]^d$ *and consider the tracking function* $g_{\mathbf{x}} : \{0,1\}^d \to \{0,1\}$*. Fix a coordinate* $j \in [d]$*. The influence of* $g_{\mathbf{x}}$ *at a subset* S *along the* j*th coordinate is defined as*

$$I^{=j}_{g_{\mathbf{x}}}(S) = 1 \ \textit{iff} \ g_{\mathbf{x}}(S) \neq g_{\mathbf{x}}(S \oplus j) \quad \textit{that is} \quad (S \circ f)(\mathbf{x}) \neq (S \oplus j \ \circ f)(\mathbf{x})$$

In plain English, the influence of the jth coordinate at a subset S is 1 if the function value (the hypergrid function) changes when we include the dimension j to be sorted. Once again, note that the same sensitive edge $(S, S \oplus j)$ is contributing towards both $I^{=j}_{g_{\mathbf{x}}}(S)$ and $I^{=j}_{g_{\mathbf{x}}}(S \oplus j)$. We define a robust, colored version of these influences.

Definition 1.8 (Colorful Influence of the Tracking Functions)**.** *Fix a* $\mathbf{x} \in [n]^d$ *and consider the tracking function* $g_{\mathbf{x}} : \{0,1\}^d \to \{0,1\}$*. Fix any arbitrary coloring* $\xi_{\mathbf{x}} : E(2^{[d]}) \to \{0,1\}$ *of the Boolean* hypercube*. Fix a coordinate* $j \in [d]$*. The influence of* $g_{\mathbf{x}}$ *at a subset* S *along the* j*th coordinate is defined as*

$$I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) = 1 \ \textit{iff} \ g_{\mathbf{x}}(S) \neq g_{\mathbf{x}}(S \oplus j) \quad \textbf{\textit{and}} \quad g_{\mathbf{x}}(S) = \xi_{\mathbf{x}}(S, S \oplus j)$$

The colorful total influence at the point S *in* $g_{\mathbf{x}}$ *is defined as*

$$I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) := \sum_{j=1}^{d} I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)$$

As before, for a sensitive edge $(S, S \oplus j)$ of $g_{\mathbf{x}}$, we count it towards the influence of the endpoint whose value equals the color $\xi_{\mathbf{x}}(S, S \oplus j)$. The main technical contribution of this paper is proving that for any function $f : [n]^d \to \{0,1\}$ and any arbitrary coloring $\chi : E \to \{0,1\}$ of the hypergrid edges, for every $\mathbf{x} \in [n]^d$ there exists a coloring $\xi_{\mathbf{x}} : E(2^{[d]}) \to \{0,1\}$ of the Boolean *hypercube* edges, such that

$$T_{\Phi_\chi}(f) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right] \gtrapprox \mathbf{E}_{\mathbf{x}\in[n]^d}\mathbf{E}_{S\subseteq[d]}\left[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}\right] \tag{H1}$$

We explain the $\approx$ in the above inequality in the next subsection.

Why is a statement like (H1) useful? Because the RHS terms are Talagrand objectives on colored influences on the usual undirected hypercube. Therefore, we can apply undirected Talagrand bounds (known from KMS, Theorem 2.8) to get an upper bound on the variance.

Corollary 1.9 (Corollary of Theorem 1.8 in [KMS15])**.** *Fix $f : [n]^d \to \{0,1\}$. Fix an $\mathbf{x} \in [n]^d$ and consider the tracking function $g_{\mathbf{x}} : \{0,1\}^d \to \{0,1\}$. Consider any* arbitrary *coloring $\xi_{\mathbf{x}} : E(2^{[d]}) \to \{0,1\}$ of the Boolean* hypercube*. Then, for every $\mathbf{x} \in [n]^d$, we have*

$$\mathbf{E}_{S\subseteq[d]}\left[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}\right] = \Omega(\mathrm{var}(g_{\mathbf{x}}))$$

The final piece of the puzzle connects $\mathrm{var}(g_{\mathbf{x}})$'s with the distance to monotonicity. Ideally, we would have liked to have a statement such as the following true.

$$\mathbf{E}_{\mathbf{x}\in[n]^d}\left[\mathrm{var}(g_{\mathbf{x}})\right] \approx \Omega(\varepsilon_f) \tag{H2}$$

We now see that (H1), Corollary 1.9, and (H2) together implies Theorem 1.4 (indeed without the $\log n$).

1.3.2 High level description of our approaches

Addressing the $\approx$ in (H1) via semisorting. As stated, we do not know if (H1) is true. However, we establish (H1) for *semisorted* functions $f : [n]^d \to \{0,1\}$. A function f is semisorted if on any line ℓ, the restriction of the function on the first half is sorted and the restriction on the second half is sorted. This may seem like a simple subclass of functions, but note that all functions on the Boolean hypercube ($n = 2$) are vacuously semisorted. Thus, proving Theorem 1.4 on semi-sorted functions is already a generalization of the [KMS15] result. Theorem 3.2 is the formal restatement of (H1).

We reduce Theorem 1.4 on general functions to the same bound for semisorted functions. Consider semisorting f, which means we sort f on each half of every line. Suppose the Talagrand objective did not increase *and* the distance to monotonicity did not decrease. Then Theorem 1.4 on the semisorted version of f implies Theorem 1.4 on f. What we can prove is that: given the semisorted function, one can find a *recoloring* of the hypergrid edges such that the Talagrand objective doesn't increase. The precise statement is given in Lemma 3.1. We comment on our techniques to prove such a statement in a later paragraph.

Although semisorting can't increase the Talagrand objective, it can clearly reduce the distance to monotonicity. However, a relatively simple inductive argument proves Theorem 1.4 with a $\log n$ loss. Any function can be turned into a completely sorted (aka monotone) function by performing "$\log n$ semisorting steps" at varying scales. In each scale, we consider many disjoint small hypergrids, and convert a semisorted function defined over a small hypergrid to another semisorted function over a hypergrid of double the size (the next scale). In one of these scales, we will find a semisorted function that has $\Omega(\varepsilon/\log n)$ distance from its sorted version. One can average Theorem 1.4 over all the small hypergrids at this scale to bound the Talagrand objective of the whole function by $\Omega(\varepsilon/\log n)$. This is the step where we incur the $\log n$-factor loss. This argument is not complicated, and we provide illustrated details in Section 3.

The real work happens in proving Theorem 3.3, that is, (H1) for semisorted functions.

Approach to proving (H1) for semisorted functions. Recall, we have a fixed adversarial coloring $\chi : E \to \{0,1\}$. The proof follows a "hybrid argument" where we define a potential that is modified over $d+1$ rounds. At the beginning of round 0 it takes the value $\mathbf{E}_{\mathbf{x}\in[n]^d}[\sqrt{\Phi_{f,\chi}(\mathbf{x})}]$ which is the LHS of (H1). At the end of round d it takes the value $\mathbf{E}_{\mathbf{x}\in[n]^d}\mathbf{E}_{S\subseteq[d]}[\sqrt{I_{g_\mathbf{x},\xi_\mathbf{x}}(S)}]$ which is the RHS of (H1). The proof follows by showing that the potential decreases in each round.

Let us describe the potential. Let us first write this without any reference to the colorings (so no χ's and $\xi_\mathbf{x}$'s), and then subsequently address the colorings. At stage i, fix a subset $S \subseteq [i]$. Define

$$R_i(S) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\sum_{j=1}^{i} I^{=j}_{g_\mathbf{x}}(S) \;+\; \sum_{j=i+1}^{d} \Phi_{S\circ f}(\mathbf{x};j)}\right] \tag{Hybrid}$$

We remind the reader that $S \circ f$ is the function f after the dimensions corresponding to $i \in S$ have been sorted. Thus, $R_i(S)$ is a "hybrid" Talagrand objective, with two different kinds of influences being summed. Consider point $\mathbf{x} \in [n]^d$. On the first i coordinates, we sum the undirected influence (along these coordinates) of S on the function $g_\mathbf{x}$. On the coordinates $i+1$ to d, we sum to directed influence along these coordinates in the function $S \circ f$. The potential is $\Lambda_i := \mathbf{E}_{S\subseteq[i]}[R_i(S)]$.

To make some sense of this, consider the extreme cases of $i = 0$ and $i = d$. When $i = 0$, we only have the second $\Phi_{S\circ f}$ term. Furthermore, S is empty since $S \subseteq [i]$. So Λ_0 is precisely the original directed Talagrand objective, the LHS of (H1). When $i = d$, we only have the $I^{=j}_{g_\mathbf{x}}$ terms. Taking expectation over $S \subseteq [d]$ to get Λ_d, we deduce that Λ_d is the RHS of (H1).

We will prove $\Lambda_{i-1} \geq \Lambda_i$ for all $1 \leq i \leq d$. To choose a uar set in $[i]$, we can choose a uar subset of $[i-1]$ and then add i with $1/2$ probability. Hence, $\Lambda_i = (\mathbf{E}_{S\subseteq[i-1]}[R_i(S) + R_i(S+i)])/2$, while $\Lambda_{i-1} = \mathbf{E}_{S\subseteq[i-1]}[R_{i-1}(S)]$. So, if we prove that $R_{i-1}(S)$ is at least than both $R_i(S)$ and $R_i(S+i)$, then $\Lambda_{i-1} \geq \Lambda_i$. The bulk of the technical work in this paper is involved in proving these two inequalities, so let us spend a little time explaining what proving this entails.

Let's take the inequality $R_{i-1}(S) \geq R_i(S)$. Refer again to (Hybrid). When we go from $R_{i-1}(S)$ to $R_i(S)$, under the square root, the term $\Phi_{S\circ f}(\mathbf{x};i)$ is replaced by $I^{=i}_{g_\mathbf{x}}(S)$. To remind the reader, the former term is the indicator of whether $\mathbf{x}$ participates in a i-violation after the coordinates in $S \subseteq [i-1]$ have been sorted. The latter term is whether $g_\mathbf{x}(S+i)$ equals $g_\mathbf{x}(S)$, that is, whether the (hypergrid) function value at $\mathbf{x}$ changes between sorting on coordinates in S and $S+i$. Just by parsing the definitions, one can observe that $\Phi_{S\circ f}(\mathbf{x};i) \geq I^{=i}_{g_\mathbf{x}}(S)$; if a point is modified on sorting in the i-coordinate, then it must be participating in some i-violation (note that vice-versa may not be true and thus we have an inequality and not an equality). The quantity under the square-root *point-wise* dominates (ie, for every $\mathbf{x}$) when we move from $R_{i-1}(S)$ to $R_i(S)$. Thus, the $R_{i-1}(S) \geq R_i(S)$.

The other inequality $R_{i-1}(S) \geq R_i(S+i)$, however, is much trickier to establish. In $R_i(S+i)$, the second summation under the square-root, the Φ terms, are actually on a *different* function. The $\Phi_{S\circ f}(\mathbf{x};j)$ terms in $R_{i-1}(S)$ are the thresholded influences of the function after sorting on coordinates in S. But in $R_i(S+i)$, these terms are $\Phi_{(S+i)\circ f}(\mathbf{x};j)$, the thresholded influences of $\mathbf{x}$ for the function after sorting on $S+i$. Although, it is true that sorting on more coordinates cannot increase the total number of violations along any dimension, this fact is *not* true point-wise. So, a point-wise argument as in the previous inequality is not possible.

The argument for this inequality proceeds *line-by-line*. One fixes an i-line ℓ and considers the vector of "hybrid function" values on this line. We then consider this vector when moving from $R_{i-1}(S)$ to $R_i(S+i)$, and we need to show that the *sum of square roots* can get only smaller. This is where one of our key insights comes in: the theory of majorization can be used to assert these bounds. Roughly speaking, a vector $\mathbf{a}$

(weakly) majorizes a vector $\mathbf{b}$ if the sum of the k-largest coordinates of $\mathbf{a}$ dominates the sum of the k-largest coordinates of $\mathbf{b}$, for every k. A less balanced vector majorizes a more balanced vector. If the ℓ_1-norms of these vectors are the same, then the sum of square roots of the entries of $\mathbf{a}$ is at most the sum of square roots of that of $\mathbf{b}$. This follows from concavity of the square-root function.

Our overarching mantra throughout this paper is this: whenever we perform an operation and the hybrid-influence-vector induced by a line changes, the new vector majorizes the old vector. Specifically, these vectors are generated by look at the terms of $R_{i-1}(S)$ and $R_i(S+i)$ restricted to i-lines.

To prove this vector-after-operation majorizes vector-before-operation, we need some structural assumptions on the function. Otherwise, it's not hard to construct examples where this just fails. The structure we need is precisely the *semisortedness* of f. When a function is semisorted, then majorization argument goes through. At a high level, when f is semisorted, the vector of influences (along a line) satisfy various monotonicity properties. In particular, when we (fully) sort on some coordinate i, we can show the points losing violations had low violations to begin with. That is, the vector of violations becomes less balanced, and the majorization follows.

The above discussion disregarded the colors. With colors, the situation is noticeably more difficult. Although the function f is assumed to be semisorted, the coloring $\chi : E \to \{0,1\}$ is adversarial. So even though the vector of influences may have monotonicity properties, the colored influences may not have this structure. So a point with high influence could have much lower colored influence. Note that the sort operator is insensitive to the coloring. So the majorization argument discussed above might not hold when looking at colored influences.

With colors, (Hybrid) is replaced by the actual quantity (Colorful Hybrid) described in Section 5. To carry out the majorization argument, we need to construct a family of colorings $\xi_{\mathbf{x}}$ on the n^d different hypercubes. We also need 2^d many different auxiliary colorings χ_S of the hypergrid, constructed after every sort operation. The argument is highly technical. But all colorings are chosen to follow our mantra: vector after operation should majorize vector before operation. The same principle is also used to prove Lemma 3.1 which claims that semisorting an interval can only decrease the Talagrand objective, after a recoloring.

The details of the actual $R_i(S)$ hybrid function and the strategy to use them is presented in Section 5. The most technical part of the paper is in Section 6, which proves that the potential decreases in each round.

Addressing the $\approx$ in (H2) via random sorts. To finally complete the argument, we need (H2) that relates the average variance of the $g_{\mathbf{x}}$ functions to the distance to monotonicity of f. As discussed earlier, (H2) is false, even for the case of hypercubes. Nevertheless, one can use (H1) and Corollary 1.9 to prove a lower bound on $T_{\Phi_\chi}(f)$ with respect to ε_f. This is the telescoping argument of KMS, refined in [PRW22]. We describe the main ideas below. The first observation (see Theorem 4.1) is that $\mathbf{E}_{\mathbf{x}\in[n]^d}[\mathrm{var}(g_{\mathbf{x}})]$ is roughly $\mathbf{E}_S[\Delta(S\circ f, \overline{S}\circ f]$ where S is a uniform random subset of coordinates. The distance to monotonicity ε_f is approximated by $\Delta\left(f, S\circ\overline{S}\circ f\right)$ which, by triangle inequality, is at most $\Delta(f, S\circ f) + \Delta(S\circ f, \overline{S}\circ f)$. Thus, we get a relation between ε_f, the expected $\mathrm{var}(g_{\mathbf{x}})$, and the distance between f and a "random sort" of f. Therefore, if (H2) is not true, then a random sort of f must be still far from being monotone, and then one can repeat the whole argument on just this random sort itself. In one of these $\log d$ "repetitions", the (H2) must be true since in the end we get a monotone function (which can't be far from being monotone). And this suffices to establish Theorem 1.4. We re-assert that the main ideas are already present in [KMS15, PRW22]. However, we require a more general presentation to make things work for hypergrids. These details can be found in Section 4.

1.4 Related Work

Monotonicity testing has seen much activity since its introduction around 25 years ago [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, BCS20, BKR20, HY22].

We have already covered much of the previous work on Boolean monotonicity testing over the hypercube, but give a short recap. For convenience of presentation, in some results, we subsume ε-dependencies using the notation O_ε. The problem was introduced by Goldreich et al. [GGL+00] and Raskhodnikova [Ras99], who described an $O(d/\varepsilon)$-query tester. Chakrabarty and Seshadhri [CS14a] achieved the first sublinear in dimension query complexity of $\widetilde{O}_\varepsilon(d^{7/8})$ using directed isoperimetric inequalities. Chen, Servedio, and Tan [CST14] improved the analysis to $\widetilde{O}_\varepsilon(d^{5/6})$ queries. Fischer et al. [FLN+02] had first shown an $\Omega(\sqrt{d})$-query lower bound for non-adaptive, one-sided testers, by a short and neat construction. The non-adaptive, two-sided $\widetilde{\Omega}(\sqrt{d})$ lower bound is much harder to attain, and was done by Chen, Waingarten, and Xie [CWX17], improving on the $\Omega(d^{1/2-c})$ bound from [CDST15], which itself improved on the $\widetilde{\Omega}(d^{1/5})$ bound of [CST14]. [KMS15] gave an $\widetilde{O}_\varepsilon(\sqrt{d})$-query tester, via the robust directed Talagrand inequality.

While this resolves the non-adaptive testing complexity (up to $\mathrm{poly}(\varepsilon^{-1}\log d)$ factors) for the hypercube, the adaptive complexity is still open. The first polynomial lower bound of $\widetilde{\Omega}(d^{1/4})$ for adaptive testers was given by Belovs and Blais [BB16] and has since been improved to $\widetilde{\Omega}(d^{1/3})$ by Chen, Waingarten, and Xie [CWX17]. Chakrabarty and Seshadhri [CS19] gave an adaptive $\widetilde{O}_\varepsilon(I_f)$-query tester, thereby showing that adaptivity can help in monotonicity testing. The $d^{1/3}$ vs $\sqrt{d}$ query complexity gap is an outstanding open question in property testing.

There has been work on approximating the distance to monotonicity in $\mathrm{poly}(d, \varepsilon_f)$-queries. Fattal and Ron [FR10] gave the first non-trivial result of an $O(d)$-approximation, and Pallavoor, Raskhodnikova, and Waingarten [PRW22] gave a non-adaptive $O(\sqrt{d})$-approximation (all running in $\mathrm{poly}(d, \varepsilon_f)$ time). They also show that non-adaptive $\mathrm{poly}(d)$-time algorithms cannot beat this approximation factor.

The above discussion is only for Boolean valued functions on the hypercube. For arbitrary ranges, the original results on monotonicity testing gave an $O(d^2/\varepsilon)$-query tester [GGL+00, DGL+99]. Chakrabarty and Seshadhri [CS13] proved that $O(d/\varepsilon)$-queries suffices for monotonicity testing, matching the lower bound of $\Omega(d)$ of Blais, Brody, and Matulef [BBM12]. The latter bound holds even when the range size is $\sqrt{d}$. A recent result of Black, Kalemaj, and Raskhodnikova showed a smooth trade-off between the $\sqrt{d}$ bound for the Boolean range and the d bound for arbitrary ranges ([BKR20]). Consider functions $f : \{0,1\}^d \to [r]$. They gave a tester with query complexity $\widetilde{O}_\varepsilon(r\sqrt{d})$, achieved by extending the directed Talagrand inequality to arbitrary range functions. Their techniques are quite black-box and carry over to other posets. We note that their techniques can also be ported to our setting, so we can get an $\widetilde{O}_\varepsilon(rn\sqrt{d})$-query monotonicity tester for functions $f : [n]^d \to [r]$.

We now discuss monotonicity testing on the hypergrid. We discuss more about the ε-dependencies, since there have been interesting relevant discoveries. As mentioned above, [DGL+99] gives 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 showed an interesting adaptivity gap for 2D functions $f : [n]^2 \to \{0,1\}$: there exists an $O(1/\varepsilon)$-query adaptive tester (in fact, for any constant dimension d), and they show an $\Omega(\log(1/\varepsilon)/\varepsilon)$ lower bound for non-adaptive testers. Previous work [BCS18] by the authors gave an $\widetilde{O}_\varepsilon(d^{5/6}\log n)$-query tester, by proving a directed Margulis inequality on augmented hypergrids. Another work [BCS20] of the authors, and subsequently a work [HY22] by Harms and Yoshida, designed domain reduction methods for monotonicity testing, showing how n can be reduced to $\mathrm{poly}(\varepsilon^{-1}, d)$ by subsampling the hypergrid.

For hypergrid functions with arbitrary ranges, the optimal complexity is known to be $\Theta(d \log n)$ [CS13, CS14b]. When the range is $[r]$ and $d = 1$, one can get $O(\log r)$-query testers [PRV18]. As we mentioned earlier, the techniques of [BKR20], combined with our Theorem 1.4, yield a $\widetilde{O}_\varepsilon(rn\sqrt{d})$ query tester for functions $f\colon [n]^d \to [r]$.

2 Preliminaries

A central construct in our proof is the *sort* operator.

Definition 2.1. *Consider a Boolean function on the line $h : [n] \to \{0,1\}$. The sort operator $\mathtt{sort}()$ is defined as follows.*

$$\mathtt{sort}(h)(b) = \begin{cases} 0 & \text{if } b < n - \|h\|_1 \\ 1 & \text{if } b \geq n - \|h\|_1 \end{cases}$$

Thus, the sort operator "moves" the values on a line to ensure that it is sorted. Note that $\mathtt{sort}(h)$ and h have exactly the same number of zero/one valued points. We can now define the sort operator for any dimension i. This operator takes a hypergrid function and applies the sort operator on every i-line.

Definition 2.2. *Let i be a dimension and $f : [n]^d \to \{0,1\}$. The sort operator for dimension i, $\mathtt{sort}_i()$, is defined as follows. For every i-line ℓ, $\mathtt{sort}_i(f)|_\ell = \mathtt{sort}(f|_\ell)$.*

Let S be an ordered list of dimensions, denoted $(i_1, i_2, \ldots, i_k)$. The function $S \circ f$ is obtained by applying the $\mathtt{sort}_i()$ operator in the order given by S. Namely,

$$S \circ f = \mathtt{sort}_{i_k}(\mathtt{sort}_{i_{k-1}}(\ldots \mathtt{sort}_{i_1}(f)))$$

Somewhat abusing notation, we will treat the ordered list of dimensions S as a set, with respect to containing elements. The key property of the sort operator is that it preserves the sortedness of *other* dimensions.

Claim 2.3. *The function $S \circ f$ is monotone along all dimensions in S.*

Proof. We will prove the following statement: if f is monotone along dimension i, then $\mathtt{sort}_j(f)$ is monotone along both dimensions i and j. A straightforward induction (which we omit) proves the claim.

By construction, the function $\mathtt{sort}_j(f)$ is monotone along dimension j. Consider two arbitrary points $\mathbf{x} \preceq \mathbf{x}'$ that are i-aligned (meaning that they only differ in their i-coordinates). We will prove that $\mathtt{sort}_j(f)(\mathbf{x}) \leq \mathtt{sort}_j(f)(\mathbf{x}')$, which will prove that $\mathtt{sort}_j(f)$ is monotone along dimension i.

For convenience, let the j-lines containing $\mathbf{x}$ and $\mathbf{x}'$ be ℓ and ℓ', respectively. Note that these j-lines only differ in their i-coordinates. Let c denote the j-coordinate of $\mathbf{x}$ (and $\mathbf{x}'$). Observe that $\mathtt{sort}_j(f)\mathbf{x} = \mathtt{sort}_j(f)|_\ell(c)$ (analogously for $\mathbf{x}'$).

Note that, $\forall c \in [n]$, $f|_\ell(c) \leq f|_{\ell'}(c)$. This is because f is monotone along dimension i, and ℓ has a lower i-coordinate than that of ℓ'. Hence, $\|f|_\ell\|_1 \leq \|f|_{\ell'}\|_1$. By the definition of the sort operator, $\forall c \in [n], \mathtt{sort}(f|_\ell)(c) \leq \mathtt{sort}(f|_{\ell'})(c)$. Thus, $\mathtt{sort}_j(f)|_\ell(c) \leq \mathtt{sort}_j(f)|_{\ell'}(c)$, implying $\mathtt{sort}_j(f)(\mathbf{x}) \leq \mathtt{sort}_j(f)(\mathbf{x}')$. □

A crucial property of the sort operator is that it can never increase the distance between functions. This property, which was first established in [DGL+99] (Lemma 4), will be used in Section 4, where we apply our main isoperimetric theorem on random restrictions.We provide a proof for completeness.

Claim 2.4. *Let $f, f' : [n]^d \to \{0,1\}$ be two Boolean functions. For any ordered set $S \subseteq [d]$,*

$$\Delta(S \circ f, S \circ f') \leq \Delta(f, f')$$

Proof. It suffices to prove this bound when S is a singleton. We prove that for any $i \in [d]$, $\Delta(\mathtt{sort}_i(f), \mathtt{sort}_i(f')) \leq \Delta(f, f')$. In the following, we will use the simple fact that for monotone functions $h, h' : [n] \to \{0,1\}$, $\Delta(h, h') = \Big| \|h\|_1 - \|h'\|_1 \Big|$. Also, we use the equality $\|\mathtt{sort}(h)\|_1 = \|h\|_1$.

$$\begin{aligned}
\Delta(\mathtt{sort}_i(f), \mathtt{sort}_i(f')) &= \sum_{\ell \ i\text{-line}} \Delta(\mathtt{sort}_i(f)|_\ell, \mathtt{sort}_i(f')|_\ell) = \sum_\ell \Big| \|\mathtt{sort}_i(f)|_\ell\|_1 - \|\mathtt{sort}_i(f')|_\ell\|_1 \Big| \\
&= \sum_\ell \Big| \|f|_\ell\|_1 - \|f'|_\ell\|_1 \Big| \\
&= \sum_\ell \Big| \sum_{c \in [n]} f|_\ell(c) - \sum_{c \in [n]} f|_\ell(c) \Big| \\
&\leq \sum_\ell \sum_{c \in [n]} \Big| f|_\ell(c) - f'|_\ell(c) \Big| = \Delta(f, f')
\end{aligned}$$

□

The method of obtaining a monotone function via repeated sorting is close to being optimal. For hypercubes, this result was established by [FR10] (Lemma 4.3) and also present in [KMS15] (Lemma 3.5). The proofs goes through word-for-word applied to hypergrids.

Claim 2.5. *For any function $f : [n]^d \to \{0,1\}$,*

$$\varepsilon_f \leq \Delta(f, [d] \circ f) \leq 2\varepsilon_f$$

Proof. The first inequality is obvious since $[d] \circ f$ is monotone as established in Claim 2.3. Let h be the monotone function closest to f, that is, $\varepsilon_f = \Delta(f, h)$. So,

$$\Delta(f, [d] \circ f) \underbrace{\leq}_{\text{triangle ineq}} \Delta(f, h) + \Delta([d] \circ f, h) \underbrace{=}_{\text{since } h = [d] \circ h} \Delta(f, h) + \underbrace{\Delta([d] \circ f, [d] \circ h)}_{\leq \Delta(f,h) \text{ by Claim 2.4}} \leq 2\Delta(f, h) = 2\varepsilon_f$$

□

We provide one more simple claim about the sort operator that will be used throughout Section 6. Given $h, h' : [n] \to \{0,1\}$, define

$$\Delta^-(h, h') = |\{c \in [n] : h(c) > h'(c)\}| \text{ and } \Delta^+(h, h') = |\{c \in [n] : h(c) < h'(c)\}|.$$

Claim 2.6. *Let $h, h' : [n] \to \{0,1\}$ be any two functions. Then, $\Delta^-(\mathtt{sort}(h), \mathtt{sort}(h')) \leq \Delta^-(h, h')$.*

Proof. Observe that if $\|h\|_1 \leq \|h'\|_1$, then $\Delta^-(\mathtt{sort}(h), \mathtt{sort}(h')) = 0$ and so we are done. On the other hand if $\|h\|_1 \geq \|h'\|_1$, then we have

$$\Delta^-(\mathtt{sort}(h), \mathtt{sort}(h')) = \|h\|_1 - \|h'\|_1 = \sum_{c \in [n]} h(c) - h'(c) = \Delta^-(h, h') - \Delta^+(h, h') \leq \Delta^-(h, h').$$

□

2.1 Colorful Influences and the Talagrand Objective

We will need undirected, colorful Talagrand inequalities for proving Theorem 1.4. For the sake of completeness, we explicitly define the undirected colored influence.

Definition 2.7. *Consider a function $g : \{0,1\}^d \to \{0,1\}$ and a 0-1 coloring ξ of the edges of the hypercube $\{0,1\}^d$. The influence of $\mathbf{z} \in \{0,1\}^d$, denoted $I_{g,\xi}(\mathbf{z})$, is the number of sensitive edges incident to $\mathbf{z}$ whose color has value $f(\mathbf{z})$.*

(An edge is sensitive if both endpoints have different values.)

Talagrand's theorem asserts that $\mathbf{E}_{\mathbf{z}}[\sqrt{I_g(\mathbf{z})}] = \Omega(\mathrm{var}(g))$ [Tal93]. The robust/colored version proven by KMS asserts this to be true for arbitrary colored influences.

Theorem 2.8 (Paraphrasing Theorem 1.8 of [KMS15])**.** *(Colored Talagrand Theorem on the Undirected Hypercube) There exists an absolute constant $C > 0$ such that for any function $g : \{0,1\}^d \to \{0,1\}$ and any 0-1 coloring ξ of the edges of the hypercube,*

$$\mathbf{E}_{\mathbf{z}\in\{0,1\}^d}\left[\sqrt{I_\xi(\mathbf{z})}\right] \geq C \cdot \mathrm{var}(g)$$

It will be convenient in our analysis to formally define the Talagrand objective for colored, thresholded influences on the hypergrid.

Definition 2.9 (Colored Thresholded Talagrand Objective)**.** *Given any Boolean function $f : [n]^d \to \{0,1\}$ and $\chi : E \to \{0,1\}$, we define the Talagrand objective with respect to the colorful thresholded influence as follows.*

$$T_{\Phi_\chi}(f) := \mathbf{E}_{\mathbf{x}}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right]$$

where, $\Phi_{f,\chi}$ is defined in Definition 1.3.

2.2 Majorization

It is convenient to think of the Talagrand objective as a "norm" of a vector. Throughout the paper, we (ab)use the following notation:

$$\text{given a vector } \mathbf{v} \in \mathbb{R}^t_{\geq 0}, \quad \|\mathbf{v}\|_{1/2} := \sum_{i=1}^{t} \sqrt{\mathbf{v}_i}.$$

If we imagine an n^d dimensional vector indexed by the points of the hypergrid, we see that the Talagrand objective is precisely the norm of the vector whose $\mathbf{x}$th entry is $\Phi_{f,\chi}(\mathbf{x})$. Most often, however, we would be considering the Talagrand objective line-by-line, with the natural ordering of the line defining a natural ordering on the vector. To be more precise, fix a dimension $i \in [d]$, fix an i-line ℓ. An i-line is a set of n points which only differ in the ith coordinate. This line ℓ defines a vector $\overrightarrow{\Phi_\ell(f)} \in \mathbb{R}^n_{\geq 0}$ whose jth coordinate, for $1 \leq j \leq n$ is precisely $\Phi_{f,\chi}(\mathbf{x})$ where $\mathbf{x} \in \ell$ has $\mathbf{x}_i = j$. Note that

$$\forall i \in [d], \quad T_{\Phi_\chi}(f) = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{\Phi_\ell(f)} \right\|_{1/2}.$$

Our proof to establish (the correct version of) (H1) proceeds via a hybrid argument that modifies the function and the coloring in various stages. In each stage, we prove that the norm decreases. We use the following facts from the theory of majorization.

In the rest of this subsection all vectors, unless explicitly mentioned, live in $\mathbb{R}^t_{\geq 0}$ for some positive integer t. Given a vector $\mathbf{a}$, we use $(\mathbf{a})^{\downarrow}$ and $(\mathbf{a})^{\uparrow}$ to denote the vectors obtained by sorting $\mathbf{a}$ in decreasing and increasing order, respectively. Given two vectors $\mathbf{a}$ and $\mathbf{b}$ with the same ℓ_1 norm, we say $\mathbf{a} \succeq_{\mathsf{maj}} \mathbf{b}$ if for all $1 \leq k \leq r$, $\sum_{i \leq k} (\mathbf{a})^{\downarrow}_i \geq \sum_{i \leq k} (\mathbf{b})^{\downarrow}_i$.

Throughout this paper, when we apply majorization the LHS vector would be sorted (either increasing or decreasing) while the RHS vector would be unsorted. To be absolutely clear which is which, when $\mathbf{a}$ is sorted decreasing, we use the notation $\mathbf{a} \succeq_{\mathsf{maj}} (\mathbf{b})^{\downarrow}$ and when $\mathbf{a}$ is sorted increasing we use the notation $\mathbf{a} \succeq_{\mathsf{maj}} (\mathbf{b})^{\uparrow}$. Here is a simple standard fact that connects majorization to the Talagrand objective; it uses the fact that the sum of square roots is a symmetric concave function, and is thus Schur-concave.

Fact 2.10 (Chapter 3, [MOA11])**.** *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be two vectors such that* $\mathbf{a} \succeq_{\mathsf{maj}} \mathbf{b}$*. Then,* $\|\mathbf{a}\|_{1/2} \leq \|\mathbf{b}\|_{1/2}$.

Next, we state and prove a simple but key lemma repeatedly used throughout the analysis.

Lemma 2.11. *Let* $\vec{U} = \sum_i \mathbf{w}_i$ *be a finite sum of* t*-dimensional non-negative vectors. Let* $\vec{S} := \sum_i (\mathbf{w}_i)^{\downarrow}$*. Then,* $\vec{S} \succeq_{\mathsf{maj}} \left(\vec{U}\right)^{\downarrow}$*. Analogously, if* $\vec{S} := \sum_i (\mathbf{w}_i)^{\uparrow}$*, then* $\vec{S} \succeq_{\mathsf{maj}} \left(\vec{U}\right)^{\uparrow}$.

Proof. We prove the first statement; the second analogous statement has an absolutely analogous proof. We begin by noting $\vec{S}$ is a sorted decreasing vector since it is a sum of sorted decreasing vectors. For brevity, let's use $\vec{V} := \left(\vec{U}\right)^{\downarrow}$. Next, we note that $\left\|\vec{S}\right\|_1 = \left\|\vec{V}\right\|_1 = \sum_i \|\mathbf{w}_i\|_1$.

Now fix a $1 \leq \tau \leq t$. We need to show $\sum_{j=1}^{\tau} \vec{S}_j \geq \sum_{j=1}^{\tau} \vec{V}_j$. Consider the τ-largest coordinates of $\vec{V}$, and let them comprise $T \subseteq [t]$. Consider the $|T|$-dimensional vectors $\mathbf{w}_i[T]$ where we restrict our attention to only these coordinates. Let $\vec{S}'$ be the $|T|$-dimensional vector formed by sum of the sorted versions of $\mathbf{w}_i[T]$. Note that $\sum_{j=1}^{\tau} \vec{S}'_j = \sum_{j=1}^{\tau} \vec{V}_j$. Also note that for any $1 \leq j \leq \tau$, the number $\vec{S}'_j$ equals the $\sum_i (j$th max of $\mathbf{w}_i[T])$ and $\vec{S}_j$ equals $\sum_{i=1}^{T} (j$th max of $\mathbf{w}_i)$. Thus, $\vec{S}_j \geq \vec{S}'_j$, proving that $\sum_{j=1}^{\tau} \vec{S}_j \geq \sum_{j=1}^{\tau} \vec{V}_j$. □

3 Semisorting and Reduction to Semisorted Functions

As we mentioned earlier when we stated (H1), we do not know if this is a true statement for an arbitrary function. It is true for what we call semisorted functions, and proving this would be the bulk of the work. In this section, we define what semisorted functions are, we prove that the Talagrand objective can only decrease when one moves to a semisorted function, and therefore how one can reduce to proving Theorem 1.4 only for semisorted functions.

Fix a function $f : [n]^d \to \{0, 1\}$. Fix a coordinate i and fix an interval $I = [a, b]$. Semisorting f on this interval in dimension i leads to a function $h : [n]^d \to \{0, 1\}$ as follows. We take every i-line ℓ and consider the function restricted on the interval I on this line, and we sort it. The following lemma shows that semisorting on any (i, I) pair can only reduce the Talagrand objective. We defer its proof to Section 3.1.

Lemma 3.1 (Semisorting only decreases T_Φ.)**.** *Let* f *be any hypergrid function and let* χ *be any bi-coloring of the augmented hypergrid edges. Let* $i \in [d]$ *be any dimension and* I *be any interval* $[a, b]$.

There exists a (re)-coloring χ' of the edges of the augmented hypergrid such that

$$T_{\Phi_\chi}(f) \geq T_{\Phi_{\chi'}}(h)$$

where h is the function obtained upon semisorting f in dimension i on the interval I.

A function $f : [n]^d \to \{0,1\}$ is called *semisorted* if for any $i \in [d]$ and any i-line ℓ, the function restricted to the first $n/2$ points is sorted increasing and the function restricted to the second half is also sorted increasing. It is instructive to note that when $n = 2$, that is when the domain is the hypercube, every function is semisorted. This shows that semisorted functions form a non-trivial family. However, the semisortedness is a property that allows us to prove that (H1) holds. In particular, we prove this theorem.

Theorem 3.2 (Connecting Talagrand Objectives of f and Tracker Functions)**.** *Let $f\colon [n]^d \to \{0,1\}$ be a **semisorted** function and let $\chi\colon E \to \{0,1\}$ be an arbitrary coloring of the edges of the fully augmented hypergrid. Then for every $\mathbf{x} \in [n]^d$, one can find a coloring $\xi_{\mathbf{x}}$ of the edges of the Boolean hypercube such that*

$$T_{\Phi_\chi}(f) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right] \;\geq\; \mathbf{E}_{\mathbf{x}\in[n]^d}\,\mathbf{E}_{S\subseteq[d]}\,\left[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}\right].$$

We can use the above theorem to get set the intuition behind (H2) correct, and prove Theorem 1.4 for semisorted functions. We state this below, but we defer the proof of this to Section 4. At this point we remind the reader again that this is not at all trivial, but the proof ideas are generalizations of those present in [KMS15, PRW22] for the hypercube case.

Theorem 3.3 (Theorem 1.4 for semisorted functions.)**.** *Let $f : [n]^d \to \{0,1\}$ be a **semisorted** function that is ε-far from monotone. Let $\chi : E \to \{0,1\}$ be an arbitrary coloring of the edges of the augmented hypergrid. Then there is a constant C'' such that*

$$T_{\Phi_\chi}(f) := \mathbf{E}_{\mathbf{x}}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right] \geq C''\varepsilon$$

Lemma 3.1 shows that the Talagrand objective can't rise on semisorting. The distance to monotonicty, however, can fall. In the remainder of the section we show how we can reduce to the semisorted case with a loss of $\log n$, and in particular, we use Theorem 3.3 to prove Theorem 1.4.

Sequence of Semisorted Functions and Reduction to the Semisorted Case. We now describe a semisorting process which gives a way of getting from f to a monotone function. Without much loss of generality, let us assume $n = 2^k$ which we can assume by padding. Iteratively coarsen the domain $[n]^d = [2^k]^d$ as follows. First "chop" this hypergrid into 2^d many $[n/2]^d = [2^{k-1}]^d$ hypergrids by slicing through the "middle" in each of the d-coordinates. More precisely, these 2^d hypergrids can be indexed via $\mathbf{v} \in \{0,1\}^d$, where given such a vector, the corresponding hypergrid is

$$H_{\mathbf{v}} = \prod_{i=1}^{d} \{\mathbf{v}_i \cdot \frac{n}{2} + 1, \mathbf{v}_i \cdot \frac{n}{2} + 2, \cdots, \mathbf{v}_i \cdot \frac{n}{2} + \frac{n}{2}\}$$

Each hypergrid $H_{\mathbf{v}}$ is an $[n/2]^d = [2^{k-1}]^d$ hypergrid. Let us denote the collection of all these hypergrids as the set $\mathcal{H}_1$. So, $\mathcal{H}_1$ has 2^d many hypergrids and each hypergrid has dimension $[n/2]^d = [2^{k-1}]^d$. Repeat the above operation on each hypergrid in $\mathcal{H}_1$. More precisely, each hypergrid $H_{\mathbf{v}}$ in $\mathcal{H}_1$ will lead to 2^d hypergrids each with dimension $[n/4]^d = [2^{k-2}]^d$. The total number of such hypergrids, which we collect in the collection $\mathcal{H}_2$, is $2^d \times 2^d = (2^2)^d$. More generally, we have a family $\mathcal{H}_i$ consisting of $\left(2^i\right)^d$ many hypergrids of dimension $[n/2^i]^d = [2^{k-i}]^d$. The collection $\mathcal{H}_{k-1}$ consists of $(2^{k-1})^d$ many d-dimensional hypercubes.

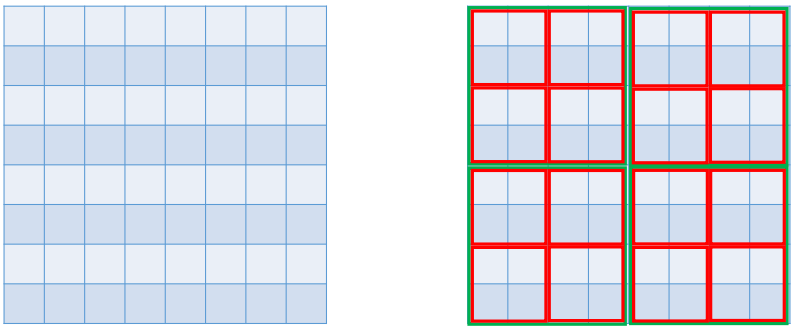


Figure 2: *In the figure, we see an example with* $d = 2$ *and* $n = 8 = 2^3$. *There are* 2^2 *many* 4×4 *green (hyper)-grids, and* 4^2 *many* 2×2 *red squares.*

Note that in any family $\mathcal{H}_i$ for $1 \leq i \leq k-1$, each $H \in \mathcal{H}_i$ is a sub-hypergrid of $[n]^d$. We let f_H denote the restriction of f only to this subset H of the domain. Also, let $\mathcal{H}_0$ denote the singleton set containing only one hypergrid, $[n]^d$. Define the function $f_1 : [n]^d \to \{0, 1\}$ as follows: consider every hypergrid[2] H in $\mathcal{H}_{k-1}$ and apply the sort operator on f_H for all these hypergrids. Note that f_1 is a monotone function when restricted to $H \in \mathcal{H}_{k-1}$. Recursively define f_i as follows: consider every hypergrid $H \in \mathcal{H}_{k-i}$ and apply the sort operator on $(f_{i-1})_H$ for all these hypergrids. Figure 3 is an illustration for $d = 2$ and $k = 3$, i.e. $n = 8$.

Claim 3.4. *There must exist an* $0 \leq j \leq k-1$ *such that* $\Delta(f_j, f_{j+1}) \geq \varepsilon_f/k$.

Proof. This follows from triangle inequality and the fact that $\Delta(f_0, f_k) \geq \varepsilon_f$. □

Proof of Theorem 1.4. We now show how Theorem 1.4 follows from Lemma 3.1 and Theorem 3.3 via an averaging argument. We fix the j as in Claim 3.4. By Lemma 3.1 we get that for any function f and any coloring χ, there exists a recoloring χ' such that $T_{\Phi_\chi}(f) \geq T_{\Phi_{\chi'}}(f_j)$. Now consider the hypergrids in $H \in \mathcal{H}_{k-j-1}$. Let $f_j|_H$ be the function restricted to this sub-domain H. Note that the function $f_j|_H$ is indeed semisorted by construction. Therefore, by Theorem 3.3 (on the coloring χ') we know that for all $H \in \mathcal{H}_{k-j-1}$,

$$T_{\Phi_{\chi'}}(f_j|_H) \geq C'' \cdot \varepsilon_{f_j|_H}$$

By Claim 2.5, we know that $2\varepsilon_{f_j|_H} \geq \Delta(f_j|_H, f_{j+1}|_H)$. Taking expectation over $H \in \mathcal{H}_{k-j-1}$, we see that the LHS is at most (at most since we only consider violations staying in H) $T_{\Phi_{\chi'}}(f_j)$, while the RHS is precisely $\Delta(f_j, f_{j+1})/2 \geq \varepsilon_f/2k$. Putting everything together, we get $T_{\Phi_\chi}(f) \geq \frac{C''\varepsilon_f}{2 \log n}$ proving Theorem 1.4. □

[2] these will be hypercubes

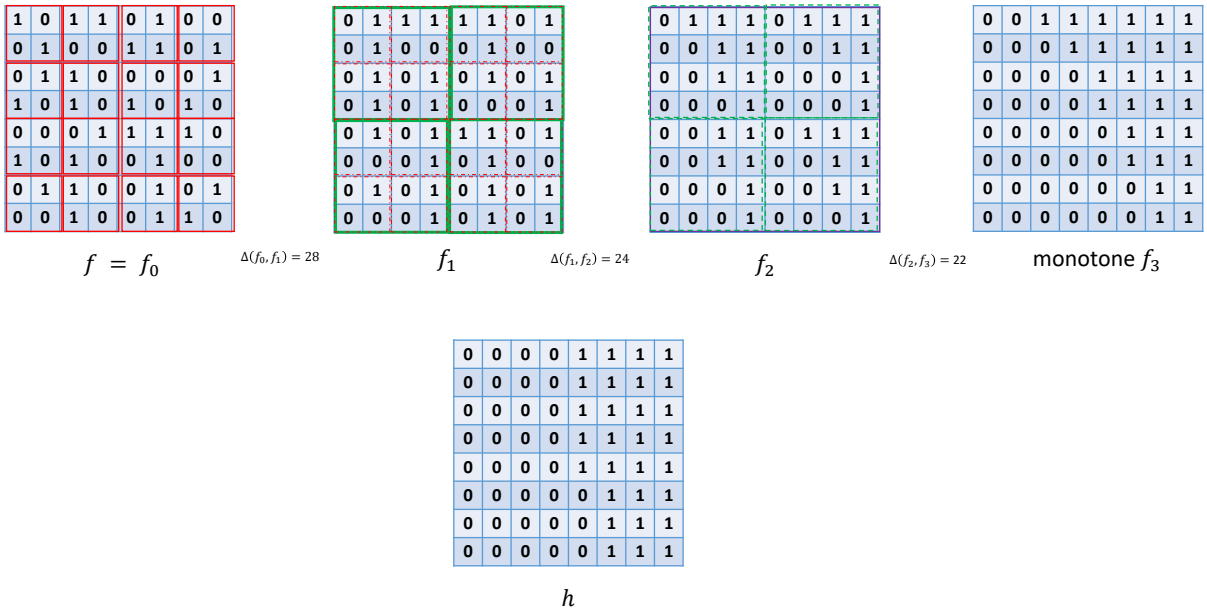


Figure 3: *The function $f = f_0$ is described to the left, and then one obtains f_1, f_2 and f_3. The function h which is obtained doing sort on the whole of f is described below. Note $h \neq f_3$.*

3.1 Semisorting only decreases the Talagrand objective: Proof of Lemma 3.1

Let us first describe the coloring χ'.

- First let us describe the recoloring of pairs of points $(\mathbf{x}, \mathbf{x}')$ which differ only in some coordinate $j \neq i$ and $\mathbf{x}_i = \mathbf{x}'_i$ lies in the interval $[a, b]$. We go over all these edges by considering pairs of i-lines which differ on a single coordinate $j \neq i$. More precisely, if $\ell = \mathbf{x} \pm t\mathbf{e}_i$ then $\ell' = \mathbf{x}' \pm t\mathbf{e}_i$ for some $\mathbf{x}' = \mathbf{x} + a\mathbf{e}_j$ with $a > 0$. We now consider re-coloring the pairs $(\mathbf{x}, \mathbf{x}' = \mathbf{x} + a\mathbf{e}_j)$ as follows.

  Let V denote the points $\mathbf{x} \in \ell$ such that (a) $\mathbf{x}_i \in I$, (b) $f(\mathbf{x}) = 1$, but (c) $f(\mathbf{x} + a\mathbf{e}_j) = 0$. That is $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j)$ is a violation. Consider all edges $E_V := \{(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \ : \ \mathbf{x} \in V\}$ and let $\vec{\chi}$ be the $|E_V|$ dimensional $0, 1$-vector which are the χ values of edges in E_V going left to right.

  Now consider the function h where I has been sorted on both ℓ and ℓ'. Let U denote the points $\mathbf{x} \in \ell$ such that (a) $\mathbf{x}_i \in I$, (b) $h(\mathbf{x}) = 1$, but (c) $h(\mathbf{x} + a\mathbf{e}_j) = 0$. That is $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j)$ is a violation in h. Firstly note that $|U| \leq |V|$ and furthermore, these $|U|$ points form a contiguous interval of I. We now describe the recoloring χ' of the edges in $E_U := \{(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \ : \ \mathbf{x} \in V\}$; all the other recolorings are immaterial since they don't contribute to $T_{\Phi_{\chi'}}(h)$ since the edges are not violating. We take the $|V|$-dimensional vector $\vec{\chi}$, sort in *decreasing* order, and then take the first $|U|$ coordinates and use them to define $\chi'(e)$ for $e \in E_U$, left to right. See Figure 4 for an illustration.

- Now we describe recoloring of pairs of points $(\mathbf{x}, \mathbf{y})$ which only differ in coordinate i. First, if both $\mathbf{x}_i$ and $\mathbf{y}_i$ lie in I, or if they both lie outside I, then we leave their colors unchanged. Furthermore, if $(\mathbf{x}, \mathbf{y})$ is *not* a violating pair in f, then we leave its color unchanged. Now consider a $\mathbf{y}$ to the right of I, that is, $\mathbf{y}_i > b$ and $f(\mathbf{y}) = 0$. Consider the $\mathbf{x}$'s with $\mathbf{x}_i$ in I with $f(\mathbf{x}) = 1$, each of which forms a violation with $\mathbf{y}$. Suppose there are k many of them, of which k_0 of them are colored 0 and k_1 of

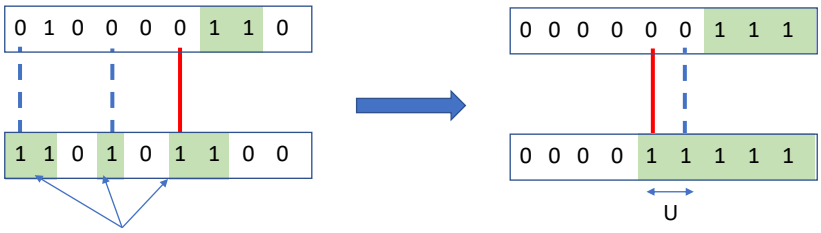


Figure 4: *We are considering only the interval I. The line below is ℓ and the line above is ℓ'. The green shaded zones correspond to where the function evaluates to* 1*s. The situation to the right is after sorting. Only the violating edges are marked. On the left, the red solid edges are colored $\chi(e) = 1$ while the blue dashed are colored $\chi(e) = 0$. On the right, the color-coding is the same but for χ'. All other unmarked edges inherit the same colors as χ.*

them are colored 1. We now consider the picture in h, and once again there are exactly k (possibly different) points in the interval which are violating with $\mathbf{y}$ in h. Going from left to right, we color the first k_1 of them 1 and the next k_0 of them 0, in χ'. We now do a similar thing for a $\mathbf{z}$ to the left of I, that is, $\mathbf{z}_i < a$ and $f(\mathbf{z}) = 1$. We now consider the $\mathbf{x}$'s with $\mathbf{x}_i \in I$ with $f(\mathbf{x}) = 0$, each of which forms a violation with $\mathbf{z}$. As before, suppose there are k many of them k_1 of them colored 1 and k_0 of them colored 0. In g also there are k locations with which $\mathbf{z}$ is a violation. We, once again, going from left to right, color the first k_1 of them 1 and the next k_0 of them 0, in χ'. See Figure 5 for an illustration.

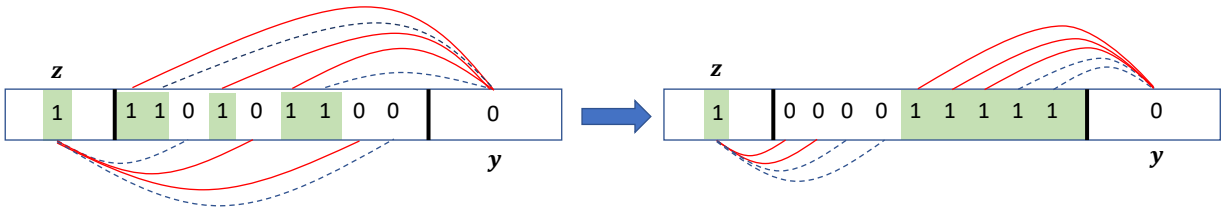


Figure 5: *The two vertical black lines demarcate I. The green shaded zones correspond to where the function evaluates to* 1*s. The situation to the right is after sorting. $\mathbf{y}$ is a point with $f(\mathbf{y}) = 0$ to the right of I; $\mathbf{z}$ is a point with $f(\mathbf{z}) = 1$ to the left of I. Only the violating edges incident to $\mathbf{y}$ and $\mathbf{z}$ are marked. On the left, the red solid edges are colored $\chi(e) = 1$ while the blue dashed are colored $\chi(e) = 0$. On the right, the color-coding is the same but for the recoloring χ'. All other unmarked edges incident of $\mathbf{y}$ or $\mathbf{z}$ inherit the same colors as χ. Edges with both endpoints in I or both endpoints outside I also inherit the same color.*

Now we prove the lemma "line-by-line". In particular, we want to prove for any i-line ℓ, we have

$$\sum_{\mathbf{x}\in\ell} \sqrt{\Phi_{f,\chi}(\mathbf{x})} \geq \sum_{\mathbf{x}\in\ell} \sqrt{\Phi_{h,\chi'}(\mathbf{x})}$$

Note that it suffices to prove the above for $\mathbf{x}$ whose $\mathbf{x}_i \in I$.

To prove the above inequality, it is best to consider the two vectors $\overrightarrow{\Phi_\chi(f)}$ and $\overrightarrow{\Phi_{\chi'}(h)}$ which are $|I|$-dimensional whose $\mathbf{x}$th coordinate is precisely $\Phi_{f,\chi}(\mathbf{x})$ and $\Phi_{h,\chi'}(\mathbf{x})$ respectively. We want to prove

$$\left\|\overrightarrow{\Phi_\chi(f)}\right\|_{1/2} \geq \left\|\overrightarrow{\Phi_{\chi'}(h)}\right\|_{1/2} \tag{1}$$

First we divide the $|I|$ coordinates of $\overrightarrow{\Phi_\chi(f)}$ into $O \cup Z$ corresponding to when $f(\mathbf{x}) = 1$ and $f(\mathbf{x}) = 0$. Let's call these two vectors $\overrightarrow{\Phi^{(1)}_\chi(f)}$ and $\overrightarrow{\Phi^{(0)}_\chi(f)}$. The former vector is $|O|$ dimensional, the latter is $|Z|$ dimensional, and $\overrightarrow{\Phi_\chi(f)}$ is obtained by some splicing of these two vectors. We will do the same for the coordinates of $\overrightarrow{\Phi_{\chi'}(h)}$ to obtain $\overrightarrow{\Phi^{(1)}_{\chi'}(h)}$ and $\overrightarrow{\Phi^{(0)}_{\chi'}(h)}$. Note that since sorting doesn't change the number of 0s or 1, both these vectors are $|O|$ and $|Z|$ dimensional, respectively. We now set to prove

$$\left\|\overrightarrow{\Phi^{(1)}_\chi(f)}\right\|_{1/2} \geq \left\|\overrightarrow{\Phi^{(1)}_{\chi'}(h)}\right\|_{1/2} \quad \text{and} \quad \left\|\overrightarrow{\Phi^{(0)}_\chi(f)}\right\|_{1/2} \geq \left\|\overrightarrow{\Phi^{(0)}_{\chi'}(h)}\right\|_{1/2} \tag{2}$$

and this will prove (1). We prove the first inequality; the proof of the second is analogous. For brevity's sake, for the rest of the section we drop the superscript (1) from $\overrightarrow{\Phi^{(1)}}$.

The plan is to write $\overrightarrow{\Phi_\chi(f)}$ as a sum of (Boolean) vectors, and then show that $\overrightarrow{\Phi_{\chi'}(h)}$ is dominated by the sum of sorts of those Boolean vectors. Then we invoke Lemma 2.11.

We write $\overrightarrow{\Phi_\chi(f)}$ as a sum of Boolean vectors as follows. Fix any other i-line $\ell' := \ell + a\mathbf{e}_j$ for some $j \neq i$ and $a > 0$. Define the following $(0,1)$-vector also indexed by elements of O.

$$\mathbf{u}_{\ell'}(\mathbf{x}) = 1 \text{ if } f(\mathbf{x} + a\mathbf{e}_j) = 0 \text{ and } \chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1$$

That is, $\mathbf{u}_{\ell'}(\mathbf{x}) = 1$ if the projection of $\mathbf{x}$ onto ℓ', $(\mathbf{x}, \mathbf{x}' := \mathbf{x} + a\mathbf{e}_j)$, is a violating edge in f with χ-color 1.

Define the following vector $\overrightarrow{A^\perp}$ as follows.

Definition 3.5. *For any $\mathbf{x} \in O$,*

$$\overrightarrow{A^\perp}(\mathbf{x}) = \sum_{j \neq i} \underbrace{\min\left(1, \sum_{\ell' = \ell + a\mathbf{e}_j} \mathbf{u}_{\ell'}(\mathbf{x})\right)}_{\text{let's call this } \mathbf{w}_j(\mathbf{x}) \in \{0,1\}} =: \sum_{j \neq i} \mathbf{w}_j(\mathbf{x})$$

Finally, for $\mathbf{x} \in O$, define

$\overrightarrow{A^\parallel}(\mathbf{x}) = 1$ if there is some $\mathbf{y}$ to its right, potentially outside the interval I with $f(\mathbf{y}) = 0$ and $\chi(\mathbf{x}, \mathbf{y}) = 1$

Using the vectors, we can write

Observation 3.6. *For any $\mathbf{x} \in O$,*

$$\overrightarrow{\Phi_\chi(f)}(\mathbf{x}) = \overrightarrow{A^\perp}(\mathbf{x}) + \overrightarrow{A^\parallel}(\mathbf{x})$$

Now let's consider the situation after I is sorted. The ones of O now "shift around"; indeed, they are the $|O|$ many right most points. Let's call these locations O' and note $|O'| = |O|$.

Now define the $|O'| = |O|$ dimensional vector $\mathbf{v}_{\ell'}$ where for $\mathbf{x} \in O'$

$$\mathbf{v}_{\ell'}(\mathbf{x}) = 1 \text{ if } h(\mathbf{x} + a\mathbf{e}_j) = 0 \text{ and } \chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1$$

Now we will use the property of the recoloring we performed. We claim two things:

Claim 3.7. *The number of* 1*s in* $\mathbf{v}_{\ell'}$ *is at most the number of* 1*s in* $\mathbf{u}_{\ell'}$*, and* $\mathbf{v}_{\ell'}$ *is sorted decreasing.*

Proof. The number of 1s in $\mathbf{u}_{\ell'}$ is precisely the number of violating edges of the form $(\mathbf{x}, \mathbf{x}')$ in f, where $\mathbf{x}_i \in I$ and $\mathbf{x}' = \mathbf{x} + a\mathbf{e}_j$ and $\chi(\mathbf{x}, \mathbf{x}') = 1$. Similarly, the number of 1s in $\mathbf{u}_{\ell'}$ are precisely the number of violating edges of the form $(\mathbf{x}, \mathbf{x}')$ in h, where $\mathbf{x}_i \in I$ and $\mathbf{x}' = \mathbf{x} + a\mathbf{e}_j$ and $\chi'(\mathbf{x}, \mathbf{x}') = 1$. When we recolored to get χ' we made sure by property (a) that the latter number is smaller.

Take $\mathbf{x}$ and $\mathbf{y}$ in O, with $\mathbf{x}_i < \mathbf{y}_i$, but suppose, for the sake of contradiction, $\mathbf{v}_{\ell'}(\mathbf{x}) = 0$ and $\mathbf{v}_{\ell'}(\mathbf{y}) = 1$. The latter implies $h(\mathbf{y}' := \mathbf{y} + a\mathbf{e}_j) = 0$ and $\chi'(\mathbf{y}, \mathbf{y}') = 1$. Since h is sorted on ℓ', $h(\mathbf{x}' := \mathbf{x} + a\mathbf{e}_j) = 0$ as well. Since $\mathbf{x} \in O$, $h(\mathbf{x}) = 1$ which means $(\mathbf{x}, \mathbf{x}')$ is a violating edge in h. $\mathbf{v}_{\ell'}(\mathbf{x}) = 0$ implies $\chi'(\mathbf{x}, \mathbf{x}') = 0$. But this violates property (b) of χ'. □

What we need is the following corollary.

$$\text{For any } \ell' = \ell + a\mathbf{e}_j, \quad \mathbf{v}_{\ell'} \ \leq_{\text{coor}} (\mathbf{u}_{\ell'})^{\downarrow} \tag{3}$$

where recall that $(z)^{\downarrow}$ is the sorted-decreasing version of z.

Just as we defined $\overrightarrow{A^{\perp}}$, define the $|O|$-dimensional vector $\overrightarrow{B^{\perp}}$ as follows.

Definition 3.8. *For any* $\mathbf{x} \in O'$*,*

$$\overrightarrow{B^{\perp}}(\mathbf{x}) = \sum_{j \neq i} \underbrace{\min\left(1, \sum_{\ell' = \ell + a\mathbf{e}_j} \mathbf{v}_{\ell'}(\mathbf{x})\right)}_{\text{let's call this } \mathbf{z}_j(\mathbf{x}) \in \{0,1\}} =: \sum_{j \neq i} \mathbf{z}_j(\mathbf{x})$$

Note that for every $j \neq i$, $\mathbf{w}_j$ and $\mathbf{z}_j$ are $|O| = |O'|$ dimensional Boolean vectors which we index by $\mathbf{x} \in O$ and $\mathbf{x} \in O'$, respectively.

Claim 3.9. *For all* j*,* $\mathbf{z}_j \leq_{\text{coor}} (\mathbf{w}_j)^{\downarrow}$.

Proof. Follows from (3), and the defintions of $\mathbf{z}_j$ and $\mathbf{w}_j$ as described in Definition 3.5 and Definition 3.8. □

Finally, for $\mathbf{x} \in O'$, define the $|O'| = |O|$ dimensional vector $\overrightarrow{B^{\parallel}}$ as

$\overrightarrow{B^{\parallel}}(\mathbf{x}) = 1$ if there is some $\mathbf{y}$ to its right, outside the interval I with $h(\mathbf{y}) = f(\mathbf{y}) = 0$ and $\chi'(\mathbf{x}, \mathbf{y}) = 1$.

Just as in Observation 3.6, note that

Observation 3.10. *For any* $\mathbf{x} \in O'$*,*

$$\overrightarrow{\Phi_{\chi'}(h)}(\mathbf{x}) = \overrightarrow{B^{\perp}}(\mathbf{x}) + \overrightarrow{B^{\parallel}}(\mathbf{x})$$

We now connect $\overrightarrow{A^{\parallel}}$ and $\overrightarrow{B^{\parallel}}$ as follows.

Claim 3.11. $\overrightarrow{B^{\parallel}} \leq_{\text{coor}} \left(\overrightarrow{A^{\parallel}}\right)^{\downarrow}$

Proof. Similar to Claim 3.7, this follows from the following claim.

Claim 3.12. *The number of* 1*s in* $\overrightarrow{B^{\parallel}}$ *is at most that in* $\overrightarrow{A^{\parallel}}$*, and* $\overrightarrow{B^{\parallel}}$ *is sorted decreasing.*

Proof. This also follows from the way we recolor χ' the pairs of the form $(\mathbf{x}, \mathbf{y})$ with $\mathbf{y}$ lying to the right of I and $f(\mathbf{y}) = 0$. First let's show $\overrightarrow{B^{\|}}$ is sorted decreasing. Take two points $\mathbf{x}$ and $\mathbf{z}$ with $a < \mathbf{x}_i < \mathbf{z}_i < b$ both evaluating to 1 in g. Say, $\overrightarrow{B^{\|}}(\mathbf{z}) = 1$ implying there is some $\mathbf{y}$ with $g(\mathbf{y}) = f(\mathbf{y}) = 0$ to the right of I s.t. $\chi'(\mathbf{z}, \mathbf{y}) = 1$. However, the way we recolor the edges incident on $\mathbf{y}$, this implies $\chi'(\mathbf{x}, \mathbf{y}) = 1$ as well. But that would imply $\overrightarrow{B^{\|}}(\mathbf{x}) = 1$.

The first part of the claim also follows from the way we recolor. Suppose the number of ones in $\overrightarrow{A^{\|}}$ is t. That is, only t of the points in O have 1-colored edges going to the right of the interval. Consider the subset W of these outer endpoints. The function value, both f and g, are 0 here. Note that none of these points in W have more than t edges incident on them which are colored 1 in χ. Now note that in χ', this number of 1-edges are conserved, and so for every $\mathbf{w} \in W$, the number of 1-colored violating edges is still $\leq t$. Now suppose for contradiction $\overrightarrow{B^{\|}}$ has $(t+1)$ ones. Take the right most point $\mathbf{x}$ and consider the violating edge $(\mathbf{x}, \mathbf{y})$ which is colored 1 in χ'. By construction, this $\mathbf{y}$ must have 1-colored edges to all the $(t+1)$ points (since we color them 1 left-to-right). This contradicts the number of 1-edges incident on $\mathbf{y}$. □

□

To summarize, we have from Observation 3.6 and Definition 3.5,

$$\overrightarrow{\Phi_\chi(f)} = \sum_{j \neq i} \mathbf{w}_j + \overrightarrow{A^{\|}}$$

that is, we have written the LHS as a sum of Boolean vectors. And, we have from Observation 3.10 and Definition 3.8, followed by Claim 3.9 and (3.11) that

$$\overrightarrow{\Phi_{\chi'}(h)} = \sum_{j \neq i} \mathbf{z}_j + \overrightarrow{B^{\|}} \quad \leq_{\text{coor}} \quad \underbrace{\sum_{j \neq i} (\mathbf{w}_j)^{\downarrow} + \left(\overrightarrow{A^{\|}}\right)^{\downarrow}}_{\text{call this } \overrightarrow{s\Phi}}$$

Trivially, we have $\left\|\overrightarrow{\Phi_{\chi'}(h)}\right\|_{1/2} \leq \left\|\overrightarrow{s\Phi}\right\|_{1/2}$, and from Lemma 2.11, we get $\left\|\overrightarrow{s\Phi}\right\|_{1/2} \leq \left\|\overrightarrow{\Phi_\chi(f)}\right\|_{1/2}$, completing the proof of the first part of (2).

4 Connecting with the Distance to Monotonicity: Proof of Theorem 3.3

In this section, we set the intuition behind (H2) straight. We show how the isoperimetric theorem Theorem 3.2 on semisorted functions can be used to prove Theorem 3.3. We begin by recalling the corollary of the undirected, colored Talagrand objective on the hypercube.

Corollary 1.9 (Corollary of Theorem 1.8 in [KMS15])**.** *Fix $f : [n]^d \to \{0,1\}$. Fix an $\mathbf{x} \in [n]^d$ and consider the tracking function $g_{\mathbf{x}} : \{0,1\}^d \to \{0,1\}$. Consider any* arbitrary *coloring $\xi_{\mathbf{x}} : E(2^{[d]}) \to \{0,1\}$ of the Boolean* hypercube*. Then, for every $\mathbf{x} \in [n]^d$, we have*

$$\mathbf{E}_{S \subseteq [d]} \left[\sqrt{I_{g_{\mathbf{x}}, \xi_{\mathbf{x}}}(S)}\right] = \Omega(\mathrm{var}(g_{\mathbf{x}}))$$

As mentioned earlier, one can't show (H2), that is, $\mathbf{E}_{\mathbf{x}}[\mathrm{var}(g_{\mathbf{x}})] = \Omega(\varepsilon_f)$. Indeed, there are examples of functions even over the hypercube where the above bound does *not* hold. KMS deal with this problem by

applying Theorem 3.2 to random restrictions of f. One can show that there is some restriction where the corresponding $\mathbf{E}_{\mathbf{x}}[\mathrm{var}(g_{\mathbf{x}})]$ is large. They referred to these calculations as the "telescoping argument". This argument was quantitatively improved by Pallavoor-Raskhodnikova-Waingarten [PRW22].

In this section, we port that argument to the hypergrid setting. Our proof is different in its presentation, though the key ideas are the same as KMS. Our first step is to convert Theorem 3.2 to a more convenient form, using the undirected Theorem 2.8.

Theorem 4.1. *There exists a constant $C' > 0$ such that for any semisorted function $f : [n]^d \to \{0,1\}$ and any arbitrary coloring $\chi : E \to \{0,1\}$ of the augmented hypergrid, we have*

$$T_{\Phi_\chi}(f) \geq C' \cdot \mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)].$$

Proof. By Theorem 3.2, there exists some colorings $\xi_{\mathbf{x}}$ such that $T_{\Phi_\chi}(f) \geq \mathbf{E}_{\mathbf{x}}\mathbf{E}_S[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}]$. By the undirected Talagrand bound Theorem 2.8, $\mathbf{E}_S[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}] \geq C \cdot \mathrm{var}(g_{\mathbf{x}})$.

$$\begin{aligned}
\mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)] &= \mathbf{E}_S\mathbf{E}_{\mathbf{x}}[\mathbb{1}((S \circ f)(\mathbf{x}) \neq (\overline{S} \circ f)(\mathbf{x}))] \\
&= \mathbf{E}_S\mathbf{E}_{\mathbf{x}}[\mathbb{1}(g_{\mathbf{x}}(S) \neq g_{\mathbf{x}}(\overline{S}))] \\
&= \mathbf{E}_{\mathbf{x}}\mathbf{E}_S[\mathbb{1}(g_{\mathbf{x}}(S) \neq g_{\mathbf{x}}(\overline{S}))] \leq 4\mathbf{E}_{\mathbf{x}}[\mathrm{var}(g_{\mathbf{x}})] \qquad (4)
\end{aligned}$$

(The final inequality uses Claim 4.2, stated below.) Hence, $\mathbf{E}_{\mathbf{x}}\mathbf{E}_S[\sqrt{I_{g_{\mathbf{x}},\xi}(S)}] \geq (C/4)\mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$.

Claim 4.2. *For any Boolean function $h : \{0,1\}^d \to \{0,1\}$, $\Pr_S[h(S) \neq h(\overline{S})] \leq 4\mathrm{var}(h)$.*

Proof. Recall that $\mathrm{var}(h) = 4 \Pr_S[h(S) = 0] \Pr_S[h(S) = 1]$. Hence, $\mathrm{var}(h) = 4 \max_{b \in \{0,1\}} \Pr_S[h(S) = b] \min_{b \in \{0,1\}} \Pr_S[h(S) = b]$. Since one of the values is taken with probability at least $1/2$, $\mathrm{var}(h) \geq 2 \min_{b \in \{0,1\}} \Pr_S[h(S) = b]$.

Let $\boldsymbol{S} = \{S \mid h(S) \neq h(\overline{S})\}$. Observe that half the sets in $\boldsymbol{S}$ have an h-value of 1, and the other half have value zero. Hence, $\Pr_S[h(S) \neq h(\overline{S})] \leq 2 \min_{b \in \{0,1\}} \Pr_S[h(S) = b]$. Combining with the bound from the previous paragraph, $\Pr_S[h(S) \neq h(\overline{S})] \leq 4\mathrm{var}(h)$. □

□

We now give some definitions and claim regarding the Talagrand objective of random restrictions of functions.

Definition 4.3. *Let $S \subseteq [d]$ be a subset of coordinates. The* distribution of restrictions on S, *denoted $\mathcal{R}_S$, is supported over functions and generated as follows. We pick a uar setting of the coordinates in $\overline{S}$, and output the function under this restriction. (Hence, $h \sim \mathcal{R}_S$ has domain $[n]^S$.)*

The isoperimetric theorem of Theorem 3.2 holds for any ordering of the coordinates. In this section, we will need to randomize the ordering of the sort operators. We will represent an ordering as a permutation π over $[d]$. Abusing notation, for any subset $S \subseteq [d]$, $\pi(S)$ is the induced ordered list of S.

Definition 4.4. *For any function $h : [n]^k \to \{0,1\}$, define $\delta(h)$ to be $\mathbf{E}_\pi[\Delta(h, \pi([k]) \circ h)]$.*

By Claim 2.3, sorting on all coordinates leads to a monotone function. Thus, $\delta(h)$ is at least the distance of h to monotonicity. We will perform our analyses in terms of $\delta(f)$, since it is more amenable to a proof by induction over domain size.

The following claim is central to the final induction, and relates $\delta(f)$ to $\mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$. This is the (only) claim where we need to permute the coordinates. All other claims and theorems hold for an arbitrary ordering of the coordinates (when defining $S \circ f$).

Claim 4.5. $\delta(f) \leq \mathbf{E}_S \mathbf{E}_{h\sim\mathcal{R}(S)}[\delta(h)] + \mathbf{E}_\pi \mathbf{E}_S[\Delta(\pi(S)\circ f, \pi(\overline{S})\circ f)]$

Proof. Let us consider an arbitrary ordering of dimensions. By triangle inequality,

$$\Delta(f, S\circ\overline{S}\circ f) \leq \Delta(f, S\circ f) + \Delta(S\circ f, S\circ\overline{S}\circ f)$$

Observe that $S\circ S\circ f = S\circ f$, since sorting repeatedly on a dimension does not modify a function. Hence, $\Delta(S\circ f, S\circ\overline{S}\circ f) = \Delta(S\circ S\circ f, S\circ\overline{S}\circ f) \leq \Delta(S\circ f, \overline{S}\circ f)$. The latter inequality holds because switching only reduces the Hamming distance between functions (Claim 2.4). Plugging this bound in and taking expectations over ordered subset S of dimensions:

$$\mathbf{E}_S[\Delta(f, S\circ\overline{S}\circ f)] \leq \mathbf{E}_S[\Delta(f, S\circ f)] + \mathbf{E}_S[\Delta(S\circ f, \overline{S}\circ f)] \tag{5}$$

Observe that $S\circ f$ only changes the function in the dimensions in S, and can be thought to act on the restrictions of f (to S). Hence $\mathbf{E}_S[\Delta(f, S\circ f)] = \mathbf{E}_{h\sim\mathcal{R}(S)}[\Delta(h, S\circ h)]$. Roughly speaking, the quantity $\Delta(f, S\circ\overline{S}\circ f)$ is $\varepsilon(f)$ and $\mathbf{E}_{h\sim\mathcal{R}(S)}[\Delta(h, S\circ h)]$ is $\mathbf{E}_{h\sim\mathcal{R}S}\varepsilon(h)$. So we would hope that (5) implies $\varepsilon(f) \leq \varepsilon(h) + \mathbf{E}_S[\Delta(S\circ f, \overline{S}\circ f)]$.

Unfortunately, the quantities are only constant factor approximations of $\varepsilon(f), \varepsilon(h)$. So by converting (5) in terms of $\varepsilon(f)$, we would potentially lose a constant factor in (5).

To avoid this problem, we deal with $\delta(f)$ instead. By randomly permuting S and taking expectations, the quantities in (5) can be replaced by $\delta(\cdot)$ terms. Taking expectations over a uar π, (5) implies

$$\mathbf{E}_\pi \mathbf{E}_S[\Delta(f, \pi(S)\circ\pi(\overline{S})\circ f)] \leq \mathbf{E}_\pi \mathbf{E}_S[\Delta(f, \pi(S)\circ f)] + \mathbf{E}_\pi \mathbf{E}_S[\Delta(\pi(S)\circ f, \pi(\overline{S})\circ f)] \tag{6}$$

Note that the switching order in the LHS, $\pi(S)\circ\pi(\overline{S})$, is uniformly random. Moreover,

$$\mathbf{E}_\pi \mathbf{E}_S \mathbf{E}_{h\sim\mathcal{R}(S)}[\Delta(h, \pi(S)\circ h)] = \mathbf{E}_S \mathbf{E}_h \mathbf{E}_\pi[\Delta(h, \pi(S)\circ h)] = \mathbf{E}_S \mathbf{E}_h[\delta(h)]$$

Combining all our bounds, we get that $\delta(f) \leq \mathbf{E}_S \mathbf{E}_{h\sim\mathcal{R}(S)}[\delta(h)] + \mathbf{E}_\pi \mathbf{E}_S[\Delta(\pi(S)\circ f, \pi(\overline{S})\circ f)]$. □

We prove a useful claim about the Talagrand objective of restrictions, made in [PRW22].

Claim 4.6. *Let $p \in (0,1)$, and $\mathcal{H}(p)$ be the distribution of subsets of $[d]$ generated by selecting each element with iid probability p. Then, $T_{\Phi_\chi}(f) \geq (1/\sqrt{p}) \cdot \mathbf{E}_{S\sim\mathcal{H}(p)} \mathbf{E}_{h\sim\mathcal{R}_S}[T_{\Phi_\chi}(h)]$.*

Proof. Fix a set S. For any subset S of coordinates, let the define the influence in S as $\Phi_{f,\chi}(\mathbf{x}; S) := \sum_{i\in S} \Phi_{f,\chi}(\mathbf{x}; i)$. We are just summing the influences over the coordinates of S.

Consider the quantity $\mathbf{E}_{h\sim\mathcal{R}_S}[T_{\Phi_\chi}(h)] = \mathbf{E}_{h\sim\mathcal{R}_S} \mathbf{E}_{\mathbf{z}}[\sqrt{\Phi_{h,\chi}(\mathbf{z})}]$. Note that $\mathbf{z}$ denotes a uar setting of the coordinates in S. The colorings of h are inherited from the coloring of f. Each function h is indexed by a (uar) setting of $\overline{S}$. Hence,

$$\mathbf{E}_{h\sim\mathcal{R}_S} \mathbf{E}_{\mathbf{z}}[\sqrt{\Phi_{h,\chi}(\mathbf{z})}] = \mathbf{E}_{\mathbf{x}}[\sqrt{\Phi_{f,\chi}(\mathbf{x}; S)}] \tag{7}$$

The point $\mathbf{x}$ is uar in the entire domain $[n]^d$. Note that $\mathbf{E}_{S\sim\mathcal{H}(p)}[\Phi_{f,\chi}(\mathbf{x}; S)]$ is precisely $p \cdot \Phi_{f,\chi}(\mathbf{x}; S)$, since each coordinate is independently picked in S with probability p.

$$\begin{aligned}
\mathbf{E}_{S\sim\mathcal{H}(p)} \mathbf{E}_{h\sim\mathcal{R}_S}[T_{\Phi_\chi}(h)] &= \mathbf{E}_S \mathbf{E}_{\mathbf{x}}[\sqrt{\Phi_{f,\chi}(\mathbf{x}; S)}] \\
&= \mathbf{E}_{\mathbf{x}} \mathbf{E}_S[\sqrt{\Phi_{f,\chi}(\mathbf{x}; S)}] \\
&\leq \mathbf{E}_{\mathbf{x}}\left[\sqrt{\mathbf{E}_S[\Phi_{f,\chi}(\mathbf{x}; S)]}\right] = \mathbf{E}_{\mathbf{x}}\left[\sqrt{p \cdot \Phi_{f,\chi}(\mathbf{x}; S)}\right] = \sqrt{p} \cdot T_{\Phi_\chi}(f)
\end{aligned}$$

The inequality above is a consequence of the concavity of the square root function and Jensen's inequality. □

Now we have all the ingredients to prove Theorem 3.3 whice we restate below for convenience.

Theorem 3.3 (Theorem 1.4 for semisorted functions.)**.** *Let $f : [n]^d \to \{0,1\}$ be a **semisorted** function that is ε-far from monotone. Let $\chi : E \to \{0,1\}$ be an arbitrary coloring of the edges of the augmented hypergrid. Then there is a constant C'' such that*

$$T_{\Phi_\chi}(f) := \mathbf{E}_{\mathbf{x}}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right] \geq C''\varepsilon$$

Proof. The proof is by induction over the dimension d of the domain. Formally, we will prove a lower bound of $(C'/10)\varepsilon$, where C' is the constat of Theorem 4.1.

Let us first prove the base case, when $d \leq 10$. Note that $\Phi_{f,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Phi_{f,\chi}(\mathbf{x}; i)$, where each term in the summation is 0-1 valued. Hence, by the l_1-l_2-inequality, $\sqrt{\Phi_{f,\chi}(\mathbf{x})} \geq \sum_{i=1}^{d} \Phi_{f,\chi}(\mathbf{x}; i)/d = \Phi_{f,\chi}(\mathbf{x})/d$.

Thus, $T_{\Phi_\chi}(f) \geq \mathbf{E}_{\mathbf{x}}[\Phi_{f,\chi}(\mathbf{x})]/d$. Furthermore, $\mathbf{E}_{\mathbf{x}}[\Phi_{f,\chi}(\mathbf{x})] = \sum_{i=1}^{d} \mathbf{E}_{\mathbf{x}}[\Phi_{f,\chi}(\mathbf{x}; i)]$. We can break the expectation over $\mathbf{x}$ into lines as follows.

$$\mathbf{E}_{\mathbf{x}}[\Phi_{f,\chi}(\mathbf{x})] = \sum_{i=1}^{d} \mathbf{E}_{\ell \text{ uar } i\text{-line}} \mathbf{E}_c[\Phi_{f|_\ell,\chi}(c)]$$

(The coordinate c is uar in $[n]$.) Now, for a Boolean function $f|_\ell$ on a line, if the distance to monotonicity is ε, then there are at least εn violating pairs [EKK+00], and thus for any coloring χ, we have $\mathbf{E}_c[\Phi_{f|_\ell,\chi}(c)] \geq \varepsilon(f|_\ell)$, and $\sum_{i=1}^{d} \mathbf{E}_{\ell \text{ uar } i\text{-line}} \varepsilon(f|_\ell) = \Omega(\varepsilon(f))$.

Hence, $T_{\Phi_\chi}(f) = \Omega(\varepsilon/d)$. For $d \leq 10$, the lemma holds, and so henceforth we assume $d \geq 10$.

Now for the induction step. We now break into cases.

Case 1, $\mathbf{E}_\pi \mathbf{E}_S[\Delta(\pi(S) \circ f, \pi(\overline{S}) \circ f)] \geq \delta(f)/10$: By Theorem 4.1, $T_{\Phi_\chi}(f) \geq c \cdot \mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$ (for any ordering of coordinates). So $T_{\Phi_\chi}(f) \geq c \cdot \mathbf{E}_\pi \mathbf{E}_S[\Delta(\pi(S) \circ f, \pi(\overline{S}) \circ f)] \geq (c/10) \cdot \delta(f)$.

Case 2, $\mathbf{E}_\pi \mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)] < \delta(f)/10$: By Claim 4.5, $\mathbf{E}_S \mathbf{E}_{h \sim \mathcal{R}(S)}[\delta(h)] \geq \delta(f) - \mathbf{E}_\pi \mathbf{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$. In this case, we can lower bound $\mathbf{E}_S \mathbf{E}_{h \sim \mathcal{R}(S)}[\delta(h)] \geq (9/10)\delta(f)$. Note that S is drawn from the distribution $\mathcal{H}(1/2)$. When $S \neq [d]$, we can apply induction to $T_{\Phi_\chi}(h)$ for $h \sim \mathcal{R}(S)$. Hence,

$$\begin{aligned}
\mathbf{E}_{S \sim \mathcal{H}(1/2)} \mathbf{E}_{h \sim \mathcal{R}(S)}[T_{\Phi_\chi}(h)] &\geq 2^{-d} \sum_{S \neq [d]} \mathbf{E}_{h \sim \mathcal{R}(S)}[T_{\Phi_\chi}(h)] \geq 2^{-d} \cdot (c/10) \cdot \sum_{S \neq [d]} \mathbf{E}_{h \sim \mathcal{R}(S)}[\delta(h)] \\
&= 2^{-d} \cdot (c/10) \cdot \Big(\sum_{S \subseteq [d]} \mathbf{E}_{h \sim \mathcal{R}(S)}[\delta(h)] - \mathbf{E}_{h \sim \mathcal{R}([d])}[\delta(h)] \Big) \\
&= (c/10) \Big(\mathbf{E}_S \mathbf{E}_{h \sim \mathcal{R}(S)}[\delta(h)] - 2^{-d} \delta(f) \Big) \quad (h \sim \mathcal{R}([d]) \text{ is } f) \\
&\geq (c/10) \cdot (9/10) \cdot \delta(f) - 2^{-d} \cdot (c/10) \cdot \delta(f) \quad \text{(by case condition)} \\
&= (9/10 - 2^{-d}) \cdot (c/10) \cdot \delta(f) \geq (4/5) \cdot (c/10) \cdot \delta(f) \qquad (8)
\end{aligned}$$

By Claim 4.6, $T_{\Phi_\chi}(f) \geq \sqrt{2} \cdot \mathbf{E}_{S \sim \mathcal{H}(1/2)} \mathbf{E}_{h \sim \mathcal{R}(S)}[T_{\Phi_\chi}(h)]$. Combining with the inequality of (8), $T_{\Phi_\chi}(f) \geq (\sqrt{2} \cdot 4/5) \cdot (c/10) \cdot \delta(f) \geq (c/10) \cdot \delta(f)$. □

5 Connecting Talagrand Objectives of f and the Tracker Functions

In this section and the next, we establish our main technical result Theorem 3.2 relating the Talagrand objectives on the colorful thresholded influence of the hypergrid function $f : [n]^d \to \{0,1\}$ and the Talagrand objectives on the undirected influence of the tracker functions. We restate the theorem below for convenience.

Theorem 3.2 (Connecting Talagrand Objectives of f and Tracker Functions)**.** *Let $f\colon [n]^d \to \{0,1\}$ be a **semisorted** function and let $\chi\colon E \to \{0,1\}$ be an arbitrary coloring of the edges of the fully augmented hypergrid. Then for every $\mathbf{x} \in [n]^d$, one can find a coloring $\xi_{\mathbf{x}}$ of the edges of the Boolean hypercube such that*

$$T_{\Phi_\chi}(f) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right] \;\geq\; \mathbf{E}_{\mathbf{x}\in[n]^d}\,\mathbf{E}_{S\subseteq[d]}\,\left[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}\right].$$

To prove Theorem 3.2 we need to describe the coloring $\xi_{\mathbf{x}}$ for each $\mathbf{x}$ in $[n]^d$. We proceed doing so in d stages.

- For every $i \in \{0,1,\ldots,d\}$ and for every $\mathbf{x} \in [n]^d$, we define a *partial* edge coloring $\xi^{(i)}_{\mathbf{x}}$ of the hypercube which assigns a $\{0,1\}$ value to every hypercube edge of the form $(T, T \oplus j)$ for all $j \leq i$, and for all $T \subseteq [i]$. The process will begin with the null coloring, $\xi^{(0)}_{\mathbf{x}}$, and end with a complete coloring, $\xi_{\mathbf{x}} := \xi^{(d)}_{\mathbf{x}}$, for every $\mathbf{x} \in [n]^d$.
- For every $i \in \{0,1,\ldots,d\}$ and every $S \subseteq [i]$ we will also define a coloring $\chi^{(i)}_S$ of the edges of the augmented hypergrid. We start with $\chi^{(0)}_\emptyset := \chi$ where χ is the original coloring which, recall, is adversarially chosen.

For every $i \in \{0,1,\ldots,d\}$ and $S \subseteq [i]$ we will use the above colorings to define the (i,S)-*hybrid Talagrand objective*

$$R_i(S) := \mathbf{E}_{\mathbf{x}\in[n]^d}\sqrt{\sum_{j=1}^{i} I^{=j}_{g_{\mathbf{x}},\xi^{(i)}_{\mathbf{x}}}(S) \;+\; \sum_{j=i+1}^{d} \Phi_{S\circ f,\chi^{(i)}_S}(\mathbf{x};j)}. \qquad \text{(Colorful Hybrid)}$$

Recall that $S \circ f$ is the function obtained after sorting f on the coordinates in S. Note that $R_i(S)$ is well-defined given the partial colorings $\xi^{(i)}_{\mathbf{x}}$ for each $x \in [n]^d$ as defined above. Also observe that since $\chi^{(0)}_\emptyset := \chi$, the arbitrary coloring specified in the theorem statement, we have that $R_0(\emptyset)$ is precisely the LHS in the statement of Theorem 3.2, that is, $R_0(\emptyset) = \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right]$. Additionally, since we use $\xi_{\mathbf{x}} := \xi^{(d)}_{\mathbf{x}}$, observe that $\mathbf{E}_{S\subseteq[d]}[R_d(S)]$ is precisely the statement of Theorem 3.2.

With the above setup in mind, we show that the following Lemma 5.1 suffices to prove Theorem 3.2.

Lemma 5.1 (Potential Drop Lemma)**.** *Fix $i \in \{1,\ldots,d\}$, $\xi^{(i-1)}_{\mathbf{x}}$ for all $\mathbf{x} \in [n]^d$, and $\chi^{(i-1)}_S$ for every $S \subseteq [i-1]$, which all satisfy the specifications described in the previous paragraph. There exists a choice of $\xi^{(i)}_{\mathbf{x}}$ for every $\mathbf{x} \in [n]^d$ and $\chi^{(i)}_S$, $\chi^{(i)}_{S+i}$ for every $S \subseteq [i-1]$ all satisfying the specifications*

described in the previous paragraph, such that for all $S \subseteq [i-1]$, we have (a) $R_{i-1}(S) \geq R_i(S)$ and (b) $R_{i-1}(S) \geq R_i(S+i)$.

Proof of Theorem 3.2: Consider the following binary tree with $d+1$ levels. Each level $i \in \{0, 1, \ldots, d\}$ has 2^i nodes indexed by subsets $S \subseteq [i]$. Every such node is associated with a coloring $\chi_S^{(i)}$ of the augmented hypergrid edges. The level i is also associated with a partial coloring $\xi_{\mathbf{x}}^{(i)}$ for every $\mathbf{x} \in [n]^d$.

The 0'th level contains a single node indexed by $\emptyset$. The associated augmented hypergrid coloring is $\chi_\emptyset^{(0)} := \chi$. The partial coloring $\xi_{\mathbf{x}}^{(0)}$ is null for all $\mathbf{x} \in [n]^d$. We associate the value $R_0(\emptyset) = T_{\Phi_\chi}(f)$ with the root.

For $1 \leq i \leq d$, we describe the children of each node in level $i-1$. Each node in level $i-1$ is indexed by some $S \subseteq [i-1]$. We associate this node with the value $R_{i-1}(S)$. This node has two children at level i: one, the left child, indexed by S and the other, the right child, indexed by $S+i$. The coloring of the hypergrid edges at the left child is defined as $\chi_S^{(i)}$ from the lemma, and that of the hypergrid edges at the right child is defined as $\chi_{S+i}^{(i)}$ from the lemma. The left and right children hold the quantites $R_i(S)$ and $R_i(S+i)$, respectively. At level i, the partial coloring $\xi_{\mathbf{x}}^{(i-1)}$ is also extended to $\xi_{\mathbf{x}}^{(i)}$ for every $\mathbf{x} \in [n]^d$ as stated in the lemma. From the lemma, we have $R_{i-1}(S) \geq R_i(S)$ and $R_{i-1}(S) \geq R_i(S+i)$. This immediately implies the following:

$$\text{For all } i \in \{1, \ldots, d\}, \text{ we have } \mathbf{E}_{S \subseteq [i-1]}[R_{i-1}(S)] \geq \mathbf{E}_{S \subseteq [i]}[R_i(S)]$$

and chaining these d inequalities together yields $R_0(\emptyset) \geq \mathbf{E}_{S \subseteq [d]}[R_d(S)]$.

Now consider the leaf nodes of this tree, which hold the values $R_d(S)$ for every $S \subseteq [d]$. Observe that $R_d(S) = \mathbf{E}_{\mathbf{x} \in [n]^d}\left[\sqrt{I_{g_{\mathbf{x}}, \xi_{\mathbf{x}}}(S)}\right]$ since $\xi_{\mathbf{x}} := \xi_{\mathbf{x}}^{(d)}$. Recalling that $R_0(\emptyset) = T_{\Phi_\chi}(f)$ yields

$$T_{\Phi_\chi}(f) = R_0(\emptyset) \geq \mathbf{E}_{S \subseteq [d]}[R_d(S)] = \mathbf{E}_{S \subseteq [d]} \mathbf{E}_{\mathbf{x} \in [n]^d}\left[\sqrt{I_{g_{\mathbf{x}}, \xi_{\mathbf{x}}}(S)}\right]$$

and this establishes the claim after exchanging the expectations. □

6 Proof of Potential Drop Lemma 5.1

Recall $i \in \{1, \ldots, d\}$ is fixed. For brevity's sake, we will fix a set $S \subseteq [i-1]$ and call $h := (S \circ f)$. Let's refer to $\chi_S^{(i-1)}$ as simply χ without confusing with the original χ in the theorem. The two colorings $\chi_S^{(i)}$ and $\chi_{S+i}^{(i)}$ that we construct will be simply called χ' and χ'', respectively. Let's call the partial colorings $\xi_{\mathbf{x}}^{(i-1)}$ as simply $\xi_{\mathbf{x}}$. We will call the coloring $\xi_{\mathbf{x}}^{(i)}$ which we need to construct simply $\xi'_{\mathbf{x}}$ in the latter. Recall that $\xi_{\mathbf{x}}$ is defined on all edges $(T, T \oplus j)$ for $T \subseteq [i-1]$ and $j \leq i-1$ and in order to prove the lemma we will need to define $\xi'_{\mathbf{x}}$ on all edges $(T \oplus j)$ for $T \subseteq [i]$ and $j \leq i$.

Fix an i-line ℓ. We prove the lemma line-by-line. To be precise, let us consider the following vectors. First,

$$\vec{L}_\ell := \left(\underbrace{\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}}, \xi_{\mathbf{x}}}(S)}_{\vec{L^{(1)}}_\ell} + \underbrace{\Phi_{h,\chi}(\mathbf{x}; i)}_{\vec{L^{(2)}}_\ell} + \underbrace{\sum_{j=i+1}^{d} \Phi_{h,\chi}(\mathbf{x}; j)}_{\vec{L^{(3)}}_\ell} \;:\; \mathbf{x} \in \ell \right) \tag{9}$$

Observe that

$$R_{i-1}(S) = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \vec{L}_\ell \right\|_{1/2} = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{L^{(1)}}_\ell + \overrightarrow{L^{(2)}}_\ell + \overrightarrow{L^{(3)}}_\ell \right\|_{1/2} \tag{10}$$

where, recall, we are (ab)using the notation $\|v\|_{1/2} := \sum_i \sqrt{v_i}$.
Define

$$\vec{R}_\ell := \left(\underbrace{\sum_{j=1}^{i-1} I_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}{}^{=j}(S)}_{\overrightarrow{R^{(1)}}_\ell} + \underbrace{I^{=i}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S)}_{\overrightarrow{R^{(2)}}_\ell} + \underbrace{\sum_{j=i+1}^{d} \Phi_{h,\chi'}(\mathbf{x}; j)}_{\overrightarrow{R^{(3)}}_\ell} \quad : \mathbf{x} \in \ell \right) \tag{11}$$

where we have denoted, in red, the recolorings that we need to define. The "first" RHS term is

$$R_i(S) := \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \vec{R}_\ell \right\|_{1/2} = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{R^{(1)}}_\ell + \overrightarrow{R^{(2)}}_\ell + \overrightarrow{R^{(3)}}_\ell \right\|_{1/2} \tag{12}$$

Similarly, define

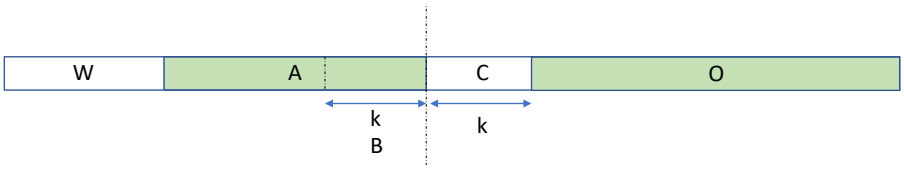
$$\vec{M}_\ell := \left(\underbrace{\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S+i)}_{\overrightarrow{M^{(1)}}_\ell} + \underbrace{I^{=i}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S+i)}_{\overrightarrow{M^{(2)}}_\ell} + \underbrace{\sum_{j=i+1}^{d} \Phi_{ioh,\chi''}(\mathbf{x}; j)}_{\overrightarrow{M^{(3)}}_\ell} \quad : \mathbf{x} \in \ell \right) \tag{13}$$

and notice that the "second" RHS term is

$$R_i(S+i) := \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \vec{M}_\ell \right\|_{1/2} = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{M^{(1)}}_\ell + \overrightarrow{M^{(2)}}_\ell + \overrightarrow{M^{(3)}}_\ell \right\|_{1/2} \tag{14}$$

Observe now that it suffices to prove that there exists colorings χ', χ'', and $\xi'_{\mathbf{x}}$'s such that $\left\| \vec{L}_\ell \right\|_{1/2} \geq \left\| \vec{R}_\ell \right\|_{1/2}$ and $\left\| \vec{L}_\ell \right\|_{1/2} \geq \left\| \vec{M}_\ell \right\|_{1/2}$ for all i-lines ℓ. Thus, we now fix an i-line ℓ and drop the subscript, ℓ, from all the previously defined vectors for brevity. We define $\mathsf{LHS} := \left\| \vec{L} \right\|_{1/2}$, $\mathsf{RHS}_1 := \left\| \vec{R} \right\|_{1/2}$, $\mathsf{RHS}_2 := \left\| \vec{M} \right\|_{1/2}$, and set out to prove $\mathsf{LHS} \geq \mathsf{RHS}_1$ and $\mathsf{LHS} \geq \mathsf{RHS}_2$.

A Picture of the Line. Since h is semisorted, the picture of h restricted to ℓ looks like this. The green zone is where the function is 1. Without loss of generality we assume ℓ has more ones than zeros. We use A to denote the ones on the left and C to denote the zeros on the right. We use $k := |C|$, and $B \subseteq A$ are the k right most ones in the left side. Throughout, we will use the notation $\vec{A}_X$ to denote the sub-vector of $\vec{A}$ defined on ℓ with coordinates restricted to $x \in X$; we will always use this notation when X is a contiguous interval. Indeed, these X's will be always picked from $\{W, A, C, O, B, A \setminus B\}$ or unions of these, always making sure they form a contiguous interval.



High Level Idea. Before we venture into proving the inequalities, we would like to remind the reader again of the proof strategy discussed in Section 1.3. We need to define the colorings χ', χ'', and also $\xi_{\mathbf{x}}^{(i)}$'s such that the objective after recoloring satisfy the inequality we desire to prove. This going to hinge upon showing that the vector obtained after operation either majorizes or is coordinate-wise dominated by a vector that majorizes the vector before the operation. In particular, these are the conditions (a)-(d) and (e)-(h) mentioned below in the grey boxes. To show these properties, we would be crucially using the property that the function f is semi-sorted which leads to certain monotonicity properties that allows us to claim them. In particular, we would be using Lemma 2.11 when establishing almost all the conditions mentioned above. There is a certain sense of repetition in which these arguments are made, however, we have provided all the details for completeness.

6.1 Proving $\mathsf{LHS} \geq \mathsf{RHS}_1$

During the proof of $\mathsf{LHS} \geq \mathsf{RHS}_1$, we will define the coloring χ' on all edges of the fully augmented hypergrid and $\xi'_{\mathbf{x}}(S, S \oplus j)$ where $j \leq i$ for all $\mathbf{x} \in [n]^d$. We will not specify $\xi'_{\mathbf{x}}(S+i, S+i \oplus j)$ since these won't be needed to prove this inequality; we will describe them when we prove $\mathsf{LHS} \geq \mathsf{RHS}_2$.

Before we describe the recolorings, it is useful to describe the plan of the proof. This will motivate why we recolor as we do. We will actually consider

$$\mathsf{LHS} = \left\|\vec{L}_W\right\|_{1/2} + \left\|\vec{L}_A\right\|_{1/2} + \left\|\vec{L}_C\right\|_{1/2} + \left\|\vec{L}_O\right\|_{1/2}$$

and

$$\mathsf{RHS}_1 = \left\|\vec{R}_W\right\|_{1/2} + \left\|\vec{R}_A\right\|_{1/2} + \left\|\vec{R}_C\right\|_{1/2} + \left\|\vec{R}_O\right\|_{1/2}$$

and argue domination term-by-term.

More precisely, we find recolorings χ', ξ' such that

(a) $\overrightarrow{R_A^{(q)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_A^{(q)}}\right)^{\downarrow}$ and $\overrightarrow{R_O^{(q)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_O^{(q)}}\right)^{\downarrow}$, for $q \in \{1,3\}$,

(b) $\exists \overrightarrow{L_A'^{(2)}}$ such that $\overrightarrow{L_A'^{(2)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_A^{(2)}}\right)^{\downarrow}$ and $\overrightarrow{L_A'^{(2)}} \succeq_{\mathsf{coor}} \overrightarrow{R_A^{(2)}}$,

(c) $\overrightarrow{R_W^{(q)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_W^{(q)}}\right)^{\uparrow}$ and $\overrightarrow{R_C^{(q)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_C^{(q)}}\right)^{\uparrow}$, for $q \in \{1,3\}$,

(d) $\exists \overrightarrow{L_C'^{(2)}}$ such that $\overrightarrow{L_C'^{(2)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_C^{(2)}}\right)^{\uparrow}$ and $\overrightarrow{L_C'^{(2)}} \succeq_{\mathsf{coor}} \overrightarrow{R_C^{(2)}}$.

Let us see why the above conditions suffice to prove the inequality. The second part of (b) implies that $\left\|\overrightarrow{R_A}\right\|_{1/2} \leq \left\|\overrightarrow{R_A^{(1)}} + \overrightarrow{R_A^{(3)}} + \overrightarrow{L_A'^{(2)}}\right\|_{1/2}$. Part (a) and the first part of (b), along with Lemma 2.11, implies

$\overrightarrow{R_A^{(1)}} + \overrightarrow{R_A^{(3)}} + \overrightarrow{L_A'^{(2)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_A}\right)^{\downarrow}$. And so, $\|L_A\|_{1/2} \geq \|R_A\|_{1/2}$. A similar argument using (c) and (d) implies $\|L_C\|_{1/2} \geq \|R_C\|_{1/2}$.

One last observation is needed to complete the proof. Note that $R_W^{(2)}$ is the **zero** vector: the points $x \in W$ don't change value even when ℓ is sorted. Also note that $L_W^{(2)}$ is the zero vector; the points $x \in W$ don't participate in a violation in direction i. And therefore, part (c) along with Lemma 2.11 implies $\overrightarrow{R_W} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_W}\right)^{\uparrow}$ implying $\|L_W\|_{1/2} \geq \|R_W\|_{1/2}$. Similarly, $R_O^{(2)} \equiv L_O^{(2)} \equiv \mathbf{0}$, and thus part (a) along with Lemma 2.11 implies $\|L_O\|_{1/2} \geq \|R_O\|_{1/2}$.

6.1.1 Proving (a) and (c) for $q = 3$

Defining the Coloring χ': We will now describe the coloring χ' on all edges of the form $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j)$ where $j \geq i+1$, $h(\mathbf{x}) = 1$ and $h(\mathbf{x} + a\mathbf{e}_j) = 0$. For all other edges e, we simply define $\chi'(e) = \chi(e)$ as these edges do not play a role in proving the inequality.

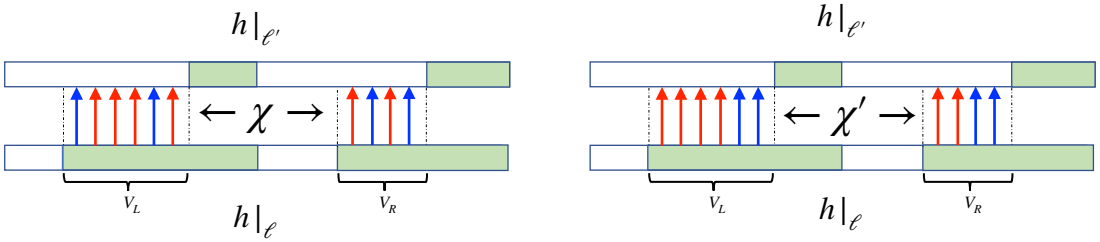
Given a pair of i-lines ℓ and $\ell' = \ell + a\mathbf{e}_j$ for $j \geq i+1$ and $a > 0$, we consider the set of violations from ℓ to ℓ' in h:

$$V := \{(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in \ell,\, h(\mathbf{x}) = 1, \text{ and } h(\mathbf{x} + a\mathbf{e}_j) = 0\}. \tag{15}$$

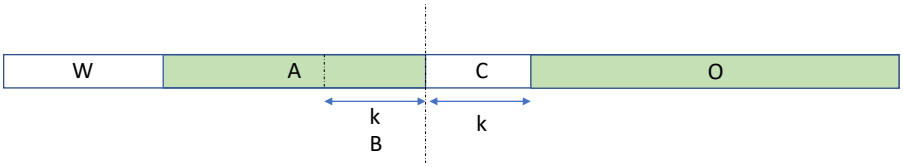
Since h is semi-sorted, it's clear that we can write $V = V_L \cup V_R$ as a union of two intervals, in the sense that $\{\mathbf{x} \colon (\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \in V_L\}$ is an interval in the lower half of ℓ and $\{\mathbf{x} \colon (\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \in V_R\}$ is an interval in the upper half of ℓ. Similarly, the upper endpoints form two intervals in ℓ'. We then obtain χ' by down-sorting χ on each of these intervals, moving left-to-right:

$$(\chi'(e) \colon e \in V_L) = (\chi(e) \colon e \in V_L)^{\downarrow} \text{ and } (\chi'(e) \colon e \in V_R) = (\chi(e) \colon e \in V_R)^{\downarrow}.$$

We provide the following illustration for clarity. The white and green intervals represent where $h = 0$ and $h = 1$, respectively. The vertical arrows represent violated edges. Blue edges have color 0 and red edges have color 1. The left picture depicts the original coloring, χ, and the right picture depicts the recoloring χ'.



We now return to our fixed i-line ℓ and set out to prove parts (a) and (c) for $q = 3$, given this coloring χ'. Let's recall our illustration of ℓ and our definition of the intervals W, A, C, O.



Proving (a) for $q = 3$: Fix $j \geq i+1$ and a i-line $\ell' := \ell + a\mathbf{e}_j$. Let $A' := \{\mathbf{x} \in A \ : \ h(\mathbf{x} + a\mathbf{e}_j) = 0\}$ and $O' := \{\mathbf{x} \in O \ : \ h(\mathbf{x} + a\mathbf{e}_j) = 0\}$. Since h is semi-sorted, it is not hard to see that A' and O' are prefixes of A and O, respectively.

Claim 6.1. *If $\mathbf{x}_i < \mathbf{x}'_i$ in A such that $\mathbf{x}' \in A'$, then $\mathbf{x} \in A'$. The same is true for O and O'.*

Proof. Since h is semisorted, $h(\mathbf{x}' + a\mathbf{e}_j) = 0$ implies $h(\mathbf{x} + a\mathbf{e}_j) = 0$. □

Moreover, observe that our definition of χ' gives us

$$(\chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in A') = \big(\chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in A'\big)^{\downarrow}$$

and

$$(\chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in O') = \big(\chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in O'\big)^{\downarrow}.$$

Let's investigate what this leads to. These are key properties.

Definition 6.2. *Fix $j \geq i+1$ and fix an i-line $\ell' := \ell + a\mathbf{e}_j$ for $a > 0$. Define the following two boolean vectors*

$$\mathbf{v}^R_{j,a} := \big(\mathbf{1}\big(h(\mathbf{x} + a\mathbf{e}_j) = 0 \ \textbf{\textit{and}} \ \ \chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1\big) \ \ : \ \mathbf{x} \in A\big)$$

and

$$\mathbf{v}^L_{j,a} := (\mathbf{1}(h(\mathbf{x} + a\mathbf{e}_j) = 0 \ \textbf{\textit{and}} \ \ \chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1) \ \ : \ \mathbf{x} \in A)$$

Observe, for $\mathbf{x} \in A$,

$$\Phi_{h,\chi'}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^R_{j,a}(\mathbf{x})\right) \ \text{ and } \ \Phi_{h,\chi}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right) \tag{16}$$

Claim 6.3. *Fix a $j \geq i+1$ and $a > 0$. For any two $\mathbf{x}_i < \mathbf{x}'_i$ in A, we have $\mathbf{v}^R_{j,a}(\mathbf{x}) \geq \mathbf{v}^R_{j,a}(\mathbf{x}')$. That is, the vector $\mathbf{v}^R_{j,a}$ is sorted decreasing.*

Proof. Since h is semisorted $h(\mathbf{x}' + a\mathbf{e}_j) = 0$ implies $h(\mathbf{x} + a\mathbf{e}_j) = 0$. Furthermore, since both these are violations, by design $\chi'(\mathbf{x}', \mathbf{x}' + a\mathbf{e}_j) = 1$ implies $\chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1$. □

Claim 6.4. *Fix a $j \geq i+1$ and $a > 0$. The vectors $\mathbf{v}^R_{j,a}$ and $\mathbf{v}^L_{j,a}$ are permutations of one another.*

Proof. This is precisely how χ' is defined: it only permutes the colorings on the violations incident on A. □

In conclusion, using the observation (16), we conclude that we can write

$$\overrightarrow{L^{(3)}_A} = \left(\sum_{j=i+1}^{d} \Phi_{h,\chi}(\mathbf{x}; j) \ \ : \ \ \mathbf{x} \in A\right)$$

as a weighted sum of Boolean vectors, and the above two claims imply that the vector

$$\overrightarrow{R^{(3)}_A} = \left(\sum_{j=i+1}^{d} \Phi_{h,\chi'}(\mathbf{x}; j) \ \ : \ \ \mathbf{x} \in A\right)$$

is the same weighted sum of the *sorted decreasing* orders of those Boolean vectors. Therefore, we can conclude using Lemma 2.11,

$$\overrightarrow{R_A^{(3)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_A^{(3)}}\right)^{\downarrow} \tag{17}$$

An absolutely analogous argument with O's replacing A's gives us

$$\overrightarrow{R_O^{(3)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_O^{(3)}}\right)^{\downarrow} \tag{18}$$

Proving (c) for $q = 3$: The picture is similar, but reversed, when we consider the points in $W \cup C$, where $h(\mathbf{x}) = 0$. Recall the definition of W and C as in the illustration. Fix $j \geq i+1$ and a i-line $\ell'' := \ell - a\mathbf{e}_j$. Let $W' := \{\mathbf{x} \in W \ : \ h(\mathbf{x} - a\mathbf{e}_j) = 1\}$ and $C' := \{\mathbf{x} \in C \ : \ h(\mathbf{x} - a\mathbf{e}_j) = 1\}$. It is not hard to see that W' and C' are suffixes of W and C, respectively.

Claim 6.5. *If $\mathbf{x}_i < \mathbf{x}'_i$ in W such that $\mathbf{x} \in W'$, then $\mathbf{x}' \in W'$. The same is true for C and C'.*

Proof. Since h is semisorted, $h(\mathbf{x} - a\mathbf{e}_j) = 1$ implies $h(\mathbf{x}' - a\mathbf{e}_j) = 1$. □

Again, observe that our definition of χ' gives us

$$(\chi'(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) \colon \mathbf{x} \in W') = \big(\chi(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) \colon \mathbf{x} \in W'\big)^{\downarrow}$$

and

$$(\chi'(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) \colon \mathbf{x} \in C') = \big(\chi(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) \colon \mathbf{x} \in C'\big)^{\downarrow}$$

Definition 6.6. *Fix $j \geq i+1$ and fix an i-line $\ell'' := \ell - a\mathbf{e}_j$ for $a > 0$. Define the following two boolean vectors*

$$\mathbf{v}^R_{j,a} := \big(\mathbf{1}\big(h(\mathbf{x} - a\mathbf{e}_j) = 1 \ \textbf{\textit{and}} \ \ \chi'(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) = 0\big) \ \ : \ \mathbf{x} \in C\big)$$

and

$$\mathbf{v}^L_{j,a} := (\mathbf{1}(h(\mathbf{x} - a\mathbf{e}_j) = 1 \ \textbf{\textit{and}} \ \ \chi(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) = 0) \ \ : \ \mathbf{x} \in C)$$

Observe, for $\mathbf{x} \in C$,

$$\Phi_{h,\chi'}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^R_{j,a}(\mathbf{x})\right) \ \text{ and } \ \Phi_{h,\chi}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right) \tag{19}$$

Claim 6.7. *Fix a $j \geq i+1$ and $a > 0$. For any two $\mathbf{x}_i > \mathbf{x}'_i$ in C, we have $\mathbf{v}^R_{j,a}(\mathbf{x}) \geq \mathbf{v}^R_{j,a}(\mathbf{x}')$. That is, the vector $\mathbf{v}^R_{j,a}$ is sorted increasing when considered left to right.*

Proof. Since h is semisorted $h(\mathbf{x}' - a\mathbf{e}_j) = 1$ implies $h(\mathbf{x} - a\mathbf{e}_j) = 1$. Furthermore, since both these are violations, by design $\chi'(\mathbf{x}', \mathbf{x}' + a\mathbf{e}_j) = 0$ implies $\chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 0$. □

Claim 6.8. *Fix a $j \geq i+1$ and $a > 0$. The vectors $\mathbf{v}^R_{j,a}$ and $\mathbf{v}^L_{j,a}$ are permutations of one another.*

A similar argument to the one given above now implies $\overrightarrow{L_C^{(3)}}$ is a sum of Boolean vectors, and $\overrightarrow{R_C^{(3)}}$ is the sum of the *sorted increasing* orders of those Boolean vectors. Using Lemma 2.11, we can conclude

$$\overrightarrow{R_C^{(3)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_C^{(3)}}\right)^{\uparrow} \tag{20}$$

And an absolutely analogous argument gives

$$\overrightarrow{R_W^{(3)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_W^{(3)}}\right)^{\uparrow} \tag{21}$$

This finishes the proofs of $q = 3$ for (a) and (c).

6.1.2 Proving (a) and (c) for $q = 1$

Defining $\xi'_{\mathbf{x}}(S, S \oplus j)$ for $S \subseteq [i-1]$ and $j \leq i-1$: We now define the partial coloring $\xi'_{\mathbf{x}} := \xi_{\mathbf{x}}^{(i)}$ on all edges $(S, S \oplus j)$ where $S \subseteq [i-1]$ and $j \leq i-1$ for all $\mathbf{x} \in [n]^d$. These are exactly the relevant edges for the proof of parts (a) and (c) for $q = 1$. Note that the partial coloring $\xi_{\mathbf{x}} := \xi_{\mathbf{x}}^{(i-1)}$ is defined over precisely these edges for each $\mathbf{x} \in [n]^d$. The color of $\xi'_{\mathbf{x}}$ on the edges $(S, S+i)$ for $S \subseteq [i-1]$ will be defined when we prove parts (b) and (d). The color of $\xi'_{\mathbf{x}}$ on the edges $(S+i, S+i \oplus j)$ for $S \subseteq [i-1]$ and $j \leq i-1$ will be defined when we prove $\mathsf{LHS} \geq \mathsf{RHS}_2$.

Fix $j \leq i-1$, $S \subseteq [i-1]$, and a i-line ℓ. We consider the set of $\mathbf{x} \in \ell$ such that $(S, S \oplus j)$ is influential in $g_{\mathbf{x}}$:

$$V := \{\mathbf{x} \in \ell \colon g_{\mathbf{x}}(S) = 1 \text{ and } g_{\mathbf{x}}(S \oplus j) = 0\}. \tag{22}$$

Note that since f is semi-sorted, we have that $(S \circ f)$ and $(S \oplus j \circ f)$ are both semi-sorted. Thus, we can write $V = V_L \cup V_R$ where V_L and V_R are intervals contained in the left and right half of ℓ, respectively. We again obtain $\xi'_{\mathbf{x}}$ by down-sorting the original coloring on these intervals:

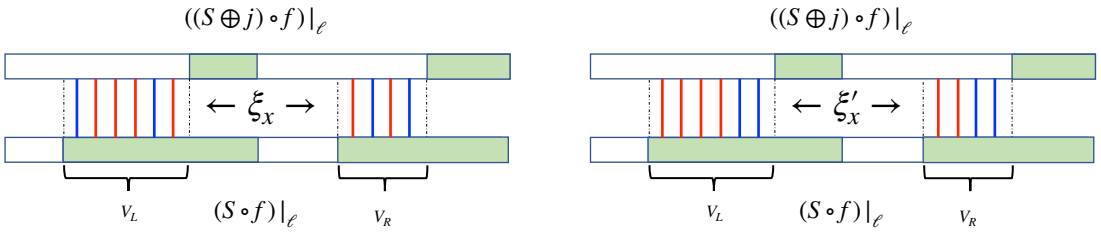
$$(\xi'_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_L) = (\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_L)^{\downarrow}$$

and similarly

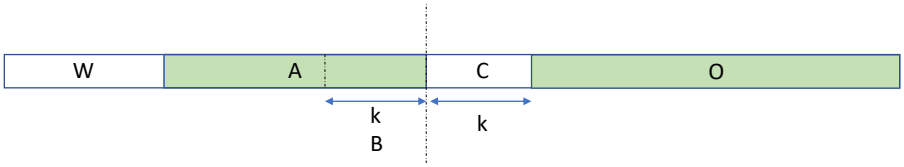
$$(\xi'_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_R) = (\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_R)^{\downarrow}.$$

For all $\mathbf{x} \in \ell \setminus V$, we define $\xi'_{\mathbf{x}}(S, S \oplus j) := \xi_{\mathbf{x}}(S, S \oplus j)$. This completely describes $\xi'_{\mathbf{x}}(S, S \oplus j)$ for every $\mathbf{x} \in [n]^d$.

We provide the following illustration for clarity. Note that the picture is quite similar to the one provided in Section 6.1.1, when we defined χ'. The key difference is that the bottom and top segments represent the same line ℓ, but with different functions $S \circ f$ and $(S \oplus j) \circ f$, respectively. The vertical lines are no longer arrows to emphasize that they represent *undirected edges in the hypercube* as opposed to directed edges in the augmented hypergrid.



We now return to our fixed i-line ℓ and set out to prove parts (a) and (c) for $q = 1$, given the colorings $\xi'_{\mathbf{x}}$. Let's recall our illustration of ℓ and our definition of the intervals W, A, C, O. Recall that $g_{\mathbf{x}} = h(\mathbf{x})$ and so the definition of these intervals is the same.



Proof of Part (a) for $q = 1$: Fix $j \le i-1$ and let $A' = \{\mathbf{x} \in A \colon g_{\mathbf{x}}(S \oplus j) = 0\}$ and $O' = \{\mathbf{x} \in O \colon g_{\mathbf{x}}(S \oplus j) = 0\}$, which are prefixes of A and O, respectively. From our definition of $\xi'_{\mathbf{x}}(S, S \oplus j)$ above, we have

$$(\xi'_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in A') = \left(\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in A'\right)^{\downarrow}$$

and similarly

$$(\xi'_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in O') = \left(\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in O'\right)^{\downarrow}.$$

Claim 6.9. $\left(I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S) \ : \ \mathbf{x} \in A\right)$ *is a sorted decreasing vector, and is a permutation of* $\left(I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) \ : \ \mathbf{x} \in A\right)$.

Proof. Take $\mathbf{x}_i < \mathbf{x}'_i$ in A. Note that $g_{\mathbf{x}}(S) = 1$ for both $\mathbf{x}, \mathbf{x}'$. Thus,

$$I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S) = \mathbf{1}\left(g_{\mathbf{x}}(S \oplus j) = 0 \textbf{ and } \xi'_{\mathbf{x}}(S, S \oplus j) = 1\right)$$

and

$$I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) = \mathbf{1}\left(g_{\mathbf{x}}(S \oplus j) = 0 \textbf{ and } \xi_{\mathbf{x}}(S, S \oplus j) = 1\right)$$

The two vectors are Boolean vectors with number of ones equal to the number of ones in $(\xi_{\mathbf{x}}(S, S \oplus j) \ : \ \mathbf{x} \in A')$ which equals the number of ones in $(\xi'_{\mathbf{x}}(S, S \oplus j) \ : \ \mathbf{x} \in A')$. Thus, they are permutations. By design of $\xi'_{\mathbf{x}}$'s, this vector is sorted decreasing on A', and all zeros in $A \setminus A'$ (which come to the right of A'). □

Observing that

$$\overrightarrow{L^{(1)}_A} = \left(\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) \ : \ \mathbf{x} \in A\right) \quad \text{and} \quad \overrightarrow{R^{(1)}_A} = \left(\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S) \ : \ \mathbf{x} \in A\right)$$

using Lemma 2.11 and the claim above, we get

$$\overrightarrow{R^{(1)}_A} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_A}\right)^{\downarrow} \tag{23}$$

Absolutely analogously, we get

$$\overrightarrow{R^{(1)}_O} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_O}\right)^{\downarrow} \tag{24}$$

Proof of Part (c) for $q = 1$: The picture is similar, but reversed when we consider the points in $W \cup C$, where $g_{\mathbf{x}}(S) = 0$. Fix $j \leq i-1$ and define $W' := \{\mathbf{x} \in W : g_{\mathbf{x}}(S \oplus j) = 1\}$ and $C' := \{\mathbf{x} \in C : g_{\mathbf{x}}(S \oplus j) = 1\}$ which are suffixes of W and C, respectively. From our definition of $\xi'_{\mathbf{x}}(S, S \oplus j)$ above, made from the perspective of the set $S \oplus j$, we have

$$(\xi'_{\mathbf{x}}(S, S \oplus j) : \mathbf{x} \in W') = \left(\xi_{\mathbf{x}}(S, S \oplus j) : \mathbf{x} \in W'\right)^{\downarrow}$$

and similarly

$$(\xi'_{\mathbf{x}}(S, S \oplus j) : \mathbf{x} \in C') = \left(\xi_{\mathbf{x}}(S, S \oplus j) : \mathbf{x} \in C'\right)^{\downarrow}.$$

Analogous to Claim 6.9, we have the following claim.

Claim 6.10. *$\left(I^{=j}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S) \;:\; \mathbf{x} \in W\right)$ is a sorted increasing vector, and is a permutation of $\left(I^{=j}_{g_{\mathbf{x}}, \xi_{\mathbf{x}}}(S) \;:\; \mathbf{x} \in W\right)$.*

Arguing similarly to the proof of Equation (23) we get

$$\overrightarrow{R^{(1)}_W} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_W}\right)^{\uparrow} \tag{25}$$

and absolutely analogously, we get

$$\overrightarrow{R^{(1)}_C} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_C}\right)^{\uparrow} \tag{26}$$

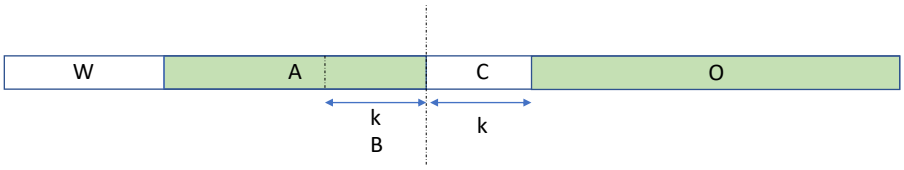
(17), (18), (20), (21), and (23), (24), (25), (26) establish (a) and (c).

6.1.3 Proving (b) and (d):

Finally, we need to establish (b) and (d). Let us recall these and also draw the picture of ℓ that we have been using.

(b) $\exists \overrightarrow{L'^{(2)}_A}$ such that $\overrightarrow{L'^{(2)}_A} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(2)}_A}\right)^{\downarrow}$ and $\overrightarrow{L'^{(2)}_A} \succeq_{\mathsf{coor}} \overrightarrow{R^{(2)}_A}$.

(d) $\exists \overrightarrow{L'^{(2)}_C}$ such that $\overrightarrow{L'^{(2)}_C} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(2)}_C}\right)^{\uparrow}$ and $\overrightarrow{L'^{(2)}_C} \succeq_{\mathsf{coor}} \overrightarrow{R^{(2)}_C}$.



We remind the reader that $\overrightarrow{L^{(2)}}(\mathbf{x}) = \Phi_{h,\chi}(\mathbf{x}; i)$ for all $\mathbf{x} \in \ell$. We begin with an observation which strongly uses the "thresholded" nature of the definition of Φ.

Claim 6.11. *No matter how χ is defined, either $\overrightarrow{L^{(2)}_A}$ is the all 1s vector, or $\overrightarrow{L^{(2)}_C}$ is the all 1s vector.*

Proof. Suppose for the sake of contradiction, there exists $\mathbf{x} \in A$ and $\mathbf{y} \in C$ such that $\Phi_{h,\chi}(\mathbf{x}; i) = \Phi_{h,\chi}(\mathbf{y}; i) = 0$. But the edge $(\mathbf{x}, \mathbf{y})$ is a violation, and if $\chi(\mathbf{x}, \mathbf{y}) = 1$ then $\Phi_{h,\chi}(\mathbf{x}; i) = 1$, otherwise $\Phi_{h,\chi}(\mathbf{y}; i) = 1$. Contradiction. □

Next we remind the reader that $\overrightarrow{R^{(2)}}(\mathbf{x}) = I^{=i}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S)$. We now define the $\xi'_{\mathbf{x}}(S, S+i)$ colorings for $\mathbf{x} \in A \cup C$ using the above claim in the following simple manner.

$$\text{If } \overrightarrow{L^{(2)}_A} \equiv \mathbf{1}, \text{ then } \xi'_{\mathbf{x}}(S, S+i) = 1 \ \ \forall \mathbf{x} \in A \cup C \tag{27}$$

otherwise,

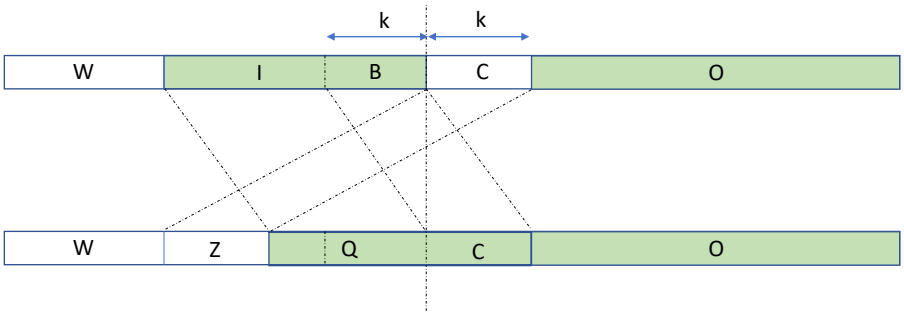
$$\text{we have } \overrightarrow{L^{(2)}_C} \equiv \mathbf{1}, \text{ and so we define } \xi'_{\mathbf{x}}(S, S+i) = 0 \ \ \forall \mathbf{x} \in A \cup C \tag{28}$$

In the former case, we have $\overrightarrow{R^{(2)}_A} = (\underbrace{111\cdots1}_{k \text{ many}}0000)$ and $\overrightarrow{L^{(2)}_A} \equiv \mathbf{1}$ and so we pick $\overrightarrow{L'^{(2)}_A} = \overrightarrow{L^{(2)}_A}$. Also note that we have $\overrightarrow{R^{(2)}_C}$ as the all zeros vector, and so we pick $\overrightarrow{L'^{(2)}_C} = \left(\overrightarrow{L^{(2)}_C}\right)^{\uparrow}$. These satisfy (b) and (d). In the latter case the argument is analogous. Thus, in either case we have established (b) and (d), and this completes the proof of $\mathsf{LHS} \geq \mathsf{RHS}_1$.

We remind the reader that we have now defined $\xi'_{\mathbf{x}}(S, S \oplus j)$ for all subsets $S \subseteq [i-1]$ and $1 \leq j \leq i$. In the next subsection, when we prove $\mathsf{LHS} \geq \mathsf{RHS}_2$, we will need to define $\xi'_{\mathbf{x}}(S+i, S+i \oplus j)$ for all $j \leq i-1$. Note that for $j = i$, we have $(S+i, S+i \oplus j) = (S+i, S)$ and the coloring $\xi'_{\mathbf{x}}$ has already been defined for these edges in (27) or (28).

6.2 Proving $\mathsf{LHS} \geq \mathsf{RHS}_2$

This inequality is a bit trickier to establish because the function h itself now changes to $i \circ h$ in RHS_2. For instance, focusing on the illustration we have been using, upon sorting the picture looks like this.



We have now partitioned the interval A into $I \cup B$ where B is the k-ones closest to the semi-sorting boundary. After sorting, we think of the ones in B moving into C, and the ones in I shifting and moving to $Q \subseteq A$. The first k entries of A, which we call Z, takes the value 0 after sorting this line.

To argue $\mathsf{LHS} \geq \mathsf{RHS}_2$, we break the vector $\vec{L}$ as

$$\left\|\vec{L}\right\|_{1/2} = \left\|\vec{L}_W\right\|_{1/2} + \left\|\vec{L}_{I \cup B \cup O}\right\|_{1/2} + \left\|\vec{L}_C\right\|_{1/2}$$

and the vector $\vec{M}$ as

$$\left\|\vec{M}\right\|_{1/2} = \left\|\overrightarrow{M_W}\right\|_{1/2} + \left\|\overrightarrow{M_{Q \cup C \cup O}}\right\|_{1/2} + \left\|\overrightarrow{M_Z}\right\|_{1/2}$$

and argue vector-by-vector. The plan of the proof is similar to the previous case. We want to find recolorings χ'' and $\xi'_{\mathbf{x}}$ such that

(e) $\exists \overrightarrow{M'^{(q)}_{QCO}}$ such that $\overrightarrow{M'^{(q)}_{QCO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(q)}_{IBO}}\right)^{\downarrow}$ and $\overrightarrow{M'^{(q)}_{QCO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(q)}_{QCO}}$, for $q \in \{1,3\}$.

(f) $\exists \overrightarrow{L'^{(2)}_{IBO}}$ such that $\overrightarrow{L'^{(2)}_{IBO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(2)}_{IBO}}\right)^{\downarrow}$ and $\overrightarrow{L'^{(2)}_{IBO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(2)}_{QCO}}$.

(g) $\exists \overrightarrow{M'^{(q)}_{WZ}}$ such that $\overrightarrow{M'^{(q)}_{WZ}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(q)}_{WC}}\right)^{\uparrow}$ and $\overrightarrow{M'^{(q)}_{WZ}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(q)}_{WZ}}$, for $q \in \{1,3\}$.

(h) $\exists \overrightarrow{L'^{(2)}_{WC}}$ such that $\overrightarrow{L'^{(2)}_{WC}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(2)}_{WC}}\right)^{\uparrow}$ and $\overrightarrow{L'^{(2)}_{WC}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(2)}_{WZ}}$.

Let us see why the above conditions suffice to prove the inequality. The second part of (f) implies that $\left\|\overrightarrow{M_{QCO}}\right\|_{1/2} \leq \left\|\overrightarrow{M^{(1)}_{QCO}} + \overrightarrow{M^{(3)}_{QCO}} + \overrightarrow{L'^{(2)}_{IBO}}\right\|_{1/2}$. Part (e) and the first part of (f), along with Lemma 2.11, implies $\overrightarrow{M^{(1)}_{QCO}} + \overrightarrow{M^{(3)}_{QCO}} + \overrightarrow{L'^{(2)}_{IBO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_{IBO}}\right)^{\downarrow}$. And so, $\|L_{IBO}\|_{1/2} \geq \|M_{IBO}\|_{1/2}$. Now, by the second part of (g) and the second part of (h) we have $\left\|\overrightarrow{M_{WC}}\right\|_{1/2} \leq \left\|\overrightarrow{M'^{(1)}_{WC}} + \overrightarrow{M'^{(3)}_{WC}} + \overrightarrow{L'^{(2)}_{WC}}\right\|_{1/2}$ and by the first part of (g) and (h) we have $\left\|\overrightarrow{M'^{(1)}_{WC}} + \overrightarrow{M'^{(3)}_{WC}} + \overrightarrow{L'^{(2)}_{WC}}\right\|_{1/2} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_{WC}}\right)^{\uparrow}$. Thus, $\|L_{WZ}\|_{1/2} \geq \|M_{WZ}\|_{1/2}$.

6.2.1 Proving (e) and (g) for $q = 3$

Defining the Coloring χ'': We now describe the coloring χ'' on all edges of the form $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j)$ where $j \geq i+1$, $(i \circ h)(\mathbf{x}) = 1$ and $(i \circ h)(\mathbf{x} + a\mathbf{e}_j) = 0$. For all other edges e, we simply define $\chi''(e) = \chi(e)$.

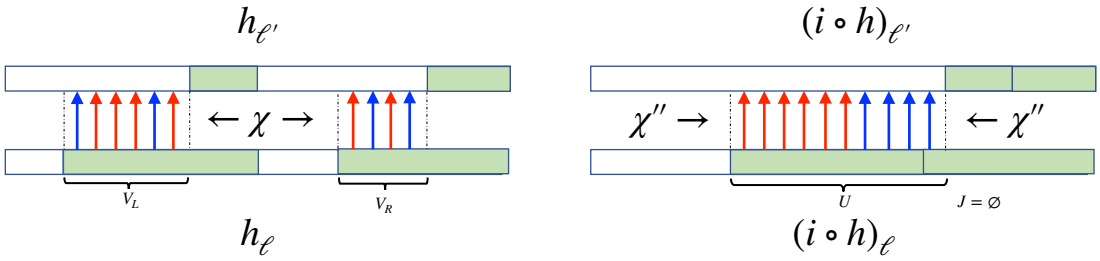
Given a pair of i-lines ℓ and $\ell' = \ell + a\mathbf{e}_j$ for $j \geq i+1$ and $a > 0$ we consider the set of violations from ℓ to ℓ' in h and in $i \circ h$. As before, the violations in h form two a union of two intervals $V = V_L \cup V_R$. Recall the definition of V in (15). Since $(i \circ h)$ is sorted in dimension i, the violations from ℓ to ℓ' in $(i \circ h)$ form a single interval which we will call U:

$$U := \{(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in \ell,\ (i \circ h)(\mathbf{x}) = 1), \text{ and } (i \circ h)(\mathbf{x} + a\mathbf{e}_j) = 0\}.$$

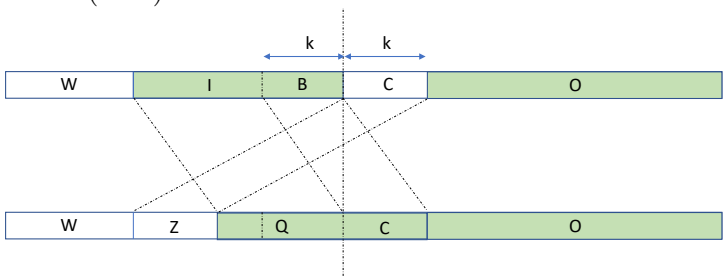
Since the sort operator can only reduce the number of violations in a dimension, we have $|U| \leq |V|$ (Claim 2.6 applied to $h|_\ell$ and $h|_{\ell'}$). We define J to be the interval of $|V| - |U|$ points directly to the right of U so that $U \cup J$ is an interval of size $|V|$. We then define

$$(\chi''(e) \colon \mathbf{x} \in U \cup J) = (\chi(e) \colon e \in V)^{\downarrow}.$$

We now have a complete description of χ''. We provide the following illustration for clarity. The white and green intervals represent where $h = 0$ and $h = 1$, respectively. The vertical arrows represent violated edges. Blue edges have color 0 and red edges have color 1. The left picture depicts the original coloring, χ, and the original function, h. The right picture depicts the recoloring, χ'', and the function after sorting, $i \circ h$.



We now return to our fixed i-line ℓ and set out to prove (e) and (g) for $q = 3$, given this coloring χ''. Let's recall our illustration of h and $(i \circ h)$ restricted to ℓ and our definition of the intervals W, I, B, C, O, Z, Q.



Proving (e) for $q = 3$: Recall the definition of $A = I \cup B$, O, and $Q \cup C \cup O$ as in the illustration. Fix $j \geq i+1$ and a i-line $\ell' = \ell + a\mathbf{e}_j$. Let $A' := \{\mathbf{x} \in A\colon h(\mathbf{x}+a\mathbf{e}_j) = 0\}$, $O' := \{\mathbf{x} \in O\colon h(\mathbf{x}+a\mathbf{e}_j) = 0\}$, and $U := \{\mathbf{x} \in Q \cup C \cup O\colon (i \circ h)(\mathbf{x} + a\mathbf{e}_j) = 0\}$. Again, applying Claim 2.6 to $h|_\ell$ and $h|_{\ell'}$, we have $|U| \leq |A'| + |O'|$. Let J denote the interval of size $|A'| + |O'| - |U|$ directly to the right of U so that $U \cup J$ is an interval of size $|A'| + |O'|$. Observe that by our definition of χ'' above, we have

$$(\chi''(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j)\colon \mathbf{x} \in U \cup J) = \big(\chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j)\colon \mathbf{x} \in A' \cup O'\big)^{\downarrow}.$$

Let's see what this leads to.

Definition 6.12. *Fix $j \geq i+1$ and fix an i-line $\ell' := \ell + a\mathbf{e}_j$ for $a > 0$. Define the following two boolean vectors:*

$$\mathbf{v}^M_{j,a} := \big(\mathbf{1}\big((i \circ h)(\mathbf{x} + a\mathbf{e}_j) = 0 \ \textbf{\textit{and}} \ \ \chi''(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1\big) \ : \ \mathbf{x} \in Q \cup C \cup O\big)$$

and

$$\mathbf{v}^L_{j,a} := (\mathbf{1}(h(\mathbf{x} + a\mathbf{e}_j) = 0 \ \textbf{\textit{and}} \ \ \chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1) \ : \ \mathbf{x} \in I \cup B \cup O).$$

Observe, for $\mathbf{x} \in Q \cup C \cup O$,

$$\Phi_{i\circ h, \chi''}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^M_{j,a}(\mathbf{x})\right) \tag{29}$$

and for $\mathbf{x} \in I \cup B \cup O$,

$$\Phi_{h,\chi}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right). \tag{30}$$

Claim 6.13. *Fix $j \geq i+1$ and $a > 0$. For any two $\mathbf{x}_i < \mathbf{x}'_i$ in $Q \cup C \cup O$, we have $\mathbf{v}^M_{j,a}(\mathbf{x}) \geq \mathbf{v}^M_{j,a}(\mathbf{x}')$. That is, the vector $\mathbf{v}^M_{j,a}$ is sorted decreasing.*

Proof. Since $(i \circ h)$ is semisorted $(i \circ h)(\mathbf{x}' + a\mathbf{e}_j) = 0$ implies $(i \circ h)(\mathbf{x} + a\mathbf{e}_j) = 0$. Furthermore, since both these are violations, by design $\chi''(\mathbf{x}', \mathbf{x}' + a\mathbf{e}_j) = 1$ implies $\chi''(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 1$. □

Claim 6.14. *Fix $j \geq i+1$ and $a > 0$. The vector $\mathbf{v}^M_{j,a}$ has at most as many 1s as $\mathbf{v}^L_{j,a}$ and thus $\left(\mathbf{v}^L_{j,a}\right)^{\downarrow} \succeq_{\mathsf{coor}} \mathbf{v}^M_{j,a}$.*

Proof. This is precisely how χ'' is defined: it only permutes the colorings on the violations incident on $I \cup B \cup O$, and this number can only decrease upon sorting (Claim 2.6 applied to χ restricted to the edges going from ℓ to ℓ'). □

In conclusion, we can write

$$\overrightarrow{L^{(3)}_{IBO}} = \left(\sum_{j=i+1}^{d} \Phi_{h,\chi}(\mathbf{x}; j) \;\; : \;\; \mathbf{x} \in I \cup B \cup O \right)$$

as a sum of Boolean vectors, and the above two claims imply that the vector

$$\overrightarrow{M^{(3)}_{QCO}} = \left(\sum_{j=i+1}^{d} \Phi_{(i\circ h),\chi''}(\mathbf{x}; j) \;\; : \;\; \mathbf{x} \in Q \cup C \cup O \right)$$

is *coordinate wise* dominated by the sum of the *sorted decreasing* orders of those Boolean vectors. Defining $\overrightarrow{M'^{(3)}_{QCO}}$ to be the sum of the sorted decreasing orders, using Lemma 2.11, we establish part (e) for $q = 3$. Namely, we get

$$\exists \overrightarrow{M'^{(3)}_{QCO}} : \;\; \overrightarrow{M'^{(3)}_{QCO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_{IBO}}\right)^{\downarrow} \text{ and } \overrightarrow{M'^{(3)}_{QCO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(3)}_{QCO}} \tag{31}$$

Proving (g) for $q = 3$: A similar argument but working with the zeros establishes part (g) for $q = 3$. The picture is similar, but reversed, when we consider the points in $W \cup C$, where $h(\mathbf{x}) = 0$. Fix a dimension $j \geq i+1$ and some $\ell'' = \ell - ae_j$. Let $W' = \{\mathbf{x} \in W : h(\mathbf{x} - a\mathbf{e}_j) = 1\}$, $C' = \{\mathbf{x} \in C : h(\mathbf{x} - a\mathbf{e}_j) = 1\}$, and $U = \{\mathbf{x} \in W \cup Z : (i \circ h)(\mathbf{x} - a\mathbf{e}_j) = 1\}$. Note that $|U| \leq |W'| + |C'|$ (Claim 2.6 applied to $h|_{\ell''}$ and $h|_{\ell}$). Let J denote the interval of $|W'| + |C'|$ directly to the right of $|U|$ so that $U \cup J$ is an interval of size $|W'| + |C'|$. Observe that by our definition of χ'' above, we have

$$(\chi''(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) : \mathbf{x} \in U \cup J) = \left(\chi(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) : \mathbf{x} \in W' \cup C'\right)^{\downarrow}.$$

Let's see what this leads to.

Definition 6.15. *Fix $j \geq i+1$ and fix an i-line $\ell'' := \ell - a\mathbf{e}_j$ for $a > 0$. Define the following two boolean vectors:*

$$\mathbf{v}^M_{j,a} := \left(\mathbf{1}\left((i \circ h)(\mathbf{x} - a\mathbf{e}_j) = 1 \;\; \textbf{\textit{and}} \;\; \chi''(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) = 0\right) \;\; : \;\; \mathbf{x} \in W \cup Z\right)$$

and

$$\mathbf{v}^L_{j,a} := \left(\mathbf{1}(h(\mathbf{x} - a\mathbf{e}_j) = 1 \;\; \textbf{\textit{and}} \;\; \chi(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) = 0) \;\; : \;\; \mathbf{x} \in W \cup C\right).$$

Observe, for $\mathbf{x} \in W \cup Z$,

$$\Phi_{i\circ h,\chi''}(\mathbf{x};j) = \min\left(1, \sum_a \mathbf{v}^M_{j,a}(\mathbf{x})\right) \tag{32}$$

and for $\mathbf{x} \in W \cup C$,

$$\Phi_{h,\chi}(\mathbf{x};j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right). \tag{33}$$

Claim 6.16. *Fix $j \geq i+1$ and $a > 0$. For any two $\mathbf{x}_i < \mathbf{x}'_i$ in $W \cup Z$, we have $\mathbf{v}^M_{j,a}(\mathbf{x}) \leq \mathbf{v}^M_{j,a}(\mathbf{x}')$. That is, the vector $\mathbf{v}^M_{j,a}$ is sorted increasing.*

Proof. Since $(i \circ h)$ is sorted in dimension i, we have $(i \circ h)(\mathbf{x} - a\mathbf{e}_j) = 1$ implies $(i \circ h)(\mathbf{x}' - a\mathbf{e}_j) = 1$. Furthermore, since both these are violations, by design $\chi''(\mathbf{x} - a\mathbf{e}_j, \mathbf{x}) = 0$ implies $\chi''(\mathbf{x}' - a\mathbf{e}_j, \mathbf{x}') = 0$. □

Claim 6.17. *Fix $j \geq i+1$ and $a > 0$. The vector $\mathbf{v}^M_{j,a}$ has at most as many 1s as $\mathbf{v}^L_{j,a}$ and thus $\left(\mathbf{v}^L_{j,a}\right)^{\uparrow} \succeq_{\mathsf{coor}} \mathbf{v}^M_{j,a}$.*

Proof. This is precisely how χ'' is defined: it only permutes the colorings on the violations incident on $W \cup C$, and this number can only decrease upon sorting (Claim 2.6 applied to χ restricted to the edges going from ℓ'' to ℓ). □

In conclusion, we can write

$$\overrightarrow{L^{(3)}_{WC}} = \left(\sum_{j=i+1}^{d} \Phi_{h,\chi}(\mathbf{x};j) \;:\; \mathbf{x} \in W \cup C\right)$$

as a sum of Boolean vectors, and the above two claims imply that the vector

$$\overrightarrow{M^{(3)}_{WZ}} = \left(\sum_{j=i+1}^{d} \Phi_{(i\circ h),\chi''}(\mathbf{x};j) \;:\; \mathbf{x} \in W \cup Z\right)$$

is *coordinate wise* dominated by the sum of the *sorted increasing* orders of those Boolean vectors. Defining $\overrightarrow{M'^{(3)}_{WZ}}$ to be the sum of the sorted decreasing orders, using Lemma 2.11, we establish part (g) for $q = 3$. Namely, we get

$$\exists \overrightarrow{M'^{(3)}_{MZ}} : \; \overrightarrow{M'^{(3)}_{MZ}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_{WC}}\right)^{\uparrow} \text{ and } \overrightarrow{M'^{(3)}_{WZ}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(3)}_{WZ}} \tag{34}$$

6.2.2 Proving (e) and (g) for $q = 1$

Defining $\xi'_{\mathbf{x}}(S+i, S+i\oplus j)$ for $S \subseteq [i-1]$ and $j \leq i-1$: We now define the partial coloring $\xi'_{\mathbf{x}} := \xi^{(i)}_{\mathbf{x}}$ on all edges $(S+i, S+i\oplus j)$ where $S \subseteq [i-1]$ and $j \leq i-1$ for all $\mathbf{x} \in [n]^d$. These are exactly the relevant edges for the proof of parts (e) and (g) for $q = 1$. Note that the partial coloring $\xi_{\mathbf{x}} := \xi^{(i-1)}_{\mathbf{x}}$ is undefined over these edges.

Fix $S \subseteq [i-1]$, $j \leq i-1$, and a i-line ℓ. We consider the set of $\mathbf{x} \in \ell$ such that $(S, S \oplus j)$ is influential in $g_{\mathbf{x}}$ and the set of edges where $(S+i, S+i \oplus j)$ is influential in $g_{\mathbf{x}}$. As before, the former is a union of two intervals $V = V_L \cup V_R$. Recall the definition of V in (22). Since $(S+i) \circ f$ and $(S+i \oplus j) \circ f$ are both sorted in dimension i, the set of $\mathbf{x} \in \ell$ such that $(S+i, S+i \oplus j)$ is influential forms a single interval which we will call U:

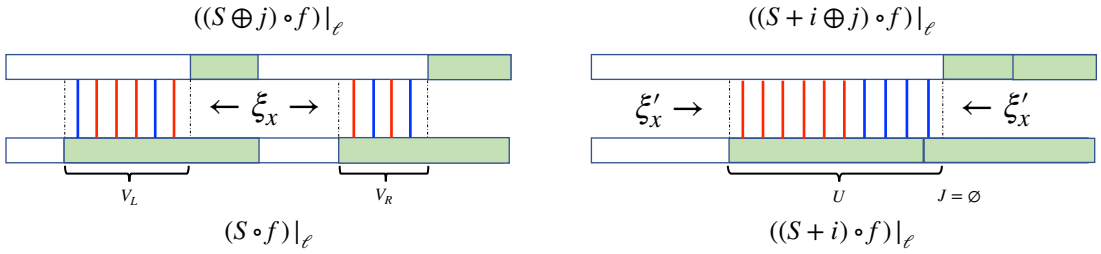
$$U := \{\mathbf{x} \in \ell \colon g_{\mathbf{x}}(S+i) = 1 \text{ and } g_{\mathbf{x}}(S+i \oplus j) = 0\}.$$

Again, we have $|U| \leq |V|$ (Claim 2.6 applied to $(S \circ f)|_\ell$ and $((S \oplus j) \circ f)|_\ell$) and we let J denote the $|V| - |U|$ points directly right of U, so that $U \cup J$ is an interval of length $|V|$. We then define

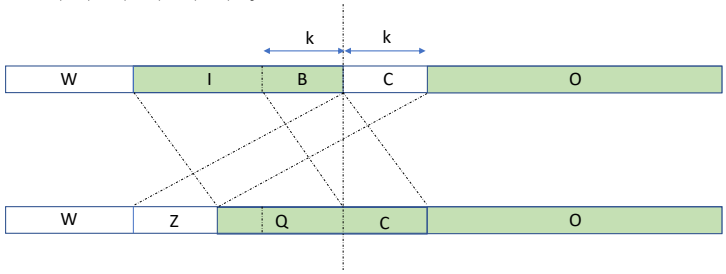
$$(\xi'_{\mathbf{x}}(S+i, S+i \oplus j) \colon \mathbf{x} \in U \cup J) = (\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V)^{\downarrow}.$$

For all $x \in \ell \setminus (U \cup J)$ we define $\xi'_{\mathbf{x}}(S+i, S+i \oplus j) = 1$. Note that this is an arbitrary choice since such edges are not influential and so they do not come in to play in the rest of the proof.

We now have a complete description of $\xi'_{\mathbf{x}}$ on $(S+i, S+i \oplus j)$ for all $\mathbf{x} \in [n]^d$. We provide the following illustration for clarity, which is quite similar to the illustration provided in Section 6.2.1 when we defined χ''. The left picture depicts the original colorings, $\xi_{\mathbf{x}}$, and the relevant functions before applying the sort operator in dimension i. The right picture depicts the recoloring, $\xi'_{\mathbf{x}}$, and the relevant functions after applying the sort operator in dimension i.



We now return to our fixed i-line ℓ and set out to prove (e) and (g) for $q = 1$, given the colorings $\xi'_{\mathbf{x}}$. Recall $g_{\mathbf{x}}(S) = h(\mathbf{x})$ and $g_{\mathbf{x}}(S+i) = (i \circ h)(\mathbf{x})$ and so we can reference the same illustration and our definition of the intervals W, I, B, C, O, Z, Q.



Proving (e) for $q = 1$: Fix $j \leq i-1$ and let $A' := \{\mathbf{x} \in A \colon g_{\mathbf{x}}(S \oplus j) = 0\}$, $O' := \{\mathbf{x} \in O \colon g_{\mathbf{x}}(S \oplus j) = 0\}$, and $U := \{\mathbf{x} \in Q \cup C \cup O \colon g_{\mathbf{x}}(S+i \oplus j) = 0\}$. As before, $|U| \leq |A'| + |O'|$ (applying Claim 2.6 to $(S \circ f)|_\ell$ and $((S \oplus j) \circ f)|_\ell$) and we define J to be the $|A'| - |O'|$ points directly to the right of U so that $U \cup J$ is a prefix of $Q \cup C \cup O$ of size $|A'| + |O'|$. From our definition of $\xi'_{\mathbf{x}}$ from above we have

$$(\xi'_{\mathbf{x}}(S+i, S+i \oplus j) \colon \mathbf{x} \in U \cup J) = \big(\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in A' \cup O'\big)^{\downarrow}.$$

We now get the following claim.

Claim 6.18. *$\left(I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) \;:\; \mathbf{x}\in Q\cup C\cup O\right)$ is a sorted decreasing vector, and has at most as many ones as the vector $\left(I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) \;:\; \mathbf{x}\in I\cup B\cup O\right)$.*

Proof. Take $\mathbf{x}_i < \mathbf{x}'_i$ in $Q\cup C\cup O$. Note that $g_{\mathbf{x}}(S+i) = g_{\mathbf{x}'}(S+i) = 1$ by definition $Q\cup C\cup O$. Thus,

$$I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) = \mathbf{1}\left(g_{\mathbf{x}}(S+i\oplus j)=0 \textbf{ and } \xi'_{\mathbf{x}}(S+i,S+i\oplus j)=1\right)$$

and

$$I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) = \mathbf{1}\left(g_{\mathbf{x}}(S\oplus j)=0 \textbf{ and } \xi_{\mathbf{x}}(S,S\oplus j)=1\right)$$

By design of the $\xi'_{\mathbf{x}}$'s, the first vector is sorted decreasing on $Q\cup C\cup O$ (it takes value 0 after U). Also by design, the number of ones in the latter vector can only be larger since we obtain ξ' by taking a permutation and possibly discarding some ones (the ones corresponding to J). □

Observing that

$$\overrightarrow{L^{(1)}_{IBO}} = \left(\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) \;:\; \mathbf{x}\in I\cup B\cup O\right) \quad\text{and}\quad \overrightarrow{M^{(1)}_{QCO}} = \left(\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) \;:\; \mathbf{x}\in Q\cup C\cup O\right)$$

we see that the latter vector is coordinate-wise dominated by a vector which is a sum of *sorted decreasing* versions of Boolean vectors which add up to the former one. Defining $\overrightarrow{M'^{(1)}_{QCO}}$ to be the sum of the sorted decreasing orders, using Lemma 2.11, we establish part (e) for $q=3$. Namely, we get

$$\exists \overrightarrow{M'^{(1)}_{QCO}}: \quad \overrightarrow{M'^{(1)}_{QCO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_{IBO}}\right)^{\downarrow} \quad\text{and}\quad \overrightarrow{M'^{(1)}_{QCO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(1)}_{QCO}}. \tag{35}$$

Proof of Part (g) for $q=1$: A similar argument but working with the zeros establishes part (g) for $q=1$. Recall the definition of the sets W, C, and Z. Let $W' = \{\mathbf{x}\in W\colon g_{\mathbf{x}}(S\oplus j)=1\}$, $C' = \{\mathbf{x}\in C\colon g_{\mathbf{x}}(S\oplus j)=1\}$, and $U = \{x\in W\cup Z\colon g_{\mathbf{x}}(S+i\oplus j)=1\}$. As before $|U|\le |W'|+|C'|$ (applying Claim 2.6 to $((S\oplus j)\circ f)|_\ell$ and $(S\circ f)|_\ell$) and we define J to be the set of $|W'|+|C'|-|U|$ points directly to the right of U so that $U\cup J$ is an interval of size $|W'|+|C'|$. Note that U is a suffix of $W\cup Z$ and J is a prefix of $Q\cup C\cup O$.

From our definition of $\xi'_{\mathbf{x}}$ above, made with the set $S\oplus j$, we have

$$(\xi'_{\mathbf{x}}(S+i,S+i\oplus j)\colon \mathbf{x}\in U\cup J) = \left(\xi_{\mathbf{x}}(S,S\oplus j)\colon \mathbf{x}\in W'\cup C'\right)^{\downarrow}.$$

Claim 6.19. *$\left(I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) \;:\; \mathbf{x}\in W\cup Z\right)$ is a sorted increasing vector, and has at most as many ones as the vector $\left(I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) \;:\; \mathbf{x}\in W\cup C\right)$.*

Proof. Take $\mathbf{x}_i < \mathbf{x}'_i$ in $W\cup Z$. Note that $g_{\mathbf{x}}(S+i) = g_{\mathbf{x}'}(S+i) = 0$ by definition $W\cup Z$. Thus,

$$I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) = \mathbf{1}\left(g_{\mathbf{x}}(S+i\oplus j)=1 \textbf{ and } \xi'_{\mathbf{x}}(S+i,S+i\oplus j)=0\right)$$

and

$$I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) = \mathbf{1}\left(g_{\mathbf{x}}(S\oplus j)=1 \textbf{ and } \xi_{\mathbf{x}}(S,S\oplus j)=0\right)$$

By design of the $\xi'_{\mathbf{x}}$'s, the first vector is sorted increasing on $W\cup Z$. Also by design, the number of ones in the latter vector can only be larger since we obtain ξ' by taking a permutation and possibly discarding some ones (the ones corresponding to J). □

Observing that

$$\overrightarrow{L^{(1)}_{WC}} = \left(\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S) \ : \ \mathbf{x} \in W \cup C \right) \ \text{ and } \ \overrightarrow{M^{(1)}_{WZ}} = \left(\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) \ : \ \mathbf{x} \in W \cup Z \right)$$

we see that the latter vector is coordinate-wise dominated by a vector which is a sum of *sorted increasing* versions of Boolean vectors which add up to the former one. Defining $\overrightarrow{M'^{(1)}_{WZ}}$ to be the sum of the sorted increasing orders, using Lemma 2.11, we establish part (e) for $q = 3$. Namely,

$$\exists \overrightarrow{M'^{(1)}_{WZ}} : \ \overrightarrow{M'^{(1)}_{WZ}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_{WC}} \right)^{\downarrow} \ \text{ and } \ \overrightarrow{M'^{(1)}_{WZ}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(1)}_{WZ}}. \tag{36}$$

6.2.3 Proving (f) and (h):

Let us now prove part (f) and (h). Note, at this point, $\xi'_{\mathbf{x}}$ is fully defined on all pairs $(S, S \oplus j)$ for $S \subseteq [i]$ and $j \leq i$. We don't have the freedom to redefine. However, we see that the definition we made in (27) and (28) suffices. Let us recall what we want to establish.

(f) $\exists \overrightarrow{L'^{(2)}_{IBO}}$ such that $\overrightarrow{L'^{(2)}_{IBO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(2)}_{IBO}} \right)^{\downarrow}$ and $\overrightarrow{L'^{(2)}_{IBO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(2)}_{QCO}}$.

(h) $\exists \overrightarrow{L'^{(2)}_{WC}}$ such that $\overrightarrow{L'^{(2)}_{WC}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(2)}_{WC}} \right)^{\uparrow}$ and $\overrightarrow{L'^{(2)}_{WC}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(2)}_{WZ}}$.

We remind the reader that $\overrightarrow{L^{(2)}}(\mathbf{x}) = \Phi_{h,\chi}(\mathbf{x}; i)$ for all $\mathbf{x} \in \ell$ and the coloring was defined as follows:

$$\text{If } \ \overrightarrow{L^{(2)}_{IB}} \equiv \mathbf{1}, \ \text{ then } \ \xi'_{\mathbf{x}}(S, S+i) = 1 \ \ \forall \mathbf{x} \in I \cup B \cup C$$

otherwise,

$$\text{we have } \ \overrightarrow{L^{(2)}_{C}} \equiv \mathbf{1}, \ \text{ and so } \ \xi'_{\mathbf{x}}(S, S+i) = 0 \ \ \forall \mathbf{x} \in I \cup B \cup C$$

We remind the reader that $\overrightarrow{M^{(2)}}(\mathbf{x}) = I^{=i}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i)$ and therefore this is 1 iff $g_{\mathbf{x}}(S+i) \neq g_{\mathbf{x}}(S)$ and $\xi'_{\mathbf{x}}(S, S+i) = g_{\mathbf{x}}(S+i)$. The former implies $\mathbf{x} \in Z \cup C$.

Suppose we are in the first case. Then, $\overrightarrow{M^{(2)}}(\mathbf{x}) = 1$ if and only if $\mathbf{x} \in C$. Since $\overrightarrow{L^{(2)}_{IB}} \equiv \mathbf{1} \succeq_{\mathsf{coor}} \overrightarrow{M^{(2)}_{QC}}$, we can set $\overrightarrow{L'^{(2)}_{IBO}}$ to be the vector that is 1s in $I \cup B$ and 0's in O. This establishes (f). To establish (h), we observe that $\overrightarrow{M^{(2)}_{WZ}}$ is the zero vector, and thus we can choose $\overrightarrow{L'^{(2)}_{WC}}$ to be $\left(\overrightarrow{L^{(2)}_{WC}} \right)^{\uparrow}$.

Suppose we are in the second case. Then, $\overrightarrow{M^{(2)}}(\mathbf{x}) = 1$ if and only if $\mathbf{x} \in Z$. Since $\overrightarrow{L^{(2)}_{C}} \equiv \mathbf{1} \succeq_{\mathsf{coor}} \overrightarrow{M^{(2)}_{Z}}$, we can set $\overrightarrow{L'^{(2)}_{WC}}$ to be the vector that is 1s in C and 0's in W. This establishes (h). To establish (f), we observe that $\overrightarrow{M^{(2)}_{QCO}}$ is the zero vector, and thus we can choose $\overrightarrow{L'^{(2)}_{IBO}}$ to be $\left(\overrightarrow{L^{(2)}_{IBO}} \right)^{\downarrow}$.

In either case, we have established (f) and (h), and thus completed the proof.

7 The Tester and it's Analysis: Proof of Theorem 1.5

With the isoperimetric theorem of Theorem 1.4 in place, we can now design and analyze the monotonicity tester for Boolean hypergrid functions. This section closely follows the analogous analysis in [KMS15], and will lift certain notions from that paper. We do have to make slight adaptations to various arguments therein to account for the hypergrid domain.

We first describe the path tester for the hypergrid.

Input: A Boolean function $f : [n]^d \to \{0,1\}$.

1. Choose $k \in_R \{0, 1, 2, \ldots, \lceil \log d \rceil\}$ uniformly at random. Set $\tau := 2^k$.
2. Choose $\mathbf{x} \in [n]^d$ uniformly at random. Denote $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_d)$.
3. Pick a uniform random subset $R \subseteq [d]$ of τ coordinates.
4. For each $r \in R$, pick uar value $c_r \in [n] \setminus \{\mathbf{x}_r\}$.
5. Generate $\mathbf{z}$ as follows. For every $r \in [d]$, if $r \in R$ *and* $c_r > \mathbf{x}_r$, set $\mathbf{z}_r = c_r$. Else, set $\mathbf{z}_r = \mathbf{x}_r$.
6. If $f(\mathbf{z}) > f(\mathbf{x})$, REJECT.

Figure 6: **Path Tester for Hypergrid Functions**

Clearly, the tester doesn't reject any monotone function. Our main theorem regarding the tester follows. A standard boosting argument gives us the $\widetilde{O}(\varepsilon^{-2} n\sqrt{d})$-query Boolean monotonicity tester on $[n]^d$, proving Theorem 1.5.

Theorem 7.1. *If f is ε-far from monotone, then the path tester for hypergrid functions rejects with probability $\Omega(\frac{\varepsilon^2}{n\sqrt{d}\log^5(nd)})$.*

7.1 Directed random walks, influences, and persistence

One can think of the above tester as obtaining $\mathbf{z}$ by a lazy directed random walk of τ steps from a uniform random $\mathbf{x}$. Note that in some steps we may not move at all; these correspond to coordinates $r \in R$ such that $c_r \leq \mathbf{x}_r$. It will be convenient to define an alternate, but equivalent process that generates the pair $(\mathbf{x}, \mathbf{z})$.

Input: A length parameter τ.

1. In each dimension i, sample a uniform random pair $a_i < b_i$ from $[n]$.
2. Let H be the hypercube formed by $\prod_{i=1}^{d} \{a_i, b_i\}$.
3. Pick a uar point $\mathbf{x}$ from H.
4. Pick a uar subset R of τ coordinates, permuted randomly.
5. Generate $\mathbf{z}$ as follows. For every $r \in [d]$, if $r \in R$ and $\mathbf{x}_r = a_r$, set $\mathbf{z}_r = b_r$. Else, set $\mathbf{z}_r = \mathbf{x}_r$.

Figure 7: **Directed random walks, by sampling hypercubes**

This process generates walks by first sampling a random hypercube H, and then doing a lazy directed walk on H. We first observe that conditioned on the walk length τ, the distribution of $(\mathbf{x}, \mathbf{z})$ pairs generated by the path tester and the above process are identical.

Observation 7.2. *Fix $\tau \in [d]$. Let $\mathcal{D}_1, \mathcal{D}_2$ denote the distributions over $[n]^d \times [n]^d$ described in Fig. 6 and Fig. 7, respectively, conditioned on the walk length τ. Then, for any pair $(\mathbf{x}, \mathbf{z})$ where $\mathbf{x} \preceq \mathbf{z}$, we have $\Pr[(\mathbf{x}, \mathbf{z}) \sim \mathcal{D}_1] = \Pr[(\mathbf{x}, \mathbf{z}) \sim \mathcal{D}_2]$.*

Proof. Let $S = \{i \in [d] \colon \mathbf{z}_i > \mathbf{x}_i\}$. The probability from $\mathcal{D}_1$ is given by

$$\begin{aligned}\Pr[(\mathbf{x}, \mathbf{z}) \sim \mathcal{D}_1] &= \frac{1}{n^d} \cdot \sum_{R \supseteq S \,:\, |R| = \tau} \binom{d}{\tau}^{-1} \prod_{i \in S} \Pr[c_i = \mathbf{z}_i] \prod_{i \in R \setminus S} \Pr[c_i \leq \mathbf{x}_i] \\ &= \frac{1}{n^d} \cdot \binom{d}{\tau}^{-1} \cdot \sum_{R \supseteq S \,:\, |R| = \tau} \prod_{i \in S} \frac{1}{n-1} \prod_{i \in R \setminus S} \frac{\mathbf{x}_i - 1}{n-1}.\end{aligned}$$

We now compute the probability for $\mathcal{D}_2$. For $i \in [d]$, let $\mathcal{E}_i$ be the event that $a_i = \mathbf{x}_i$ or $b_i = \mathbf{x}_i$. Note

$$\Pr[\neg \mathcal{E}_i] = \frac{\binom{n-1}{2}}{\binom{n}{2}} = \frac{n-2}{n} \quad \Rightarrow \Pr[\mathcal{E}_i] = \frac{2}{n}.$$

Let $\mathcal{E}_{\mathbf{x}}$ denote the event that the first point sampled according to $\mathcal{D}_2$ is $\mathbf{x}$. We have

$$\Pr[\mathcal{E}_{\mathbf{x}}] = \prod_{i=1}^{d} \Pr[\mathcal{E}_i] \cdot \frac{1}{2^d} = \left(\frac{2}{n}\right)^d \frac{1}{2^d} = \frac{1}{n^d}.$$

Let $\mathcal{E}_{\mathbf{z}}$ denote the event that the second point is $\mathbf{z}$. We have

$$\Pr\left[\mathcal{E}_{\mathbf{z}} | \mathcal{E}_{\mathbf{x}}\right] = \sum_{R \supseteq S \,:\, |R| = \tau} \binom{d}{\tau}^{-1} \prod_{i \in S} \Pr[a_i = \mathbf{x}_i \text{ and } b_i = \mathbf{z}_i \mid \mathcal{E}_{\mathbf{x}}] \cdot \prod_{i \in R \setminus S} \Pr[b_i = \mathbf{x}_i \mid \mathcal{E}_{\mathbf{x}}]$$

Fix an $i \in S$. We have

$$\Pr[a_i = \mathbf{x}_i \text{ and } b_i = \mathbf{z}_i \mid \mathcal{E}_{\mathbf{x}}] = \Pr[a_i = \mathbf{x}_i \text{ and } b_i = \mathbf{z}_i \mid \mathcal{E}_i] = \frac{\Pr[a_i = \mathbf{x}_i \text{ and } b_i = \mathbf{z}_i]}{\Pr[\mathcal{E}_i]} = \frac{1/\binom{n}{2}}{2/n} = \frac{1}{n-1}.$$

Now fix an $i \in R \setminus S$. We have

$$\Pr[b_i = \mathbf{x}_i \mid \mathcal{E}_{\mathbf{x}}] = \Pr[b_i = \mathbf{x}_i \mid \mathcal{E}_i] = \frac{\Pr[b_i = \mathbf{x}_i]}{\Pr[\mathcal{E}_i]} = \frac{\binom{\mathbf{x}_i - 1}{\binom{n}{2}}}{2/n} = \frac{\mathbf{x}_i - 1}{n-1}.$$

Therefore, $\Pr[(\mathbf{x}, \mathbf{z}) \sim \mathcal{D}_2] = \Pr[\mathcal{E}_{\mathbf{x}}] \cdot \Pr[\mathcal{E}_{\mathbf{z}} | \mathcal{E}_{\mathbf{x}}] = \Pr[(\mathbf{x}, \mathbf{z}) \sim \mathcal{D}_1]$. □

We now have that the random walk distributions described in Fig. 6 and Fig. 7 are equivalent. The former is more convenient to analyze for the final rejection probability, but the latter perspective allows us to prove various influence bounds by piggybacking on the [KMS15] hypercube analysis.

Definition 7.3. *We define the total influence and total negative influence of f as follows.*

$$I_f = n^{-d} \sum_{\mathbf{x} \in [n]^d} \sum_{i=1}^{d} \sum_{c=1}^{n} \mathbf{1}(f(\mathbf{x}) \neq f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}, c, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_d))$$

$$I_f^- = n^{-d} \sum_{\mathbf{x} \in [n]^d : f(\mathbf{x}) = 1} \sum_{i=1}^{d} \sum_{c > \mathbf{x}_i} \mathbf{1}(f(\mathbf{x}) \neq f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}, c, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_d))$$

Note these are different from the thresholded influences and for most functions will grow as n grows.

We can analogously define these influences on the hypercubes H sampled by the process described in Fig. 7. Abusing notation, we will denote these influences as $I_H := I_{f|_H}$ and $I_H^- := I_{f|_H}^-$. A simple, yet important claim relates the expected influence on H to the total influence on the hypergrid. All expectations over H are taken with respect to the distribution described in steps 1-2 of Fig. 7.

Claim 7.4. $\mathbf{E}_H[I_H] = I_f/(n-1)$ *and* $\mathbf{E}_H[I_H^-] = I_f^-/(n-1)$.

Proof. Consider any i-edge $(\mathbf{x}, \mathbf{y})$ of the fully augmented hypergrid. The probability that this edge is present in H can be computed as follows. Firstly, we need $a_i = \mathbf{x}_i$ and $b_i = \mathbf{y}_i$. This happens with probability $\binom{n}{2}^{-1}$. Then, for all $j \neq i$, one of $\{a_j, b_j\}$ needs to be $\mathbf{x}_j$ (note that $\mathbf{x}_j = \mathbf{y}_j$ since $(\mathbf{x}, \mathbf{y})$ is an i-edge). This happens with probability $(1 - (\binom{n-1}{2}/\binom{n}{2}))^{d-1} = (2/n)^{d-1}$.

The total number of influential edges in f is $n^d I_f$. Thus, by linearity of expectation, the expected number of influential edges in H is $\binom{n}{2}^{-1} \times (2/n)^{d-1} \times n^d I_f = 2^d I_f/(n-1)$.

An analogous proof holds for the negative influence. □

By using a lemma of [KMS15], we can prove that if the total influence of f is too large, then the negative influence is also large.

Claim 7.5. *If* $I_f > 9(n-1)\sqrt{d}$, *then* $I_f^- > (n-1)\sqrt{d}$.

Proof. Theorem 9.1 of [KMS15] asserts that, if $I_H > 6\sqrt{d}$, then $I_H^- > I_H/3$. (This holds for any Boolean hypercube function.) If $I_f > 9(n-1)\sqrt{d}$, then by Claim 7.4, $\mathbf{E}_H[I_H] > 9\sqrt{d}$. Hence,

$$\begin{aligned} 9\sqrt{d} < \mathbf{E}_H[I_H] &= \Pr[I_H \leq 6\sqrt{d}]\, \mathbf{E}_H[I_H | I_H \leq 6\sqrt{d}] \;+\; \Pr[I_H > 6\sqrt{d}]\, \mathbf{E}_H[I_H | I_H > 6\sqrt{d}] \\ &< 6\sqrt{d} + \Pr[I_H > 6\sqrt{d}] \mathbf{E}_H[3I_H^- | I_H > 6\sqrt{d}] \leq 6\sqrt{d} + 3\mathbf{E}_H[I_H^-] \end{aligned} \tag{37}$$

Hence, $\mathbf{E}_H[I_H^-] > \sqrt{d}$. By Claim 7.4, $I_f^- > (n-1)\sqrt{d}$. □

One of the crucial definitions is that of *persistence*.

Definition 7.6. *A point* $\mathbf{x} \in [n]^d$ *is called* τ*-persistent if* $\Pr_{\mathbf{z}}[f(\mathbf{x}) = f(\mathbf{z})] \geq 1/2$ *where* $\mathbf{z}$ *is chosen by a* τ*-length directed random walk from* $\mathbf{x}$.

Lemma 7.7. *If* $I_f \leq 9(n-1)\sqrt{d}$, *then the fraction of vertices that are not* τ*-persistent is at most* $C_{per}\tau/\sqrt{d}$ *(where* C_{per} *is an absolute constant).*

Proof. We will analyze the random walk using the distribution described in Fig. 7 and leverage the analysis that [KMS15] use to prove their Lemma 9.3. Let α denote the fraction of τ-non-persistent vertices in the fully augmented hypergrid with respect to our function f. Again, let $\mathcal{D}_1$, $\mathcal{D}_2$ denote the distributions over pairs $(\mathbf{x}, \mathbf{z})$ described in Fig. 6, Fig. 7, respectively, conditioned on the walk length τ. For a fixed H sampled in steps 1-2 of Fig. 7, let $\mathcal{D}_{2,H}$ denote the distribution over $(\mathbf{x}, \mathbf{z})$ described in steps 3-5. Using the definition of persistence and Obs. 7.2, we have

$$\frac{\alpha}{2} \leq \Pr_{(\mathbf{x},\mathbf{z})\sim\mathcal{D}_1}(f(\mathbf{x}) \neq f(\mathbf{z})) = \mathbf{E}_H\left[\Pr_{(\mathbf{x},\mathbf{z})\sim\mathcal{D}_{2,H}}(f(\mathbf{x}) \neq f(\mathbf{z}))\right]. \tag{38}$$

Let $\widehat{\mathcal{D}}_{2,H}$ denote the same distribution as $\mathcal{D}_{2,H}$ except with R being a uar subset of the 0-coordinates of $\mathbf{x}$ (recall step 4 of Fig. 7). I.e. $\widehat{\mathcal{D}}_{2,H}$ is the *non-lazy* walk distribution on H. Let $\mathbf{x} = \mathbf{x}^0, \mathbf{x}^1, \ldots, \mathbf{x}^\tau = \mathbf{z}$ be

the τ steps taken on the walk sampled by $\mathcal{D}_{2,H}$ and let $\mathbf{x} = \widehat{\mathbf{x}}^0, \widehat{\mathbf{x}}^1, \ldots, \widehat{\mathbf{x}}^\tau = \mathbf{z}$ be the τ steps taken on the walk sampled by $\widehat{\mathcal{D}}_{2,H}$. For a fixed H we have

$$\Pr_{(\mathbf{x},\mathbf{z})\sim\mathcal{D}_{2,H}}(f(\mathbf{x}) \neq f(\mathbf{z})) \leq \sum_{\ell=0}^{\tau-1} \Pr\left(f(\mathbf{x}^\ell) \neq f(\mathbf{x}^{\ell+1})\right) \leq \sum_{\ell=0}^{\tau-1} \Pr\left(f(\widehat{\mathbf{x}}^\ell) \neq f(\widehat{\mathbf{x}}^{\ell+1})\right). \tag{39}$$

The first inequality is by a union bound and the second inequality holds because the first walk is lazy and the second is not. More precisely, we can couple the $\tau' \leq \tau$ steps of the lazy-random walk where the point actually moves to the first τ' steps of the second non-lazy walk, and the remaining $\tau - \tau'$ terms of the non-lazy walk can only increase the RHS.

By Lemma 9.4 of [KMS15], the edge $(\widehat{\mathbf{x}}^\ell, \widehat{\mathbf{x}}^{\ell+1})$ is distributed approximately as a uniform random edge in H. In particular, this implies $\Pr\left(f(\widehat{\mathbf{x}}^\ell) \neq f(\widehat{\mathbf{x}}^{\ell+1})\right) \leq C \cdot 2I_H/d$ for an absolute constant C. (Note $2I_H/d$ is the probability of a uniform random edge in H being influential.) Putting (38) and (39) together yields $\alpha \leq \frac{4C\tau}{d}\mathbf{E}_H[I_H]$. By Claim 7.4 we have $\mathbf{E}_H[I_H] \leq 9\sqrt{d}$ and so setting $C_{per} := 36C$ completes the proof. □

7.2 The good subgraph and capturing violations

We now use the isoperimetric theorem of Theorem 1.4 to construct a *good subgraph*, in the parlance of [KMS15].

Theorem 7.8. *There exists a bipartite subgraph $G = (X, Y, E)$ of the fully augmented hypergrid with the following properties.*

- $|X| = \sigma n^d$ *or* $|Y| = \sigma n^d$.
- *Every vertex has degree at most k.*
- *For all* $\mathbf{x} \in X$, $f(\mathbf{x}) = 1$. *For all* $\mathbf{y} \in Y$, $f(\mathbf{y}) = 0$.
- $|E| \geq \sigma k n^d/2$.
- $\sigma\sqrt{k} = \Theta(\varepsilon/\log^2(nd))$.

Proof. Consider the bipartite subgraph consisting of all violations of the fully augmented hypergrid. Consider any bi-coloring χ of the edges of this subgraph. Let $\deg_\chi(\mathbf{x})$ denote the number of violating edges e incident to $\mathbf{x}$ for which $\chi(e) = f(\mathbf{x})$. Note that $\deg_\chi(\mathbf{x}) \geq \Phi_{f,\chi}(\mathbf{x})$. Hence, Theorem 1.4 asserts that $\sum_{\mathbf{x}} \sqrt{\deg_\chi(\mathbf{x})} \geq C' n^d \varepsilon/\log n$, for some absolute constant C'. According to Def. 6.4 of [KMS15], the bipartite graph is $C' n^d \varepsilon/\log n$-robust.

By Lemma 6.5 of [KMS15], any robust bipartite graph has a "good subgraph" satisfying the following bound. (Below, C'' is a constant.)

$$\sigma n^d \times \sqrt{k} \geq \frac{C'\varepsilon n^d}{8\log(nd)\log n} \implies \sigma\sqrt{k} \geq \frac{C''\varepsilon}{\log^2(nd)}$$

One can remove vertices from this good subgraph to ensure that $\sigma\sqrt{k} = \Theta(\varepsilon/\log^2(nd))$. □

For the rest of the analysis we will assume $|X| = \sigma n^d$, without loss of generality. The edges of the good subgraph of Theorem 7.8 are central to the tester analysis. We will need to choose τ carefully to ensure that the analysis carries through. Towards that choice, we will set a convenient bound on σ. We will use C_{lar} to denote a sufficiently large constant that is at least $100C_{per}$ and the constants of Theorem 1.4. (The constant C_{per} is from Lemma 7.7.)

Claim 7.9. *If $\sigma < C_{lar}/\sqrt{d}$, then $I_f^- = \Omega(\varepsilon^2\sqrt{d}/\log^4(nd))$.*

Proof. By the good subgraph properties in Theorem 7.8, $\sqrt{k} = \Omega(\varepsilon\sqrt{d}/\log^2(nd))$. Hence, the number of edges of the good subgraph is at least $\sigma k n^d/2 = (\sigma\sqrt{k}) \times \sqrt{k} \times n^d/2 = \Omega(\varepsilon^2\sqrt{d}n^d/\log^4(nd))$. We divide this bound by n^d to bound I_f^-. □

Essentially, for the analysis, we can ignore the case when σ is too small. With this bound in place, we can now set the right choice of τ based on the good subgraph parameters.

Definition 7.10. *For any $\sigma \geq C_{lar}/\sqrt{d}$, define τ_σ to be the power of 2 in the range $[\sigma\sqrt{d}/C_{lar}, \sigma 2\sqrt{d}/C_{lar}]$. (Since $\sigma\sqrt{d}/C_{lar} \geq 1$, the choice of τ_σ exists.)*

We will now define a particular "edge capturing event" that ensures that the tester finds a violation to monotonicity. The crucial property is that these events are uniquely associated with edges of the good subgraph, and are all disjoint. So, we can lower bound the probability of this event and multiply by the number of edges of the good subgraph.

Definition 7.11. *Let $\mathbf{x} \in X$. We call an edge $(\mathbf{x}, \mathbf{y})$ of the good subgraph* viable *if $\mathbf{y}$ is τ_σ-persistent. The set of* viable coordinates *of $\mathbf{x}$ are the dimensions containing the viable edges incident to $\mathbf{x}$.*

For a viable i-edge $e = (\mathbf{x}, \mathbf{y})$, the capturing event *$\mathcal{C}_e$ is defined as follows. Consider the sampling process of the tester, and condition on $\tau := \tau_\sigma$. We define $\mathcal{C}_e = \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3 \wedge \mathcal{E}_4 \wedge \mathcal{E}_5$ where:*

- *$\mathcal{E}_1$: The point $\mathbf{x}$ is chosen (as the first point).*
- *$\mathcal{E}_2$: The coordinate set R contains i.*
- *$\mathcal{E}_3$: The coordinate c_i is $\mathbf{y}_i$.*
- *$\mathcal{E}_4$: $R \setminus i$ does not contain any viable coordinates of $\mathbf{x}$.*
- *$\mathcal{E}_5$: $f(\mathbf{z}) = 0$.*

The main calculation is to lower bound the probability of the event $\mathcal{C}_e$, for any viable edge.

Lemma 7.12. *For any viable edge $e = (\mathbf{x}, \mathbf{y})$, $\Pr[\mathcal{C}_e | \tau = \tau_\sigma] = \Omega(n^{-d} \times n^{-1} \times \tau_\sigma/d)$.*

Proof. The probability of choosing $\mathbf{x}$ is n^{-d}. The probability that i lies in R is $\binom{d-1}{\tau_\sigma - 1}/\binom{d}{\tau_\sigma} = \tau_\sigma/d$. The probability that c_i equals $\mathbf{y}_i$ is $\frac{1}{n-1}$. That is, $\Pr[\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] = \Omega(n^{-d} \times n^{-1} \times \tau_\sigma/d)$. Thus, it remains to show that $\Pr[\mathcal{E}_4 \wedge \mathcal{E}_5 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] = \Omega(1)$.

Let T denote the set of viable coordinates of $\mathbf{x}$. Since the maximum degree in the good subgraph G is k, we have $|T| \leq k$. Thus,

$$\Pr[(R \setminus i) \cap T = \emptyset] \geq \frac{\binom{d-k}{\tau_\sigma - 1}}{\binom{d}{\tau_\sigma - 1}} = \prod_{\ell=0}^{k-1} \left(1 - \frac{\tau_\sigma - 1}{d - \ell}\right) \geq \left(1 - \frac{\tau_\sigma}{d-k}\right)^k \geq \exp\left(-\frac{k \cdot \tau_\sigma}{d-k}\right) \tag{40}$$

Recall from Theorem 7.8 that $k = \Theta(\frac{\varepsilon^2}{\log^4(nd)} \cdot \frac{1}{\sigma^2})$ and from Claim 7.9 we may assume that $\sigma \geq C_{lar}/\sqrt{d}$. Thus, we have $k = O(d \cdot \frac{\varepsilon^2}{\log^4(nd)})$ and in particular, $k = o(d)$ and so

$$\Pr[(R \setminus i) \cap T = \emptyset] \geq \exp\left(-O\left(\frac{k \cdot \tau_\sigma}{d}\right)\right) = \exp\left(-O\left(\frac{k}{d} \cdot \sigma\sqrt{d}\right)\right) = \exp(-O(\sigma\sqrt{k})) \tag{41}$$

where the second step used our definition of τ_σ and the last step simply used the fact that $k \leq d$. By Theorem 7.8 we have $\sigma\sqrt{k} = \Theta(\varepsilon/\log^2(nd))$ and so we have $\Pr[(R \setminus i) \cap T = \emptyset] \geq 9/10$ as long as nd is at least some constant.

Finally, since $\mathbf{y}$ is τ_σ-persistent, the probability that this random walk from $\mathbf{y}$ ends at $\mathbf{z}$ where $f(\mathbf{z}) = 1$ is at most $1/2$. Thus, by the union bound

$$\Pr[\mathcal{E}_4 \wedge \mathcal{E}_5 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq 1 - (1/10 + 1/2) = 2/5$$

and this completes the proof. □

7.3 Wrapping it all up

We combine all the bounds and calculations to prove that the path tester has an $\widetilde{\Omega}(\varepsilon^2/n\sqrt{d})$ probability of success.

Proof. (of Theorem 7.1) We first take care of some edge cases.

Case 1, $I_f > 9(n-1)\sqrt{d}$: By Claim 7.5, $I_f^- > (n-1)\sqrt{d}$. Thus, the total number of violated edges of the augmented hypergrid is at least $(n-1)\sqrt{d}n^d$. The total number of edges is $\frac{1}{2}(n-1)dn^d$. Hence, a uniform random edge is a violation with probability at least $1/(2\sqrt{d})$. The path tester selects $\tau = 1$ with probability at least $1/\log d$, so the rejection probability is at least $1/(2\sqrt{d}\log d)$.

Case 2, $\sigma < C_{lar}/\sqrt{d}$: By Claim 7.9, $I_f^- = \Omega(\varepsilon^2\sqrt{d}/\log^4(nd))$. Thus the probability that a uniform random edge is a violation is $\Omega(\varepsilon^2/(n\sqrt{d}\log^4(nd)))$. Similar to the above case, the rejection probability is $\Omega(\varepsilon^2/(n\sqrt{d}\log^5(nd)))$.

Case 3, $I_f \leq 9(n-1)\sqrt{d}$ and $\sigma \geq C_{lar}/\sqrt{d}$: This is the interesting case, where all the previous claims and lemmas are used. Since $\sigma \geq C_{lar}/\sqrt{d}$, we can define τ_σ using Definition 7.10. By Lemma 7.7, since $I_f \leq 9(n-1)\sqrt{d}$, the fraction of vertices that are not τ_σ-persistent is at most

$$C_{per}\tau_\sigma/\sqrt{d} \leq (2C_{per}/C_{lar}) \cdot \sigma < \sigma/4.$$

Let us now count the number of viable edges in the good subgraph promised by Theorem 7.8. There are at least $\sigma k n^d/2$ edges in the good subgraph. There are at most $(\sigma/4) \cdot n^d$ non-persistent vertices in the good subgraph, each of which has degree at most k. Thus, removing all non-persistent vertices leaves us with

$$\sigma k n^d/2 - (\sigma/4) \times n^d \times k = \sigma k n^d/4$$

edges, all of which are viable. Let us now lower bound the probability of the tester rejecting using Lemma 7.12. Recall that all the events $\mathcal{C}_e$ are disjoint and $\tau_\sigma = \Theta(\sigma\sqrt{d})$.

$$\begin{aligned}
\Pr\left[\bigcup_{e \text{ viable}} \mathcal{C}_e | \tau = \tau_\sigma\right] &= \sum_{e \text{ viable}} \Pr[\mathcal{C}_e | \tau = \tau_\sigma] \\
&\geq (\sigma k/4) n^d \times n^{-d} \times n^{-1} \times (\tau_\sigma/d) \\
&= \Omega(\sigma^2 k/(n\sqrt{d})) = \Omega(\varepsilon^2/(n\sqrt{d}\log^4(nd))).
\end{aligned}$$

The probability of setting τ to be τ_σ is $1/\log d$, so we multiply the bound above by $1/\log d$ to complete the proof. □

8 Towards a $\widetilde{O}(\sqrt{d})$ tester

In this section we describe a different notion of influence of Boolean functions on hypergrids. We conjecture a Talagrand style isoperimetric theorem for the colored version of this influence is true. If so, then we can design a tester whose query complexity has no polynomial dependence on n. More precisely, the dependence on n is only polylogarithmic, and since by results of [BCS20, HY22] one can assume $n = \mathrm{poly}(d)$, the final tester's query complexity is $\widetilde{O}(\sqrt{d})$.

Definition 8.1 (Weighted Influence)**.** *Fix $f : [n]^d \to \{0,1\}$ and a dimension $i \in [d]$. Fix a point $\mathbf{x} \in [n]^d$. The* weighted *influence of $\mathbf{x}$ along coordinate i is defined as*

$$\Psi_f(\mathbf{x}; i) := \sum_{\mathbf{y}=\mathbf{x}\pm a\mathbf{e}_i \,:\, \textit{such that } (\mathbf{x},\mathbf{y}) \textit{ is violating}} \frac{1}{a}$$

The weighted influence of $\mathbf{x}$ is $\Psi_f(\mathbf{x}) = \sum_{i=1}^{d} \Psi_f(\mathbf{x}; i)$.

Consider giving a *weight* to every edge $(\mathbf{x}, \mathbf{x} \pm a\mathbf{e}_i)$ equal to $1/a$, the reciprocal of the length of the edge. The weighted influence of a point $\mathbf{x}$ with $f(\mathbf{x}) = 1$ is the sum of the weights of *out*-edges which are violating. This is another generalization of the notion of influence in hypercubes. Also note that for any $\mathbf{x}$, the thresholded influence of any point can't be much smaller than the weighted influence; indeed, $\Phi_f(\mathbf{x}) \geq \frac{1}{H_n}\Psi_f(\mathbf{x})$ where H_n is the nth Harmonic number and is $\Theta(\log n)$. Therefore, lower bounding Ψ's also lower bounds Φ's. On the other hand, $\Phi_f(\mathbf{x}; i)$ could be as large as $(n-1) \cdot \Psi_f(\mathbf{x}; i)$ for a particular $\mathbf{x}$.

The Talagrand objective with respect to this notion is defined as the expected *square root* of the weighted influence.

Definition 8.2 (Talagrand Objective wrt Weighted Influence)**.** *Given any Boolean function $f : [n]^d \to \{0,1\}$, we define the Talagrand objective with respect to the weighted influence as*

$$T_\Psi(f) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Psi_f(\mathbf{x})}\right]$$

where Ψ_f is as defined in Definition 8.1.

We also have a notion of colorful weighted influence and the corresponding Talagrand objective.

Definition 8.3 (Colorful Weighted Influence)**.** *Fix $f : [n]^d \to \{0,1\}$ and $\chi : E \to \{0,1\}$. Fix a dimension $i \in [d]$. Fix a point $\mathbf{x} \in [n]^d$. The colorful weighted influence of $\mathbf{x}$ along coordinate i is defined as*

$$\Psi_{f,\chi}(\mathbf{x}; i) := \sum_{\mathbf{y}=\mathbf{x}\pm a\mathbf{e}_i \,:\, \textit{such that } (\mathbf{x},\mathbf{y}) \textit{ is violating and } \chi(\mathbf{x},\mathbf{y})=f(\mathbf{x})} \frac{1}{a}$$

The colorful weighted influence of $\mathbf{x}$ is $\Psi_{f,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Psi_{f,\chi}(\mathbf{x}; i)$.

Definition 8.4 (Colorful Thresholded Talagrand Objective)**.** *Given any Boolean function $f : [n]^d \to \{0,1\}$ and $\chi : E \to \{0,1\}$, we define the Talagrand objective with respect to the colorful weighted influence as*

$$T_{\Psi_\chi}(f) := \mathbf{E}_{\mathbf{x}}\left[\sqrt{\Psi_{f,\chi}(\mathbf{x})}\right]$$

where $\Psi_{f,\chi}$ is as defined in Definition 8.3.

We are now ready to state our main conjecture. Note that due to the fact that $\Psi_{f,\chi}(\mathbf{x}) = O(\log n)\Phi_{f,\chi}(\mathbf{x})$, the conjecture below generalizes Theorem 1.4 up to $\mathrm{poly}\log n$ factors.

Conjecture 8.5. *Let $f : [n]^d \to \{0,1\}$ which is ε-far from monotone, and let $\chi : E \to \{0,1\}$ be an arbitrary coloring of the edges of the augmented hypergrid. Then there exists a constant $c \geq 0$ such that*

$$T_{\Psi_\chi}(f) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Psi_{f,\chi}(\mathbf{x})}\right] = \Omega\left(\frac{\varepsilon}{\log^c n}\right)$$

where $\Psi_{f,\chi}(\mathbf{x})$ is as defined in Definition 8.3.

If the above conjecture is true, then we can design a $\widetilde{O}(\sqrt{d})$ tester. Before we state this tester, and indeed to motivate it, we first note that the analysis of the tester described in Figure 6 is tight on its dependence on n. An example is the so-called *centrist* function defined in [BCS20] (Section 8) defined as follows for the case $n = d$, that is, $f : [d]^d \to \{0,1\}$.

$$f(\mathbf{x}) = \begin{cases} 0 & \text{if } \exists i \in [d] \quad \mathbf{x}_i = 2 \\ 1 & \text{otherwise} \end{cases}$$

[BCS20] (Claim 8.2) showed that $\varepsilon_f = \Omega(1)$ for the above function. To see this, observe that the probability of a random point having $\mathbf{x}_i \neq 2$ for all $i \in [d]$ is $\left(1 - \frac{1}{d}\right)^d = \Theta(1)$. Now let's consider the rejection probability of the algorithm in Figure 6. Note that a violating pair $(\mathbf{x}, \mathbf{y})$ must satisfy $\mathbf{x}_i = 1$ and $\mathbf{y}_i = 2$ for some i, and that i needs to be picked in the random set R. Even conditioning on picking that $i \in R$, the probability c_i would be set to 2 is $\frac{1}{n-1} \approx \frac{1}{d}$ in this case. Therefore, the algorithm needs to be modified.

To fix this, think about the tester as performing a (lazy) directed random walk. And now when one picks a dimension i to move in, one doesn't have a *uniform* distribution over the length it moves in this dimension, as currently is the case in Figure 6, rather one chooses a length using a discrete Pareto distribution. That is, smaller lengths are given more weight than longer lengths. More precisely, we pick a length a to move with probability proportional to $\frac{1}{a}$; note that the constant of proportionality that we need to scale down by is at most $H_n = \Theta(\log n)$. This is the reason we defined the weighted influence as we did in Definition 8.1, and indeed, if Conjecture 8.5 is true, then we can prove that this tester is an $\widetilde{O}(\sqrt{d})$ tester; the proof technique is similar to that described in Section 7 which itself is a modification of the analysis in [KMS15].

We now show that the *uncolored* version of Conjecture 8.5 is in fact true, and follows easily using the colorful version Theorem 1.4 for thresholded influence.

Theorem 8.6. *Let $f : [n]^d \to \{0,1\}$ which is ε-far from monotone. Then,*

$$T_{\Psi}(f) := \mathbf{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Psi_{f}(\mathbf{x})}\right] = \Omega\left(\frac{\varepsilon}{\log n}\right)$$

where $\Psi_f(\mathbf{x})$ is as defined in Definition 8.1.

Proof. The key is that we can always define a coloring χ for which $T_\Psi(f) = \Omega(T_{\Phi_\chi}(f))$ simply as follows. For every edge $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_i)$ of the fully augmented hypergrid, define $\chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_i) := 1$ if the interval $[\mathbf{x}, \mathbf{x} + a\mathbf{e}_i]$ is at least half 0's and $\chi(\mathbf{x}, \mathbf{x} + a\mathbf{e}_i) := 0$ otherwise. This coloring achieves the desired property because of the following simple fact.

Fact 8.7. *Let $S \subseteq \{1, \ldots, n\}$ be of size $|S| \geq \lceil n/2 \rceil$. Then $\sum_{a\in S} \frac{1}{a} = \Omega(1)$.*

Proof. For simplicity let n be even. The sum is minimized when $S = [n] \setminus [n/2]$ and in this case the sum is equal to $H_n - H_{n/2} = \Omega(1)$. □

Now consider $\mathbf{x}$ and $i \in [d]$ such that $f(\mathbf{x}) = 1$ and $\Phi_{f,\chi}(\mathbf{x}; i) = 1$. This implies that there exists an i-edge $(\mathbf{x}, \mathbf{y})$ for which $f(\mathbf{y}) = 0$ and $\chi(\mathbf{x}, \mathbf{y}) = 1$. By definition of the coloring, the interval $[\mathbf{x}, \mathbf{y}]$ is at least half 0's and so $\Psi_f(\mathbf{x}; i) = \Omega(1)$. Similarly, consider $\mathbf{y}$ and $i \in [d]$ such that $f(\mathbf{y}) = 0$ and $\Phi_{f,\chi}(\mathbf{y}; i) = 1$. This implies that there exists an i-edge $(\mathbf{x}, \mathbf{y})$ for which $f(\mathbf{x}) = 1$ and $\chi(\mathbf{x}, \mathbf{y}) = 0$. By definition of the coloring, the interval $[\mathbf{x}, \mathbf{y}]$ is at least half 1's and so $\Psi_f(\mathbf{y}; i) = \Omega(1)$. Thus, $T_\Psi(f) = \Omega(T_{\Phi_\chi}(f))$ and so invoking Theorem 1.4 completes the proof. □

It is worthwhile to point out the challenge in generalizing the above theorem to the colored version. Note the above was a "point-by-point" argument in that we found a coloring χ of the fully augmented hypergrid edges such that for every $\mathbf{x}$ and every $i \in [d]$, we could prove $\Psi_f(\mathbf{x}; i) = \Omega(\Phi_{f,\chi}(\mathbf{x}; i))$. One would wonder if such a point-by-point analysis is possible even when we have an arbitrary coloring χ', and the LHS in the previous statement is replaced with the colored version. Unfortunately, this is not possible. One can find examples of f and χ' such that no matter how you define χ, there will be some point $\mathbf{x}$ and some dimension i such that $\Phi_{f,\chi}(\mathbf{x}; i) = 1$, but $\Psi_{f,\chi'}(\mathbf{x}; i) \approx 1/n$. These examples do not disprove the conjecture since, in these examples, for a constant fraction of $(\mathbf{x}, i)$ pairs, we do have $\Psi_{f,\chi'}(\mathbf{x}; i) = \Omega(\Phi_{f,\chi}(\mathbf{x}))$, but it does point to the need of a new argument. One could also wonder if the proof technique used to prove Theorem 1.4 can bear upon the proof of the conjecture. There are many roadblocks here, one of them primarily being that semisorting can *increase* the Talagrand objective with respect to the Ψ-influences mainly because it can bring violations closer which bumps up the Ψ-influence. Nevertheless, the authors believe that Conjecture 8.5 is true, and we leave the resolution of this as a promising direction towards getting rid of the polynomial dependence on n thereby resolving the Boolean monotonicity testing question on hypergrids.

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. 11

[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. 11

[AV15] Amirali Abdullah and Suresh Venkatasubramanian. A directed isoperimetric inequality with application to bregman near neighbor lower bounds. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2015.

[BB16] Aleksandrs Belovs and Eric Blais. A polynomial lower bound for testing monotonicity. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2016. 1, 11

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012. 11

[BCS18] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. A $o(d) \cdot \mathrm{polylog}(n)$ monotonicity tester for Boolean functions over the hypergrid $[n]^d$. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2018. 1, 2, 4, 11

[BCS20] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Domain reduction: A $o(d)$ tester for boolean functions in d-dimensions. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2020. 1, 4, 11, 50, 51

[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. 11

[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. 11

[Bha08] Arnab Bhattacharyya. A note on the distance to monotonicity of boolean functions. Technical Report 012, Electronic Colloquium on Computational Complexity (ECCC), 2008. 11

[BKR20] Hadley Black, Iden Kalemaj, and Sofya Raskhodnikova. Isoperimetric inequalities for real-valued functions with applications to monotonicity testing. *arXiv*, abs/2011.09441, 2020. 1, 11, 12

[BRY14a] Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. L_p-testing. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2014. 1, 11

[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. 11

[CCK+21] Clément L. Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 321–336, 2021. 4

[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. 1, 11

[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. 1, 11

[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. 1, 11, 12

[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. 1, 2, 11

[CS14b] Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory of Computing*, 10:453–464, 2014. 12

[CS19] Deeparnab Chakrabarty and C. Seshadhri. Adaptive boolean monotonicity testing in total influence time. In *Proceedings, Innovations in Theoretical Computer Science (ITCS)*, pages 20:1–20:7, 2019. 11

[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. 1, 11

[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. 1, 11

[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. 1, 2, 5, 6, 11, 12

[EKK+00] Funda Ergun, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. *J. Comput. System Sci.*, 60(3):717–751, 2000. 11, 25

[Fis04] Eldar Fischer. On the strength of comparisons in property testing. *Information and Computation*, 189(1):107–116, 2004. 11

[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. 11

[FR10] Shahar Fattal and Dana Ron. Approximating the distance to monotonicity in high dimensions. *ACM Trans. on Algorithms (TALG)*, 6(3), 2010. 11, 13

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samordinsky. Testing monotonicity. *Combinatorica*, 20:301–337, 2000. 1, 2, 4, 11

[HK03] Shirley Halevy and Eyal Kushilevitz. Distribution-free property testing. *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2003. 11

[HK08] Shirley Halevy and Eyal Kushilevitz. Testing monotonicity over graph products. *Random Structures Algorithms*, 33(1):44–67, 2008. 11

[HY22] Nathaniel Harms and Yuichi Yoshida. Downsampling for testing and learning in product distributions. In *Proceedings, International Colloquium on Automata, Languages and Programming (ICALP)*, volume 229, pages 71:1–71:19, 2022. 1, 4, 11, 50

[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. 1, 2, 3, 4, 5, 6, 8, 10, 11, 13, 14, 16, 22, 44, 45, 46, 47, 51

[LR01] Eric Lehman and Dana Ron. On disjoint chains of subsets. *Journal of Combinatorial Theory, Series A*, 94(2):399–404, 2001. 11

[Mar74] Grigory A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy Peredachi Informatsii*, 10(2):101–108, 1974. 2

[MOA11] Albert W Marshall, Ingram Olkin, and Barry C Arnold. *Inequalities: theory of majorization and its applications*, volume 143. Springer, 2nd edition, 2011. 15

[PRV18] Ramesh Krishnan S. Pallavoor, Sofya Raskhodnikova, and Nithin Varma. Parameterized property testing of functions. *ACM Trans. Comput. Theory*, 9(4):17:1–17:19, 2018. 12

[PRW22] Ramesh Krishnan S Pallavoor, Sofya Raskhodnikova, and Erik Waingarten. Approximating the distance to monotonicity of boolean functions. *Random Structures Algorithms*, 60(2):233–260, 2022. Prelim. version. in SODA 2020. 2, 5, 10, 11, 16, 23, 24

[Ras99] Sofya Raskhodnikova. Monotonicity testing. *Masters Thesis, MIT*, 1999. 11

[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. 11

[SS08] Michael E. Saks and C. Seshadhri. Parallel monotonicity reconstruction. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008. 11

[Tal93] Michel Talagrand. Isoperimetry, logarithmic sobolev inequalities on the discrete cube, and margulis graph connectivity theorem. *Geom. Func. Anal.*, 3(3):295–314, 1993. 2, 14