Title

Testing and Learning in High-Dimensions: Monotonicity Testing, Directed Isoperimetry, and Convex Sets

Author
Black, Hadley

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Testing and Learning in High-Dimensions: Monotonicity Testing, Directed Isoperimetry, and Convex Sets

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Computer Science

by

Hadley Black

2023

ABSTRACT OF THE DISSERTATION

Testing and Learning in High-Dimensions:
Monotonicity Testing, Directed Isoperimetry, and Convex Sets

by

Hadley Black
Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2023
Professor Raghu Meka, Chair

This thesis studies testing and learning of monotone functions, k-monotone functions, and convex sets over high-dimensional domains. Our primary focus is *monotonicity testing*, which has been one of the central problems in the field of property testing since its beginnings in the late 90's. Monotonicity testing has generated a lot of interest, partially due to its connection to isoperimetric inequalities, which are a fundamental tool in Boolean function analysis. Our secondary focus is on *testing and learning convex sets*. Our contributions are presented in four parts summarized as follows:

1. We present a nearly optimal non-adaptive monotonicity testing algorithm for Boolean functions over d-dimensional hypergrids and continuous product spaces. Among other technical contributions, a central tool in our proof is a new isoperimetric inequality for Boolean functions over hypergrids.

2. Given the impact of isoperimetric inequalities for testing monotonicity of Boolean functions, a natural question is whether these inequalities generalize to larger ranges. We give a black-box reduction showing that the known inequalities in this area generalize

to real-valued functions. We use this result to obtain nearly optimal bounds for (non-adaptive, one-sided error) monotonicity testing parameterized by the range size and an improved upper bound for approximating the distance to monotonicity of real-valued functions on the hypercube.

3. We present nearly matching upper and lower bounds for sample-based testing and learning of k-monotone functions over hypercubes and continuous product spaces.

4. Motivated by the limited understanding of convexity testing in high-dimensions we initiate the study of convex sets over the ternary hypercube, $\{-1, 0, 1\}^d$, which is the simplest high-dimensional domain where convexity is a non-trivial property. We obtain (i) nearly tight bounds on the edge-boundary of convex sets in this domain, (ii) new upper and lower bounds for sample-based testing and learning, and (iii) nearly matching upper and lower bounds for non-adaptive testing with one-sided error.

The dissertation of Hadley Black is approved.

Alexander Sherstov

Amit Sahai

Rafael Ostrovsky

Raghu Meka, Committee Chair

University of California, Los Angeles

2023

This thesis is dedicated to my parents.

To my mom, who taught me how to find purpose and joy in hard work.

To my pop, who taught me how to love life and ask big questions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am enormously grateful to my advisor, Raghu Meka, for his invaluable mentorship, kindness, patience, and encouragement throughout my PhD. Raghu has always given me a great deal of freedom to explore and wander, always having my best interests at heart. I'm inspired by his fearlessness to work on problems that most people find too daunting. Beyond his great talent, I believe he has an incredibly healthy attitude towards research and this has had a really positive influence on me.

I also wish to thank C. Seshadhri who advised me during my master's degree at UC Santa Cruz. While I was an undergraduate student Sesh gave me the opportunity to try out research, introducing me to the monotonicity testing problem, which eventually led to the work making up a significant portion of this thesis. I am immensely grateful that he found working with me to be worth his time, even back when I was just getting started. Since then his mentorship has been invaluable in my career and I've learned a ton from his optimism and energetic research style.

Many thanks to all of my coauthors, C. Seshadhri, Deeparnab Chakrabarty, Sofya Raskhodnikova, Iden Kalemaj, Eric Blais, and Nathan Harms. In particular, I want to thank Sofya and Eric who hosted me during the summers of 2020 and 2021, respectively. Both visits were extremely enjoyable and productive.

I want to also thank my friends who enriched my time at UCLA, especially Ashutosh, Karl, Paul, Davina, Eli, Kobe, Alexis, and Riddhi. I also spent some months visiting MIT and I am extremely grateful to my dear friend Leilani who helped me feel welcome and happy there. Outside of UCLA, I want to thank my unofficial brother Arion (a real doctor) for countless good times, rock climbing, and for making sure I left the house growing up.

None of this would be possible without my amazing partner, Kayla, who has put up with me and encouraged me endlessly over these last five years. Her brilliance inspires me to do my best in work and in life. She is a constant inspiration and reminder of what's most

important to me.

Finally, I want to express my gratitude to my parents, Jeff and Cathryn, for always encouraging me to pursue and cultivate my passions, to be resilient, to love life, and to keep my priorities straight.

VITA

2016	B.A. (Computer Science) and B.A. (Pure Mathematics), UC Santa Cruz. Highest Honors in both majors. GPA: 3.91 - Magna Cum Laude
2018	M.S. (Computer Science), UC Santa Cruz. GPA: 3.90. Advised by C. Seshadhri.

PUBLICATIONS

Hadley Black, Eric Blais, Nathaniel Harms. Testing and Learning Convex Sets in the Ternary Hypercube. ITCS 2024.

Hadley Black, Deeparnab Chakrabarty, C. Seshadhri. A $d^{1/2+o(1)}$ Monotonicity Tester for Boolean Functions on d-Dimensional Hypergrids. FOCS 2023. Invited to SICOMP Special Issue.

Hadley Black, Deeparnab Chakrabarty, C. Seshadhri. Directed Isoperimetric Theorems for Boolean Functions on the Hypergrid and an $\widetilde{O}(n\sqrt{d})$ Monotonicity Tester. STOC 2023.

Hadley Black, Iden Kalemaj, Sofya Raskhodnikova. Isoperimetric Inequalities for Real-Valued Functions with Applications to Monotonicity Testing. ICALP 2023.

Hadley Black. Nearly Optimal Bounds for Sample-Based Testing and Learning of k-Monotone Functions. Manuscript.

Hadley Black, Deeparnab Chakrabarty, C. Seshadhri. Domain Reduction for Monotonicity Testing: A $o(d)$ Tester for Boolean Functions in d-Dimensions. SODA 2020.

Hadley Black, Deeparnab Chakrabarty, C. Seshadhri. A $o(d)\cdot\text{polylog}\, n$ Monotonicity Tester for Boolean Functions over the Hypergrid $[n]^d$. SODA 2018.

CHAPTER 1

Introduction

"The work is mysterious and important." - Mark Scout[1]

An emergent aspect of modern computer science is the massive growth in data which our algorithms must consider. This quantity of data demands that we design algorithms whose run-times are extremely fast compared to the size of their input. Often this means our algorithms must run in *sublinear time*, meaning that they're only allowed to examine a tiny fraction of their input. To meet this run-time constraint, we are generally forced to relax various other demands of our algorithm, such as the quality of the solution it produces, or the class of inputs it is required to perform well on.

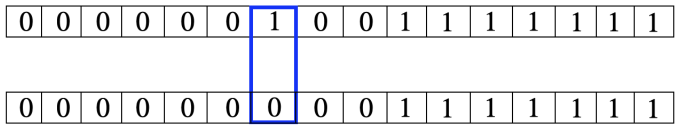
1.1 The Challenges of Massive Data

To illustrate the innate challenges of dealing with massive data, consider the following basic algorithmic question:

Is the given array of N bits sorted, or not?

An algorithm which answers the above question correctly on all inputs must answer differently on the following two arrays, which differ on *a single bit*.

[1] From *Severance* on Apple TV [SWC$^+$22].

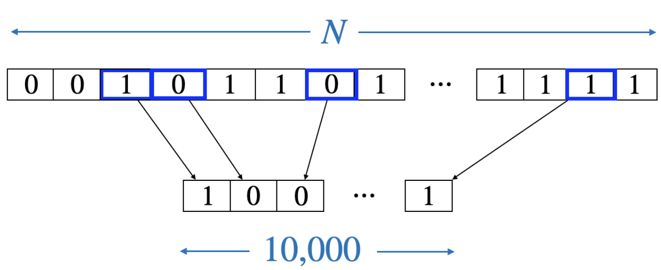


Thus, answering the above question requires the algorithm to *look at every bit* of its input and hence perform at least N operations. A conservative back-of-the-envelope calculation shows that on a modestly sized 1 terabyte dataset using a modern 3.5 GHz CPU, this would take on the order of roughly 5-10 minutes, which is far too long for many applications.

The inherent difficulty of the problem comes from the fact that we are asking our algorithm to answer correctly on *every input* and with 100% probability. Suppose instead we ask the following relaxed version of the question:

Is the given array of N bits sorted, or does it differ from every sorted N bit array on at least 1% if its entries?

The following algorithm succeeds in answering the above question with 99% probability. Check if the array is sorted on a sub-array of 10,000 randomly chosen locations. If so, output “yes”, and otherwise output “no”.



Note that now the number of operations required has no dependence on N (the size of the dataset), but instead depends on the desired probability of success (99%) and level of tolerance it has for unsorted inputs (1%). Performing 10,000 operations on the same modern CPU from the previous scenario would take less than 1 millisecond.

1.2 Property Testing and Computational Learning Theory

Property testing, originally credited to [RS96, GGR98], is a subfield of theoretical computer science (TCS) which aims to study the types of relaxed algorithmic decision problems posed in Section 1.1. In general, decision problems pose a yes/no question about a problem instance, e.g. "is this array sorted?", or "does this graph contain a clique of size k"? Property testing poses the following type of *relaxed* decision problem: "does this large object (input) have the property $\mathcal{P}$, or is it *far* from having the property $\mathcal{P}$?", where the notion of "far" is prescribed by a distance metric and a proximity parameter. Some examples of commonly studied objects are functions, graphs, and probability distributions. The property $\mathcal{P}$ is formalized as some specific subset of the objects; for example *monotone* functions, *bipartite* graphs, or *uniform* distributions. To answer such a question, a property testing algorithm is given the power to locally inspect the object at various locations, which it is either allowed to choose (query-based algorithms[2]) or which are given to it randomly according to some underlying distribution (sample-based algorithms[3]). The goal is to design an algorithm which makes as *few* inspections as possible.

Property testing is closely related to *computational learning theory*, in particular the *probably approximately correct* (PAC) learning model introduced by Valiant [Val84]. A PAC learning algorithm for a property $\mathcal{P}$ is given access to an input object $X \in \mathcal{P}$ and is required to produce a hypothesis H which is "close" to X with "high probability". Concretely, the connection between testing and learning is due to the fact that a PAC learning algorithm can always be transformed into a property testing algorithm for the same property which makes roughly the same number of inspections to its input. Therefore, learning is always at least as hard as testing.

Often times property testing and PAC learning algorithms are quite simple and intuitive. The real depth in these areas comes from asking "what are the structural properties of this

[2]This is the standard algorithmic model in property testing.

[3]This is the standard algorithmic model in learning theory.

class of objects that allow this intuitive algorithm to work efficiently?". This type of question has led to many interesting mathematical ideas in combinatorics and probability theory. A notable example is Fourier analysis of Boolean functions (see [O'D14]) which has seen a lot of attention recently due to being extremely useful for a broad class of learning problems. Another example which is quite relevant to this thesis is that of isoperimetric inequalities for Boolean functions, which have turned out to be essential for analyzing monotonicity testing algorithms.

1.3 Contributions and Organization

This thesis studies testing, learning, and structural aspects of monotone functions, k-monotone functions, and convex sets in high-dimensional domains. The thesis is divided into four parts. We give a very brief introduction for each part below. In Chapter 2 we give a survey on the monotonicity testing problem for Boolean functions over hypergrids. This serves as an introduction for Parts I and II.

Part I: Monotonicity Testing of Boolean Functions over Hypergrids: The primary topic studied in this thesis is *monotonicity testing* of Boolean functions over hypergrids, $f\colon [n]^d \to \{0,1\}$. This problem was first considered by [GGL⁺00] for the special case of $n = 2$ (the *hypercube*) and for general $n \geq 2$ (the *hypergrid*) by [DGL⁺99]. After a long line of work, [KMS18] gave a non-adaptive upper bound of $\widetilde{O}(\sqrt{d})$ for the hypercube setting, and this is known to be optimal for non-adaptive testers [CWX17]. This breakthrough was enabled by the development of new isoperimetric inequalities for Boolean functions in the directed hypercube. At the time of this result, the best upper bound for hypergrids (even $n = 3$) was $\widetilde{O}(d)$. In this thesis, we develop many new techniques for monotonicity testing over hypergrids, culminating in a $d^{1/2+o(1)}$ query non-adaptive tester for all n [BCS23a], thus resolving this question up to a $o(1)$ factor in the exponent. This result also holds for Boolean functions $f\colon \mathbb{R}^d \to \{0,1\}$ under any product measure over $\mathbb{R}^d$. The material presented appeared in the following works:

- [BCS20]: "Domain Reduction for Monotonicity Testing: A $o(d)$ Tester for Boolean Functions in d-Dimensions", with Deeparnab Chakrabarty and C. Seshadhri, appeared in SODA 2020.

- [BCS23b]: "Directed Isoperimetric Theorems for Boolean Functions on the Hypergrid and an $\widetilde{O}(n\sqrt{d})$ Monotonicity Tester", with Deeparnab Chakrabarty and C. Seshadhri, appeared in STOC 2023.

- [BCS23a] "A $d^{1/2+o(1)}$ Monotonicity Tester for Boolean Functions on d-Dimensional Hypergrids", with Deeparnab Chakrabarty and C. Seshadhri, appeared in FOCS 2023, and was invited to the SICOMP Special Issue.

Part II: Directed Isoperimetry and Monotonicity Testing of Real-Valued Functions: Isoperimetric inequalities over the directed hypercube and hypergrid have been a central tool for analyzing monotonicity testers for Boolean functions. Given the impact of these inequalities, it is natural to ask whether they extend to broader classes of functions. We give a black-box result showing essentially that any such inequality for Boolean-valued functions $f\colon D \to \{0,1\}$ over a partial order D in fact holds for all real-valued functions $f\colon D \to \mathbb{R}$. In particular, we generalize all known directed isoperimetric inequalities for $f\colon \{0,1\}^d \to \{0,1\}$ and $f\colon [n]^d \to \{0,1\}$ to $f\colon \{0,1\}^d \to \mathbb{R}$ and $f\colon [n]^d \to \mathbb{R}$. We apply these new inequalities to obtain new results for the tasks of monotonicity testing and approximating the distance to monotonicity of a function. The material presented appeared in the following work:

- [BKR23]: "Isoperimetric Inequalities for Real-Valued Functions with Applications to Monotonicity Testing", with Iden Kalemaj and Sofya Raskhodnikova, appeared in ICALP 2023.

Part III: Sample-Based Testing and Learning of k-Monotone Functions: The monotonicity testing literature has primarily focused on query-based algorithms, which are allowed to request the value of the function on points of their choosing. Another natural algorithmic model is that of *sample-based* algorithms, which are only allowed to see the function's value on points independently sampled from the underlying distribution. One of the original works on monotonicity testing of [GGL+00] initially studied this question, but their lower bound only covered a very specific parameter regime. We close this gap in the literature by giving nearly matching upper and lower bounds for this problem in all parameter regimes. In fact our results hold more generally for sample-based testing and learning of k-monotone functions with image size r, $f: \{0,1\}^d \to [r]$. We also provide nearly matching upper and lower bounds for functions $f: \mathbb{R}^d \to [r]$ when the underlying distribution is any product measure. The material presented appeared in the following work:

- [Bla23] "Nearly Optimal Bounds for Sample-Based Testing and Learning of k-Monotone Functions", unpublished manuscript.

Part IV: Testing and Learning Convex Sets in the Ternary Hypercube: Convex sets are extremely natural, appearing in a wide array of contexts. Somewhat surprisingly, many open questions still remain regarding the complexity of testing set convexity, in particular in high-dimensional domains. The "simplest" such domain where convexity is a non-trivial property is the *ternary hypercube*, $\{-1,0,1\}^d$. We study the testing and learning problems for convex sets in this domain. The material presented appeared in the following work:

- [BBH23]: "Testing and Learning Convex Sets in the Ternary Hypercube", with Eric Blais and Nathaniel Harms, appearing at ITCS 2024.

CHAPTER 2

Monotonicity Testing of Boolean Functions over Hypergrids: A Short Survey

Monotonicity testing is one of the most classic problems in property testing. Naturally, the most well-studied setting is that of testing monotonicity of Boolean functions $f\colon \{0,1\}^d \to \{0,1\}$ over the hypercube, first studied by [GGL$^+$00], and the generalized problem over the hypergrid, $[n]^d$, first studied by [DGL$^+$99]. Here $[n]$ denotes the set $\{1,\ldots,n\}$ and the partial order over the hypergrid is defined as: $x \preceq y$ iff $x_i \leq y_i$ for all $i \in [d]$. A function $f\colon [n]^d \to \{0,1\}$ is monotone if $f(x) \leq f(y)$ whenever $x \preceq y$.

The Hamming distance between two functions f and g, denoted $\Delta(f,g)$, is the fraction of points where they differ. We say a function $f\colon [n]^d \to \{0,1\}$ is ε-far from monotone if $\Delta(f,g) \geq \varepsilon$ for all monotone functions $g\colon [n]^d \to \{0,1\}$. Given a proximity parameter ε and query access to a function f, a *monotonicity tester* must satisfy the following criteria: (a) if f is monotone, then the tester accepts with probability at least $2/3$, and (b) if f is ε-far from monotone, then the tester rejects with probability at least $2/3$. If the tester accepts monotone functions with probability 1, it is called *one-sided*. If the tester decides its queries without seeing any responses, it is called *non-adaptive*.

Monotonicity testing of Boolean functions over the hypercube has attracted a great deal of attention and the query complexity of this problem is still an outstanding open question, with the current gap standing at $\widetilde{\Omega}(d^{1/3})$ vs $\widetilde{O}(\sqrt{d})$ [CWX17, KMS18]. The tester attaining this upper bound is *non-adaptive* and in fact is optimal among non-adaptive testers due to a matching lower bound of $\widetilde{\Omega}(\sqrt{d})$ also due to [CWX17]. The road to resolving the non-adaptive query complexity of this problem led to the development of novel techniques, notably the notion of *directed isoperimetric inequalities* for Boolean functions.

After the resolution of the non-adaptive question over the hypercube, the more general problem over the hypergrid, $[n]^d$, remained open, even for $n = 3$. Recently, this question was (nearly) resolved by [BCS23a], who gave an upper bound of $d^{1/2+o(1)}$ for all n, matching the non-adaptive lower bound of $\widetilde{\Omega}(\sqrt{d})$ [CWX17] up to a $d^{o(1)}$ factor. This upper bound was due to the culmination of many techniques by the authors [BCS18, BCS20, BCS23b, BCS23a] which also built upon a long line of results studying the hypercube case [GGL+00, CS14a, CST14, KMS18] and the hypergid [Ras99, DGL+99, BRY14a, BKKM23]. The purpose of this chapter is to give an overview and history of the ideas that eventually led to the $d^{1/2+o(1)}$ non-adaptive tester for hypergrids, highlighting the connection with isoperimetric inequalities. This serves as an introduction for the technical contents of parts I and II.

Domain	Co-domain	Lower Bound	Upper Bound
$\{0,1\}^d$	$\{0,1\}$	$\widetilde{\Omega}(d^{1/3})$ a. 2-s [CWX17]	$\widetilde{O}(\sqrt{d})$ n.a. 1-s [KMS18]
		$\widetilde{\Omega}(\sqrt{d})$ n.a. 2-s [CWX17]	$O(I_f)$ a. 1-s [CS19]
		$\Omega(\sqrt{d})$ n.a. 1-s [FLN+02]	
$\{0,1\}^d$	$[r]$	$\Omega(\min\{d, r^2\})$ [BBM12] a. 2-s	$\widetilde{O}(r\sqrt{d})$ n.a. 1-s [BKR23]
		$\Omega(r\sqrt{d})$ n.a. 1-s [BKR23]	
$\{0,1\}^d$	$\mathbb{R}$	$\Omega(d)$ a. 2-s [BBM12, CS14b]	$O(d)$ n.a. 1-s [CS13]
$[n]^d$	$\{0,1\}$	-	$d^{1/2+o(1)}$ n.a. 1-s [BCS23a]
$[n]^d$	$[r]$	-	-
$[n]^d$	$\mathbb{R}$	$\Omega(d \log n)$ a. 2-s [CS14b]	$O(d \log n)$ n.a. 1-s [CS13]

Table 2.1: *Current state of the art for monotonicity testing over the d-dimensional hypercube and hypergrid. All results are stated for constant ε. We use a./n.a. to denote adaptive/non-adaptive and 1-s/2-s to denote one-sided/two-sided error.*

Organization: In Section 2.1 we give a history of directed isoperimetric inequalities and upper bounds for Boolean monotonicity testing over the hypercube. The remainder of the sections pertain to the techniques developed for hypergrids: Section 2.2 discusses *domain reduction* [BRY14a, BCS20, HY22] which is the topic of Chapter 3, Section 2.3 discusses generalized directed isoperimetric inequalities over hypergrids [DGL+99, BCS18, BKKM23, BCS23b] which is the topic of Chapter 4, and Section 2.4 discusses the techniques from [BCS23a] and the $d^{1/2+o(1)}$ tester which is the topic of Chapter 5. We conclude in Section 2.5 with a discussion of some open questions and final remarks.

2.1 Monotonicity Testing and Isoperimetry on the Hypercube

The *directed hypercube* is the DAG with vertex set $\{0,1\}^d$ and edge set

$$E = \{(x,y) \colon \exists i \in [d] \text{ such that } x_i < y_i \text{ and } x_j = y_j \text{ for all } j \in [d] \setminus \{i\}\}.$$

Given a function $f \colon \{0,1\}^d \to \{0,1\}$, we use $G_f(X_f, Y_f, E_f)$ to denote the *violation graph* of f which is the subgraph of the directed hypercube defined as follows: $X_f = \{x \colon f(x) = 1\}$, $Y_f = \{y \colon f(y) = 0\}$, and $E_f = \{(x,y) \in E \colon x \in X_f,\, y \in Y_f\}$. That is, G_f is the bipartite subgraph consisting of all edges where f violates monotonicity. Observe that if a tester manages to query points $x \in X_f$, $y \in Y_f$ for which $x \prec y$, then it has found a certificate of non-monotonicity of f and it may safely reject. All of the testers which we consider in this survey are *pair testers*, meaning that they work by independently sampling pairs of points (x,y) from a distribution $\mathcal{D}$ supported over $\{(x,y) \colon x \preceq y\}$, rejecting if they ever see a violation, and accepting otherwise. Note that if T is a pair tester using distribution $\mathcal{D}$ and $\mathbb{P}_{(x,y)\sim\mathcal{D}}[x \in X_f \wedge y \in Y_f] \geq p(\varepsilon)$ whenever f is ε-far from monotone, then T is a $O(p(\varepsilon)^{-1})$ query tester since the probability of not seeing a violation across this many trials is at most $(1 - p(\varepsilon))^{O(p(\varepsilon)^{-1})} \leq 1/3$. By definition, pair testers are non-adaptive and have one-sided error.

Note that f is the characteristic function for the set X_f and E_f is the set of edges on the *boundary* of X_f, or equivalently, of f. A central theme in the monotonicity testing literature has been to prove theorems that relate the structural properties of these boundary edges to the distance to monotonicity, ε_f, and to exploit this structure to analyze testers.

2.1.1 The Edge Tester and the Directed Poincaré Inequality

The starting point for these results is the very natural *edge tester* considered by [GGL+00]: sample a uniform random edge $(x, y) \in E$ and *reject* if $(x, y) \in E_f$. The total number of edges in the hypercube is $|E| = d \cdot 2^{d-1}$ and so the probability that this test finds a violation is precisely $|E_f|/(d \cdot 2^{d-1})$. Thus, the analysis of this tester amounts to proving a lower bound on $|E_f|$. Another way to look at this quantity is in terms of *influence*.

Definition 2.1.1 (Influence)**.** *The influence and negative influence of a function* $f : \{0,1\}^d \to \{0,1\}$ *are defined as*

$$I_f = 2^{-d} \cdot \{(x, y) \in E : f(x) \neq f(y)\} \text{ and } I_f^- = 2^{-d} \cdot \{(x, y) \in E : f(x) > f(y)\} = 2^{-d} \cdot |E_f| \tag{2.1}$$

The following bound on the negative influence was proven by [GGL+00].

Theorem 2.1.2 (Directed Poincaré, [GGL+00])**.** *Every* $f : \{0,1\}^d \to \{0,1\}$ *satisfies* $I_f^- = \Omega(\varepsilon_f)$.

This bound immediately implies that the edge tester is a $O(d/\varepsilon)$ query tester since it implies a random edge violates monotonicity with probability at least $2I_f^-/d = \Omega(\varepsilon/d)$ whenever f is ε-far from monotone.

Corollary 2.1.3 (Edge tester bound, [GGL+00])**.** *For every* $f : \{0,1\}^d \to \{0,1\}$*, the edge test finds a violation of monotonicity with probability* $\Omega(\varepsilon_f/d)$*. Therefore, the edge tester is a* $O(d/\varepsilon)$ *query tester.*

It was later observed by [CS14a] that Theorem 2.1.2 can be viewed as a *directed analogue* to the following standard edge isoperimetric bound for the undirected hypercube attributed to Poincaré.

Theorem 2.1.4 (Poincaré)**.** *Every* $f : \{0,1\}^d \to \{0,1\}$ *satisfies* $I_f = \Omega(\mathrm{var}(f))$.

Going from Theorem 2.1.4 to Theorem 2.1.2, the influence I_f is replaced by the *negative* influence I_f^- and $\mathrm{var}(f)$ is replaced by ε_f. The influence measures the quantity of edges in the *undirected* hypercube which leave the set $X_f = \{x\colon f(x) = 1\}$ and the *negative* influence measures the quantity of edges in the *directed* hypercube which leave the set X_f. In this sense Theorem 2.1.4 is an edge isoperimetric bound for the directed hypercube and hence we refer to it as the "directed Poincaré inequality".

Remark 2.1.5 (On ε_f as a directed analogue of $\mathrm{var}(f)$)**.** *A useful viewpoint to appreciate the analogy between* ε_f *and* $\mathrm{var}(f)$ *is the following. First, we note that* $\mathrm{var}(f)$ *is a* 2*-approximation of the distance of* f *to the nearest constant function, which we'll denote by* $d(f, \mathsf{const})$*. This is because* $\mathrm{var}(f) = \mathbb{E}[f](1 - \mathbb{E}[f])$*, while* $d(f, \mathsf{const}) = \min(\mathbb{E}[f], 1 - \mathbb{E}[f])$*. Thus, Theorem 2.1.4 can be written as* $I_f = \Omega(d(f, \mathsf{const}))$*, which* compares the number of non-constant edges to the global distance of f to being constant*. On the other hand, Theorem 2.1.2 can be written as* $I_f^- = \Omega(d(f, \mathsf{monotone}))$*, which compares* the number of non-monotone edges to the global distance of f to being non-monotone*. Non-constant edges are those that "leave the set of* 1*'s",* $f^{-1}(1)$*, in the* undirected hypercube *and non-monotone edges are those that "leave the set of* 1*'s",* $f^{-1}(1)$*, in the* directed hypercube.

2.1.2 The Path Tester and the Directed Margulis Inequality

It is straightforward to see that the directed Poincaré inequality of [GGL+00] is tight by considering an anti-dictator function $\mathsf{AD}_1(x) = 1 - x_1$ and observing that $I^-_{\mathsf{AD}_1} = \varepsilon_{\mathsf{AD}_1} = 1/2$. Therefore, the edge tester requires $\Omega(d)$ queries. So, how can we achieve a $o(d)$ query tester? Consider the following more general type of test.

Definition 2.1.6 (τ-Length Path Test)**.** *Given* $\tau \in \mathbb{N}$*, the* τ-length path test *samples* $x \in \{0,1\}^d$ *uniformly at random, chooses a set* $T \subseteq [d]$ *of* τ *i.i.d. uniform random coordinates, and then obtains* $y \succeq x$ *by setting* $y_i = 1$ *for all* $i \in T$ *and* $y_j = x_j$ *for all* $j \in [d] \setminus T$*.*

In other words, y is the upper endpoint of a τ-length random walk from x in the directed hypercube. The edge tester corresponds to the case of $\tau = 1$ and this test catches a violation for AD_1 with probability $O(1/d)$. Suppose instead we set $\tau = \sqrt{d}$. In this case the τ-length path test catches a violation as long as (i) $x_1 = 0$ and (ii) $T \ni 1$, and the probability of these events both occurring is $\Omega(1/\sqrt{d})$, a vast improvement over the edge test[1]. At an intuitive level, the reason why this test works well for the family of anti-dicators is that, while these functions have *few boundary edges*, they have *many boundary vertices* (in fact all vertices are on the boundary). Thus, informally, at each step of the walk there is an $\Omega(1/d)$ probability of crossing the boundary and the final success probability is the sum of these probabilities.

Inspired by the above reasoning, one may ask if there is a directed isoperimetric inequality which says "when there are few boundary edges, there are many boundary vertices". Such an inequality was proven by [CS14a] which led to the first $o(d)$ query tester. To state their inequality we need to define the notion of *directed vertex boundary*, defined as $\Gamma_f^- = 2^{-d}\,|\{x\colon \exists (x,y) \in E_f\}|$. I.e. Γ_f^- is the fraction of points which are the lower endpoint of some edge violation.

Theorem 2.1.7 (Directed Margulis, [CS14a])**.** *Every* $f\colon \{0,1\}^d \to \{0,1\}$ *satisfies*

$$I_f^- \cdot \Gamma_f^- = \Omega(\varepsilon_f^2).$$

This inequality gives a tradeoff between the edge and vertex boundaries. In fact, [CS14a] prove a stronger inequality implying that either $G_f(X_f, Y_f, E_f)$ (i) contains many edges, or (ii) contains a large matching. Formally, let $\Gamma^-_{f,\mathsf{matching}}$ denote the size of the largest matching in $G_f(X_f, Y_f, E_f)$ divided by 2^d.

[1]The reader may wonder why not set $\tau \gg \sqrt{d}$. The reason we are limited to $\tau = O(\sqrt{d})$ is as follows. Consider a *truncated* anti-dictator function: $\mathsf{TAD}_1(x) = \mathsf{AD}(x)$ if $|x| \in [\frac{d-\sqrt{d}}{2}, \frac{d+\sqrt{d}}{2}]$, $\mathsf{TAD}_1(x) = 0$ if $|x| < \frac{d-\sqrt{d}}{2}$ and $\mathsf{TAD}_1(x) = 1$ if $|x| > \frac{d+\sqrt{d}}{2}$. By standard concentration bounds, we still have $I^-_{\mathsf{TAD}_1} = O(1)$ and $\varepsilon_{\mathsf{TAD}_1} = \Omega(1)$, but now if $\tau > \sqrt{d}$, then one of the endpoints of the walk will be outside of the middle layers and we will not catch a violation. In fact, truncated anti-dictator functions were used by [FLN+02] to prove an $\Omega(\sqrt{d})$ lower bound for non-adaptive, one-sided testers.

Theorem 2.1.8 ([CS14a]). *Every $f\colon \{0,1\}^d \to \{0,1\}$ satisfies*

$$I_f^- \cdot \Gamma^-_{f,\mathsf{matching}} = \Omega(\varepsilon_f^2).$$

Theorem 2.1.8 enables us to achieve $o(d)$ query complexity using the τ-length path test for $\tau \gg 1$. Originally, [CS14a] gave a $\widetilde{O}(d^{7/8}\varepsilon^{-3/2})$ tester and [CST14] later gave a $\widetilde{O}(d^{5/6}\varepsilon^{-4})$ tester. Later [KMS18] formalized the path tester as follows.

Definition 2.1.9 (Path test). *Choose $p \in [\log d]$ uniformly at random and set $\tau = 2^p$. Run the τ-length path test*[2]*.*

Using Theorem 2.1.8 one can show the following bound for the path tester, as defined in Definition 2.1.9.

Theorem 2.1.10 (Path tester bound from Margulis). *For every $f\colon \{0,1\}^d \to \{0,1\}$, the path test finds a violation of monotonicity with probability $\widetilde{\Omega}(\varepsilon_f^{4/3} \cdot d^{-5/6})$. Therefore, the path tester is a $\widetilde{O}(d^{5/6}\varepsilon^{-4/3})$ query tester.*

It turns out that Theorem 2.1.7 is a directed version of the following isoperimetric inequality due to Margulis [Mar74], and hence we refer to it as the "directed Margulis inequality". In the following theorem Γ_f denotes the vertex boundary of f in the undirected hypercube.

Theorem 2.1.11 (Margulis, [Mar74]). *Every $f\colon \{0,1\}^d \to \{0,1\}$ satisfies*

$$I_f \cdot \Gamma_f = \Omega(\mathrm{var}(f)^2).$$

2.1.3 The Directed Talagrand Inequality

The directed Margulis inequality Theorem 2.1.7 tells us, informally, that "either the edge boundary is large, or the vertex boundary is large". One may ask if there is an inequality

[2] To be completely formal, the path test actually performs a random walk going up or down, each with probability $1/2$. We ignore this detail for clarity.

which interpolates smoothly between these two cases and if this could lead to a better analysis of the path tester. Both of these questions were answered in the affirmative by [KMS18] (referred to henceforth as "KMS") who proved the following. For a point $x \in X_f$, the negative influence at x, denoted $I_f^-(x)$, is the number edges incident to x that violate monotonicity. If $x \in Y_f$, then $I_f^-(x) = 0$.

Theorem 2.1.12 (Directed Talagrand, [KMS18])**.** *Every* $f\colon \{0,1\}^d \to \{0,1\}$ *satisfies*[3]

$$\mathbb{E}_x\left[\sqrt{I_f^-(x)}\right] = \Omega(\varepsilon_f).$$

Again, Theorem 2.1.12 is a directed analogue of the following isoperimetric inequality in the undirected hypercube due to Talagrand [Tal93]. For a point $x \in X_f$, the total influence at x, denoted $I_f(x)$, is the number of sensitive edges in the undirected hypercube incident to x. If $x \in Y_f$, then $I_f(x) = 0$.

Theorem 2.1.13 (Talagrand, [Tal93])**.** *Every* $f\colon \{0,1\}^d \to \{0,1\}$ *satisfies*

$$\mathbb{E}_x\left[\sqrt{I_f(x)}\right] = \Omega(\mathrm{var}(f)).$$

We can think of the square root function in Theorem 2.1.12 as interpolating smoothly between the two extremes "the edge boundary is large" vs. "the vertex boundary is large". Two very useful examples that illustrate this tradeoff are the anti-majority function and the anti-dictator function, for which we give a discussion in Fig. 2.1.

Structurally, the hope is that this gives an interpolation between the two cases "$G_f(X_f, Y_f, E_f)$ contains many edges" vs. "$G_f(X_f, Y_f, E_f)$ contains a relatively large matching". To obtain such a structural result [KMS18] actually show a much stronger, *robust* version of Theorem 2.1.12. Note that in Theorem 2.1.12 all the edge violations are "charged" to the the points x where $f(x) = 1$. Allowing for edges to be charged to different endpoints arbitrarily

[3] The original inequality due to [KMS18] had $\Omega(\varepsilon_f/\log d)$ on the RHS and this was improved to $\Omega(\varepsilon_f)$ by [PRW22].

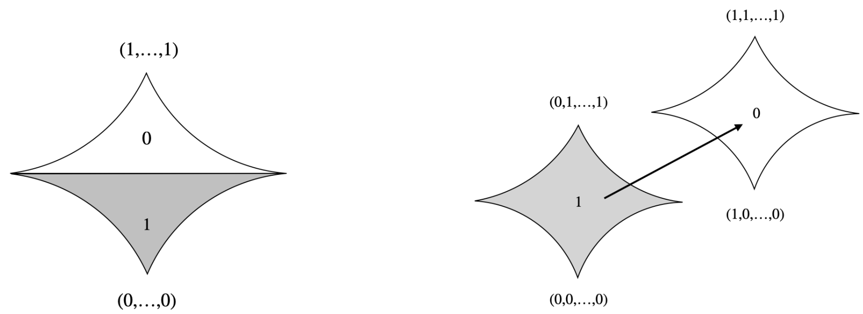


Figure 2.1: *On the left is a pictorial representation of an anti-majority function on the hypercube,* $\mathsf{AM}(x) = \mathbf{1}(|x| \leq \lfloor d/2 \rfloor)$*. On the right is a pictorial representation of a anti-dictator function* $\mathsf{AD}(x) = 1 - x_1$*. Both functions satisfy* $\varepsilon_{\mathsf{AM}} = \varepsilon_{\mathsf{AD}} = 1/2$*, while their edge-boundaries are quite different. We have* $I^-_{\mathsf{AD}} = 1/2$ *since the influential edges are exactly those crossing the first dimension cut, of which their are exactly* 2^{d-1}*. On the other hand* $I^-_{\mathsf{AM}} = \Omega(\sqrt{d})$ *since there are* $\binom{d}{\lfloor d/2 \rfloor} = \Omega(2^d/\sqrt{d})$ *vertices on the boundary of* $f^{-1}(1)$*, each incident to* $\lceil d/2 \rceil$ *influential edges. However, notice that both functions satisfy* $\mathbb{E}[\sqrt{I^-_{\mathsf{AD}}(x)}], \mathbb{E}[\sqrt{I^-_{\mathsf{AM}}(x)}] = \Theta(1)$*. Thus, both of these functions are tight examples for the directed Talagrand inequality Theorem 2.1.12, which is not the case for the directed Poincaré inequality Theorem 2.1.2.*

can greatly reduce the quantity in the LHS, due to the the square root inside the expectation. As a simple example, suppose $f(x) = 0$ iff $x = (1, 1, \ldots, 1)$. Then charging edges to the points where $f(x) = 1$ or where $f(x) = 0$ causes the LHS to equal $d \cdot 2^{-d}$ or $\sqrt{d} \cdot 2^{-d}$, respectively. [KMS18] are able to show that the inequality always holds regardless of how the edges are charged. This charging is specified by an edge coloring: given $\chi \colon E \to \{0, 1\}$, let $I^-_{f,\chi}(x)$ denote the number of violating edges (x, y) incident to x such that $\chi(x, y) = f(x)$.

Theorem 2.1.14 (Robust directed Talagrand, [KMS18])**.** *Every* $f \colon \{0,1\}^d \to \{0,1\}$ *and* $\chi \colon E \to \{0,1\}$ *satisfy*[3]

$$\mathbb{E}_x\left[\sqrt{I^-_{f,\chi}(x)}\right] = \Omega(\varepsilon_f).$$

The following lemma describes the structural property of the violation graph implied by Theorem 2.1.14.

Lemma 2.1.15 (Good subgraphs, informal, [KMS18])**.** *Suppose* $f \colon \{0,1\}^d \to \{0,1\}$ *is* $\Omega(1)$*-far from monotone. There exists* $\Delta \in [d]$ *such that* $G_f(X_f, Y_f, E_f)$ *contains a subgraph*

$G_{\mathsf{good}}(X, Y, E_{\mathsf{good}})$ *with max degree* Δ *and* $|E_{\mathsf{good}}| = \Omega(\sqrt{\Delta} \cdot 2^d)$.

If $\Delta = 1$, the good subgraph is a matching of size $\approx 2^d$ and when $\Delta = d$ the good subgraph contains at least $\approx \sqrt{d} \cdot 2^d$ edges. The analysis of the path tester (Definition 2.1.9) given by [KMS18] proves that there is a choice of $\tau \approx \sqrt{d/\Delta}$ for which the τ-length path test finds a violation with probability $\approx d^{-1/2}$. This results in the following theorem since the right choice of τ is selected with probability $1/\log d$.

Theorem 2.1.16 (Path tester bound from Talagrand, [KMS18])**.** *For every* $f \colon \{0,1\}^d \to \{0,1\}$*, the path test finds a violation of monotonicity with probability* $\widetilde{\Omega}(\varepsilon_f^2/\sqrt{d})$*. Therefore, the path tester is a* $\widetilde{O}(\sqrt{d}/\varepsilon^2)$ *query tester.*

2.1.4 The KMS Optimal Path Tester Analysis

The main ideas behind the KMS analysis of the path tester are best illustrated by the case when the good subgraph is a matching, i.e. when $\Delta = 1$ in Lemma 2.1.15.

When the good subgraph $G_{\mathsf{good}}(X, Y, E_{\mathsf{good}})$ is a matching from X to Y, KMS show that a random walk of length $\tau \approx \sqrt{d}$ succeeds in finding a violation with $\approx d^{-1/2}$ probability. A key definition in their analysis is the notion of τ*-persistence*: a vertex x is τ-persistent if a τ-length directed random walk leads to a point z where $f(x) = f(z)$ with constant probability. Using a simple argument based on the influence of the function, KMS argue that an average directed random walk has $\lessapprox \tau/\sqrt{d} = o(1)$ influential edges. Using Markov's inequality, at most $o(2^d)$ points in $\{0,1\}^d$ can be non-persistent. Thus, we can remove all non-persistent points and their matched partners from X and Y and still maintain $|X| = |Y| = \Omega(2^d)$. Thus, going forward we will assume that all points in $X \cup Y$ are persistent.

With $\Omega(1)$ probability, the tester starts from $x \in X$. Note that $f(x) = 1$. Let y denote x's partner in the matching and note that $f(y) = 0$. Let i be the dimension of the edge (x, y). With probability roughly $\tau/d \approx d^{-1/2}$, the directed walk will cross the ith dimension. Let us condition on this event. We can interpret the random walk as traversing the edge (x, y), and then taking a $(\tau-1)$-length directed walk from y to reach the destination y'. (Note that we do not care about the specific order of edges traversed by the random walk. We only care about the value at the destination.) Since y is τ-persistent[4], with $\Omega(1)$ probability the final destination y' will satisfy $f(y') = f(y) = 0$. Putting it all together, the tester succeeds with probability $\approx d^{-1/2}$.

	Undirected	Directed	Tester	Queries
Poincaré	$I_f = \Omega(\mathrm{var}(f))$	$I_f^- = \Omega(\varepsilon_f)$	Edge test	$O(d\varepsilon^{-1})$
Margulis	$I_f \cdot \Gamma_f = \Omega(\mathrm{var}(f)^2)$	$I_f^- \cdot \Gamma_f^- = \Omega(\varepsilon_f^2)$	Path test	$\widetilde{O}(d^{5/6}\varepsilon^{-4/3})$
Talagrand	$\mathbb{E}_x\left[\sqrt{I_f(x)}\right] = \Omega(\mathrm{var}(f))$	$\mathbb{E}_x\left[\sqrt{I_{f,\chi}^-(x)}\right] = \Omega(\varepsilon_f)$	Path test	$\widetilde{O}(\sqrt{d}\varepsilon^{-2})$

Table 2.2: *This table displays the isoperimetric inequalities for Boolean functions on the hypercube, $f\colon \{0,1\}^d \to \{0,1\}$, and the monotonicity testing results that follow from the directed versions. In fact, all of the directed inequalities hold for real-valued functions $f\colon \{0,1\}^d \to \mathbb{R}$ by [BKR23, Theorem 1.3].*

2.1.5 On the Relationships Between Isoperimetric Inequalities

In this section we briefly discuss the formal relationships between the isoperimetric inequalities over the hypercube covered in this section. First, each *directed inequality* implies its *undirected* analogue, e.g. Theorem 2.1.2 implies Theorem 2.1.4. A formal proof is given in [KMS18, Section 9.4]. The idea is that one can express the total influence as the sum of the negative and positive influences: $I_f(x) = I_f^-(x) + I_f^+(x)$. Then, for example, the directed Poincaré inequality yields $I_f = I_f^- + I_f^+ = \Omega(\varepsilon_f^+ + \varepsilon_f^-)$ where $\varepsilon_f^+, \varepsilon_f^-$ denote the distance of f to being monotone increasing and monotone decreasing, respectively. Thus, one just needs to show that $\varepsilon_f^+ + \varepsilon_f^- = \Omega(\mathrm{var}(f))$, or equivalently $\varepsilon_f^+ + \varepsilon_f^- = \Omega(d(f, \mathsf{constant}))$. See [KMS18, Lemma 9.6] for a proof of this.

Second, we show how the directed Talagrand inequality (Theorem 2.1.12) implies the directed Margulis inequality (Theorem 2.1.7), which implies the directed Poincaré inequality (Theorem 2.1.2). The same chain of implications holds for the undirected inequalities by

[4]Technically, one needs $(\tau-1)$-persistence, which holds from τ-persistence.

analogous arguments. These proofs are straightforward and can be originally attributed to [KMS18, Section 1.1].

Proof. Margulis $\implies$ *Poincaré*: Recall the definition of $I_f^-(x)$ and let $\Gamma_f^-(x) = \mathbf{1}(I_f^-(x) > 0)$. We have

$$I_f^- = \mathbb{E}_x[I_f^-(x)] \geq \mathbb{E}_x[\mathbf{1}(I_f^-(x) > 0)] = \mathbb{E}[\Gamma_f^-(x)] = \Gamma_f^-$$

and so $I_f^- \geq \sqrt{I_f^- \cdot \Gamma_f^-}$ which completes the proof. □

Proof. Talagrand $\implies$ *Margulis*: We have

$$I_f^- \cdot \Gamma_f^- = \mathbb{E}_x[I_f^-(x)] \cdot \mathbb{E}_x[\mathbf{1}(I_f^-(x) > 0)] \geq \mathbb{E}_x\left[\sqrt{I_f^-(x)}\right]^2$$

where the last step follows from Cauchy-Schwartz. □

2.2 Domain Reduction for Hypergrids

We turn now to the general case of hypergrids, $[n]^d$. What is the effect of n on the monotonicity testing problem? It is a folklore result that testing Boolean functions on the line, $f\colon [n] \to \{0,1\}$, can be done with $O(\varepsilon^{-1})$ queries, independent of n: (i) sample a set $T \subseteq [n]$ of $O(\varepsilon^{-1})$ random points and (ii) accept iff $f|_T$ is monotone. The analysis of this tester exploits the fact that $[n]$ is a *total order* and that the function is *Boolean*.

In contrast, testing real-valued functions on the line, $f\colon [n] \to \mathbb{R}$, requires $\Omega(\log n)$ queries (see e.g. [CS14b]). Essentially, this is because one can construct real-valued functions where all violations exist at a certain distance: consider the function assigning values $[2,1,4,3,6,5,\ldots,n,n-1]$. This function is 1/2-far from monotone while all violations occur between adjacent entries. Boolean functions on the line are quite different in that it is not possible to avoid "long-distance violations". I.e., consider a Boolean function with two violations between adjacent entries, $[\ldots,1,0,\ldots,1,0,\ldots]$. Here the left-most 1 and the right-most 0 also form a violation. Is it possible to exploit this structure for Boolean functions over the d-dimensional hypergrid, $[n]^d$? In particular, is it possible to exploit this structure to reduce monotonicity testing of $f\colon [n]^d \to \{0,1\}$ to monotonicity testing of $f\colon [k]^d \to \{0,1\}$ for $k \ll n$? We refer to this as the *domain reduction problem.*

Formally, consider sampling k i.i.d. uniform elements of $[n]$ across each dimension and restricting f to the resulting $[k]^d$ sub-grid. For k *independent* of n, can we lower bound the expected distance to monotonicity of this restriction? Note that if one can prove a lower bound of $\Omega(\varepsilon_f)$, then this reduces the problem over general $[n]^d$ to the smaller domain $[k]^d$. It turns out that one needs $k = \Omega(\sqrt{d})$ for this to hold [BCS20, Theorem 8.1] (and Theorem 3.6.1 in this thesis). In particular, a reduction to the hypercube domain is not possible using such a method. Nonetheless, we show that such a result indeed holds for $k = \mathrm{poly}(d/\varepsilon_f)$. We give a proof of this theorem in Chapter 3.

Theorem 2.2.1 (Domain Reduction Theorem for Hypergrids, [BCS20])**.** *Let $f\colon [n]^d \to \{0,1\}$ and $k \in \mathbb{Z}^+$. If $\boldsymbol{T} = T_1 \times \cdots \times T_d$ is a random sub-grid, where for each $i \in [d]$, T_i is a (multi)-set*[5] *formed by taking k independent uniform samples from $[n]$, then*

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

where $C > 0$ is a universal constant. In particular, if $k \geq \left(\frac{2Cd}{\varepsilon_f}\right)^7$, then $\mathbb{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f/2$.

For the $d = 1$ case (the line domain), [BRY14a] prove a stronger bound. More recently, [HY22] achieved a domain reduction style result using only $O((d/\varepsilon_f)^3)$ samples by choosing a random "gridding" of the domain. Both the results of [BCS20, HY22] also hold for measurable functions over continuous product spaces $f\colon \mathbb{R}^d \to \{0,1\}$.

[5] We treat duplicate elements of a multi-set as being distinct copies of that element, which are then treated as immediate neighbors in the total order.

2.3 Monotonicity Testing and Isoperimetry on the Hypergrid

Domain reduction (Theorem 2.2.1) allows us to assume that $n = \mathrm{poly}(d/\varepsilon)$, but even so the techniques discussed in Section 2.1 do not directly port to the hypergrid, even for $n = 3$. The initial proofs for the directed isoperimetric inequalities spelled out in Section 2.1 were all highly specific to the $n = 2$ case, and in fact it is not clear a priori what the right generalizations are. In this section we discuss the generalizations of the inequalities in Table 2.2 which have been proven for hypergrids and the resulting monotonicity testers.

All of the inequalities for hypercubes listed in Table 2.2 involve some notion of the negative influence and it is not clear what the right generalization of this is for functions $f\colon [n]^d \to \{0,1\}$. The typical treatment of the hypergrid as a DAG considers (x, y) to be an edge iff $\sum_{i=1}^{d} |x_i - y_i| = 1$. One may wonder why not define influence using this set of edges. Consider the function $f\colon [n] \to \{0,1\}$ defined as $f(x) = 1$ iff $x \leq n/2$. We have $\varepsilon_f = 1/2$, but only a *single* violating edge meaning that the influence would be only $O(1/n)$, precluding any inequality resembling those in Table 2.2 from holding. Thus, one needs to consider other notions of influence which consider "long-distance interactions" between points on the same line to obtain such inequalities.

2.3.1 The Line Tester and Dimension Reduction

Recall the directed Poincaré inequality Theorem 2.1.2 and the resulting bound for the edge tester, Corollary 2.1.3. A natural generalization of the edge tester is *the line test*, first considered by [DGL+99]: given $f\colon [n]^d \to \{0,1\}$, sample a random axis-parallel line ℓ in $[n]^d$ and run a $d = 1$ monotonicity tester on $f|_\ell$. A folklore result is that $O(\varepsilon^{-1})$ queries suffice to test monotonicity of Boolean functions on the line and so the query complexity of the line test depends on the distance to monotonicity, $\varepsilon_{f|_\ell}$, of this random line restriction. A lower bound on $\mathbb{E}_\ell[\varepsilon_{f|_\ell}]$ was proven by [DGL+99], who referred to this inequality as *dimension reduction*.

Theorem 2.3.1 (Directed Poincaré for Hypergrids, [DGL+99])**.** *Let* $f\colon [n]^d \to \{0,1\}$ *and let* ℓ *denote a uniform random axis-parallel line in* $[n]^d$*. Then* $\mathbb{E}_\ell[\varepsilon_{f|_\ell}] = \Omega(\varepsilon_f/d)$*.*

Note that when $n = 2$ (the hypercube) an axis-parallel line ℓ in $[n]^d$ is an edge. Moreover, if this edge is *a violation*, then $\varepsilon_{f|_\ell} = 1/2$ and otherwise $\varepsilon_{f|_\ell} = 0$. Thus, the above inequality generalizes the directed Poincaré inequality (Theorem 2.1.2). From Theorem 2.3.1 one obtains the following bound using the line tester with some technical modifications.

Theorem 2.3.2 (Line tester bound, [DGL+99, BRY14a])**.** *There is a* $\widetilde{O}(d\varepsilon^{-1})$ *query monotonicity tester for functions* $f\colon [n]^d \to \{0,1\}$*.*

2.3.2 The Directed Margulis Inequality for Hypergrids

In Section 2.1.2 we discussed how the edge tester requires $\Omega(d)$ queries and so to obtain $o(d)$ testers one needs to query pairs of points differing on more than one coordinate, i.e. *path testers*. As we discussed in Section 2.1.2, more sophisticated directed isoperimetric inequalities involving some notion of vertex boundary are needed to analyze path testers. This was first achieved for hypergrids by [BCS18] who generalized the directed Margulis inequality, Theorem 2.1.7. To capture long-distance interactions between points on the same line, they used the following notion of the "augmented hypergrid", which was first considered by [CS13].

Definition 2.3.3 (The augmented hypergrid, [CS13])**.** *Let* n *be a power of two. The* augmented hypergrid*, denoted* $\mathbf{A}_{n,d}$*, is the DAG with vertex set* $[n]^d$ *and edge set*

$$E = \{(x,y)\colon \exists i \in [d],\ m \in [\log n] \text{ such that } |x_i - y_i| = 2^m \text{ and } x_j = y_j \text{ for all } j \neq i\}.$$

For $f\colon \mathbf{A}_{n,d} \to \{0,1\}$, the negative influence I_f^- and directed vertex boundary Γ_f^- are defined analogously to the hypercube setting, but with respect to the edge set given in Definition 2.3.3. Similarly, let $\Gamma^-_{f,\mathsf{matching}}$ denote the size of the largest *matching* of decreasing edges for f in $\mathbf{A}_{n,d}$, divided by n^d.

Theorem 2.3.4 (Directed Margulis for Hypergrids, [BCS18]). *Every* $f \colon \mathbf{A}_{n,d} \to \{0,1\}$ *satisfies*

$$I_f^- \cdot \Gamma^-_{f,\mathsf{matching}} = \Omega(\varepsilon_f^2).$$

In the hypercube setting, recall that the τ-length path test samples x uniformly at random, chooses a set T of τ random coordinates, and then obtains y by incrementing x_i for all $i \in T$ where $x_i = 0$. In the hypergrid, we have a new issue to consider: by how much should we increment x_i? We will call this parameter the *step size*, which should also be chosen randomly according to some distribution. Loosely speaking, we will consider a *path test in the hypergrid* to be any generalization of Definition 2.1.9: choose $x \in [n]^d$ uniformly at random, then "increment" x in $\tau = 2^p$ coordinates where p is uniformly chosen from $[\log d]$.

In light of Theorem 2.3.4, [BCS18] choose the step size for the path tester as a random power of two in the interval $[1, n]$. Combining Theorem 2.3.4 and an appropriate generalization of the path tester analysis by [KMS18] (sketched in Section 2.1.4), [BCS18] show that this version of the path test finds a violation with probability $\Omega(\frac{\varepsilon^{4/3}}{d^{5/6}\text{polylog }(n,d,1/\varepsilon)})$. Since the dependence on n is poly-logarithmic, applying the domain reduction theorem, Theorem 2.2.1, yields the following testing result for hypergrids, generalizing Theorem 2.1.10.

Theorem 2.3.5 (Path test bound for hypergrids from domain reduction and Margulis, [BCS18, BCS20]). *For every* $f \colon [n]^d \to \{0,1\}$*, the path test using random-power-of-2 step sizes finds a violation of monotonicity with probability* $\widetilde{\Omega}(\varepsilon_f^{4/3} d^{-5/6})$*. Therefore, this is a* $\widetilde{O}(d^{5/6}\varepsilon^{-4/3})$ *query tester.*

2.3.3 The Directed Talagrand Inequality for Hypergrids

The question now is whether it is possible to generalize the directed Talagrand inequalities Theorems 2.1.12 and 2.1.14 to hypergrids. We achieve such a result in Chapter 4 (originally published in [BCS23b]) using yet another way of capturing long-distance interactions on a line, by what we call the *fully augmented hypergrid*.

Definition 2.3.6. *The* fully augmented hypergrid *is the DAG with vertex set* $[n]^d$ *and edge set*

$$E = \{(x,y)\colon \exists i \in [d] \text{ such that } x_i < y_i \text{ and } x_j = y_j \text{ for all } j \in [d] \setminus \{i\}\}.$$

Given a function $f\colon [n]^d \to \{0,1\}$, we use $G_f(X_f, Y_f, E_f)$ to denote the *violation graph* of f which is the subgraph of the fully augmented hypergrid defined as follows: $X_f = \{x\colon f(x) = 1\}$, $Y_f = \{y\colon f(y) = 0\}$, and $E_f = \{(x,y) \in E\colon x \in X_f,\ y \in Y_f\}$. An i-aligned violation is an edge $(x,y) \in E$ such that $x_i < y_i$ and $x \in X_f$, $y \in Y_f$.

Definition 2.3.7 (Thresholded Influence)**.** *Fix* $f\colon [n]^d \to \{0,1\}$ *and a dimension* $i \in [d]$. *Fix a point* $x \in [n]^d$. *The thresholded influence of* f *at* x *along coordinate* i *is denoted* $\Phi_f(x;i)$, *and has value* 1 *if there exists an* i*-aligned violation* (x,y). *The thresholded influence of* f *at* x *is* $\Phi_f(x) = \sum_{i=1}^d \Phi_f(x;i)$.

Note that the thresholded influence coincides with the definition of negative influence in the hypercube when $n = 2$. Also note that for any x, $\Phi_f(x) \in \{0,1,\ldots,d\}$ and is independent of n.

Theorem 2.3.8 (Directed Talagrand for Hypergrids, [BCS23b])**.** *Every* $f\colon [n]^d \to \{0,1\}$ *satisfies*

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

As is the case in the hypercube setting, obtaining the right structural result for the violation graph (recall Lemma 2.1.15) requires a robust version of Theorem 2.3.8. We achieve this using the following colorful version of the thresholded influence.

Definition 2.3.9 (Colorful Thresholded Influence)**.** *Fix* $f\colon [n]^d \to \{0,1\}$ *and* $\chi\colon E \to \{0,1\}$. *Fix a dimension* $i \in [d]$ *and a point* $x \in [n]^d$. *The colorful thresholded influence of* f *at* x *along coordinate* i *is denoted* $\Phi_{f,\chi}(x;i)$, *and has value* 1 *if there exists an* i*-aligned violation* (x,y) *such that* $\chi(x,y) = f(x)$, *and has value* 0 *otherwise. The colorful thresholded influence of* f *at* x *is* $\Phi_{f,\chi}(x) = \sum_{i=1}^d \Phi_{f,\chi}(x;i)$.

In Chapter 4 we prove a robust directed Talagrand isoperimetry theorem for Boolean functions on the hypergrid. It is a strict generalization of Theorem 2.1.14.

Theorem 2.3.10 (Robust Directed Talagrand for Hypergrids, [BCS23b])**.** *Every* $f\colon [n]^d \to \{0,1\}$ *and every* $\chi\colon E \to \{0,1\}$ *satisfies*

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

From Theorem 2.3.10, one obtains a generalization of Lemma 2.1.15 in the fully augmented hypergrid. Running the path tester with each step size chosen uniformly at random from $[n]$ yields the following testing result by following an appropriate generalization of the KMS analysis sketch in Section 2.1.4.

Theorem 2.3.11 (Uniform-length path test for hypergrids, [BCS23b])**.** *For every* $f\colon [n]^d \to \{0,1\}$*, the path test with uniformly random step sizes finds a monotonicity violation with probability* $\widetilde{\Omega}(\frac{\varepsilon^2}{n\sqrt{d}})$*. Thus, this is a* $\widetilde{O}(n\sqrt{d}\varepsilon^{-2})$ *query tester.*

Notice that here the dependence on n is linear, meaning that the domain reduction theorem, Theorem 2.2.1, by itself will not wash it away. Thus, a new set of ideas are required and this is where the contents of Chapter 5 (originally published in [BCS23a]) comes in.

	Directed Inequality	Tester	Queries
Dimension Reduction	$\mathbb{E}_\ell[\varepsilon_{f\|_\ell}] = \Omega(\varepsilon_f/d)$	Line test	$\widetilde{O}(d\varepsilon^{-1})$
Margulis	$I_f^- \cdot \Gamma_f^- = \Omega(\varepsilon_f^2)$	Path test in $\mathbf{A}_{n,d}$	$\widetilde{O}(d^{5/6}\varepsilon^{-4/3})$
Talagrand	$\mathbb{E}_x\left[\sqrt{\Phi_{f,\chi}(x)}\right] = \Omega(\varepsilon_f)$	Shifted path test	$d^{1/2+o(1)}\varepsilon^{-2}$

Table 2.3: *This table displays the directed isoperimetric inequalities for Boolean functions on the hypergrid,* $f\colon [n]^d \to \{0,1\}$*, and the monotonicity testing results that follow from them. In fact, all of these inequalities hold more generally for real-valued functions* $f\colon [n]^d \to \mathbb{R}$ *(by [BKR23, Theorem 1.3], or Theorem 6.0.1 in this thesis). See Theorem 6.0.3 which generalizes the robust directed Talagrand inequality in this table to real-valued functions.*

2.4 Nearly Optimal Path Tester Analysis for Hypergrids

The following question is the starting point for the most recent work of [BCS23a] for hypergrids: is it possible to remove the dependence on n from Theorem 2.3.11 with Theorem 2.3.10 being the isoperimetric inequality at the core of the analysis? We answer this question in the affirmative in Chapter 5, proving the following result, using a modified version of the path tester.

Theorem 2.4.1 (Shifted path test, [BCS23a])**.** *There is a path tester for hypergrids which, for every $f\colon [n]^d \to \{0,1\}$, finds a monotonicity violation with probability at least $\frac{\varepsilon^2}{d^{1/2+o(1)}}$. Thus, there is a $d^{1/2+o(1)}\varepsilon^{-2}$ query tester.*

We defer a definition of the path test used to prove Theorem 2.4.1 for now (see Definition 2.4.2 and Definition 2.4.3), until after we've given some context. For a formal definition see Section 5.1.

In addition to Theorem 2.3.10, the analysis leading to the above theorem requires many new ideas that go beyond the analysis due to KMS sketched in Section 2.1.4 for the hypercube case. This section is devoted to discussing the key challenges in analyzing path testers from Theorem 2.3.10 without incurring a polynomial dependence on n and how we circumvent these challenges in Chapter 5.

As we mentioned, Theorem 2.3.10 implies a similar structural result to Lemma 2.1.15 implying the existence of a "good subgraph" $G_{\mathsf{good}}(X, Y, E_{\mathsf{good}})$ in the fully augmented hypergrid. The simplest, most instructive case is again when this graph is a matching between X and Y. As in our discussion for the hypercube case, let us assume that $\varepsilon_f = \Omega(1)$ and so $|X| = |Y| = \Omega(n^d)$. Note that the matched pairs (x, y) are axis-aligned, that is, differ in exactly one coordinate i, but now $y_i - x_i$ is an integer in $\{1, 2 \dots, n-1\}$. The remainder of this section is devoted to showing how we analyze path testers in the hypergrid under the assumption that we have such a matching. First, we discuss briefly how one obtains an $O(n\sqrt{d})$ tester by a natural extension of the KMS analysis for the hypercube case. Then we discuss how we improve this to $O(\log n\sqrt{d})$ and the key challenges in doing so. Note that by domain reduction, Theorem 2.2.1, we may assume $n \leq \text{poly}(d/\varepsilon_f)$ and so $O(\log n\sqrt{d}) = \widetilde{O}(\sqrt{d})$.

An $O(n\sqrt{d})$ Tester in the Matching Case: Recall the KMS path tester analysis for the matching case which we sketched in Section 2.1.4. One can generalize the hypercube persistence arguments to again allow us to assume that every point participating in the matching is persistent while still maintaining $|X| = |Y| = \Omega(n^d)$.

The tester picks $x \in X$ with $\Omega(1)$ probability. Let y be its matched partner, which differs in the ith coordinate. If the number of steps is $\tau \approx \sqrt{d}$, then with $\tau/d \approx d^{-1/2}$ probability, the walk will choose to move along the ith coordinate. Conditioned on this event, we would like to relate the random walk to a persistent walk from y. However, there is only a $1/n$ chance that the *length* jumped along that coordinate will be exactly $y_i - x_i$. Thus, this loses an n factor, and indeed this is the argument leading to Theorem 2.3.11.

2.4.1 The BCS Shifted Path Tester Analysis for the Matching Case

In this section we sketch the arguments from [BCS23a] that lead to a $O(\log n\sqrt{d})$ query tester when the good subgraph is a *matching* of $\Omega(n^d)$ violating edges in the fully augmented hypergrid. In general, the good subgraph may not be a matching. This leads to another set of challenges, which lead us to incur an extra $d^{o(1)}$ factor. We briefly discuss these at the end of the section.

Internal Points and the Issue of Persistence: Our first idea is to try conditioning the random walk on passing through a random *internal* point between a matched pair (x, y). Formally, let $I(x, y) = \{w \colon x \preceq w \preceq y\}$. For the random walk to pick a random point in $I(x, y)$ with large enough probability, one needs to first modify the way the step sizes are chosen in each coordinate. Thus, the tester will generate walks according to the following distribution. See Definition 5.1.1 for a formal definition.

Definition 2.4.2 (Random Walk Distribution, informal)**.** *A τ-step upwalk from a point $x \in [n]^d$ generates a point $z \succeq x$ as follows. First choose $T \subset [d]$ of τ coordinates at random to increase. For each $i \in T$, the step size for coordinate i is chosen by picking $q_i \in [\log n]$ uniformly, then picking a uniform random step value $a_i \in [\min(2^{q_i}, n-x_i)]$. Then, $z_i = x_i + a_i$ for each $i \in T$ and $z_i = x_i$ for each $i \notin T$. Downwalks are defined analogously.*

The point of choosing step sizes in this way is as follows. Again, we set the walk length as $\tau \approx \sqrt{d}$ and we may assume that every point in $X \cup Y$ is persistent with respect to our new random walk distribution, Definition 2.4.2. Our tester will pick $x \in [n]^d$ uniformly at random, generate z according to the new random walk distribution, and check if $f(x) > f(z)$. Again, our starting point x belongs to X with $\Omega(1)$ probability. Let $y \in Y$ be x's partner in the matching. I.e. $y = x + s \cdot e_i$ for some $i \in [d]$, $s \in [n]$. Again, we have $i \in T$ with probability $\approx d^{-1/2}$. For the step size, we have $q_i \in [s, 2s]$ with probability $\frac{1}{\log n}$. Conditioned on these events, the random walk passes through a uniform random internal point $w \in I(x, y)$ with constant probability. We may assume that at least half of $w \in I(x, y)$ satisfy $f(w) = 0$, for otherwise we may perform a symmetric tester analysis using the version of the tester performing downwalks. Now, if we could argue that these internal points were persistent, then we would be done. Unfortunately, this is not possible; even though these edges form a matching, i.e. their endpoints are distinct, *they may share internal points arbitrarily.* Thus, it is possible that the number of internal points is extremely small and so removing non-persistent points may remove all internal points. This brings us to our main idea which circumvents this issue. The following is an informal statement of the tester. We provide a sketch for why this tester works well in the remainder of the section.

Definition 2.4.3 (Shifted Path Test, informal, [BCS23a])**.** *Sample a pair of points according to the following distributions, each with probability* $1/2$*:*

1. *(Down-shifted Upwalk) Sample $x \in [n]^d$ uniformly at random and generate z according to the τ-length upwalk distribution from x. Generate z' by a $(\tau - 1)$-length downwalk from z and let $s = z - z'$. Return $(x - s, z - s)$.*

2. *(Downwalk) Sample $x \in [n]^d$ uniformly at random and generate z according to the τ-length downwalk distribution from x. Return (z, x).*

Mostly-Zero-Below Points, and Red Edges: The following is a key definition: call a point z *mostly-zero-below* for length $\tau - 1$, or simply $(\tau-1)$-mzb, if a $(\tau-1)$-length downwalk from z leads to a zero with ≥ 0.9 probability. Suppose an upwalk of length $\tau - 1$ from a point $x \in X$ reaches an $(\tau-1)$-mzb point z. Then, a random shift $(x - s, z - s)$ has a constant probability of being a violation. The reason is (i) $\mathbb{P}[f(x-s) = f(x) = 1] \geq 0.9$ because x is $(\tau-1)$-persistent, and (ii) $\mathbb{P}[f(z-s) = 0] \geq 0.9$ because z is $(\tau-1)$-mzb. By a union bound, the tester will find a violation with constant probability (conditioned on discovering the pair (x, z)).

To formalize this analysis, we call a matching edge (x, y) *red* if for a constant fraction of the interior points $w \in I(x, y)$, a $(\tau-1)$-length upwalk ends at a $(\tau-1)$-mzb point with constant probability. If there are $\Omega(n^d)$ red matching edges, one can argue that the down-shifted upwalk test (item (1) of Definition 2.4.3) succeeds with the desired probability. Firstly, with probability $\Omega(1)$, the tester starts the walk at an endpoint x of a red edge, (x, y). With probability $\tau/d \approx d^{-1/2}$, the walk will cross the dimension corresponding to (x, y). Conditioned on this event, we can interpret the walk as first moving to a random internal point $w \in I(x, y)$ (with probability $\approx \frac{1}{\log n}$) and then taking a $(\tau-1)$-length upwalk from w to get to the point z. (Refer to the left side of Fig. 2.2.) Since the edge was red, with constant probability, z is $(\tau-1)$-mzb. Consider a random shift of (x, z), shown as $(x - s, z - s)$ in Fig. 2.2. As discussed in the previous paragraph, this shifted pair is a violation with constant probability. All in all, the tester succeeds in finding a violation with $\approx \frac{1}{\sqrt{d}\log n}$ probability. But what if there are no red edges? This brings us to the next key idea.

Translations of Violation Subgraphs, and Blue Edges. Suppose a matching edge (x, y) is non-red. So, for most internal points $w \in I(x, y)$, a $(\tau-1)$-length walk will instead

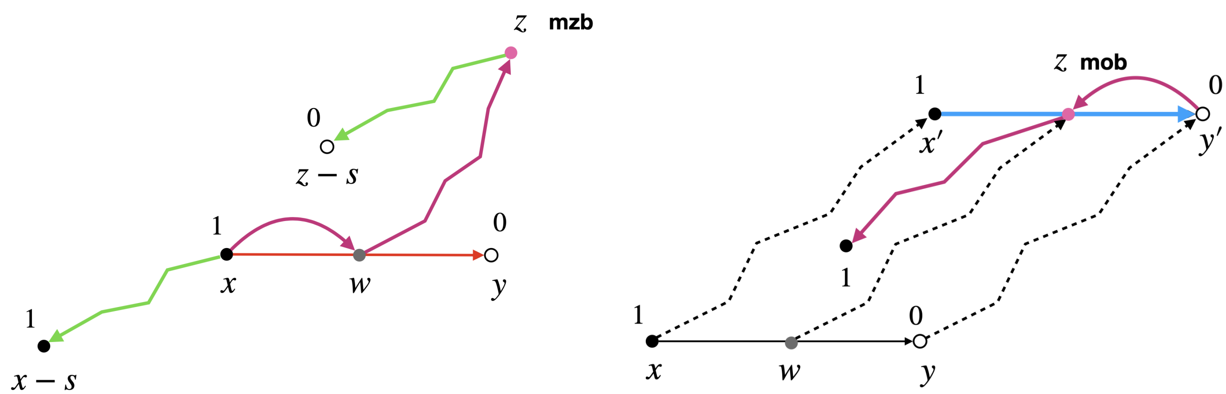


Figure 2.2: *This figure shows the key argument that either the down-shifted upwalk test or the downwalk test find a violation with probability* $\approx \frac{1}{\sqrt{d}\log n}$ *when the good subgraph is a matching of size* $\Omega(n^d)$*. The left drawing illustrates how red edges enable the analysis of the down-shifted upwalk test and the right drawing illustrates how blue edges enable the analysis of the downwalk test. In both drawings, the edge* (x, y) *is in the initial violation matching.*

reach a $(\tau-1)$-*mostly-one-below* (mob) point, z. (Refer to the right side of Fig. 2.2). Fix one such walk, which can be described by an "up-shift" s. I.e. the walk from w reaches $z := w+s$. Consider the corresponding shift of the full edge (x, y) to (x', y'), where $x' = x + s$ and $y' = y+s$. What can we say about this edge? Since both x and y are $(\tau-1)$-persistent, with high probability both $f(x') = f(x) = 1$ and $f(y') = f(y) = 0$. Observe that most internal points $z \in I(x', y')$ are $(\tau-1)$-mob. Now, performing a τ-length downward random walk from y' will pass through a mostly-one-below internal point from $I(x', y')$ with probability $\approx \frac{1}{\sqrt{d}\log n}$. This motivates the definition of a *blue* edge. A violating edge is called blue if a constant fraction of its internal points are $(\tau-1)$-mostly-one-below.

Red-blue Win-win Flow Argument: We now have two types of edges which are each good for the tester (Definition 2.4.3) for different reasons; red edges are good for the down-shifted upwalk test, and blue edges are good for the downwalk test. Again, if the matching contains a large fraction of red edges, then the down-shifted upwalk test discovers a violation with probability $\approx \frac{1}{\sqrt{d}\log n}$.

Suppose instead that most edges in the matching are non-red. Then informally, we argue that performing a random translation of the matching discovers another violation matching consisting mostly of blue edges. What does it mean to translate "all edges together"? Through the random translation, every non-red edge (x, y) in the original violation matching leads to a *distribution* over blue edges (x', y'). The idea is to treat this as a fractional flow on these blue edges. If the original matching had few red edges, one can construct a large collection of blue edges sustaining a large flow. Integrality of flow implies there must be another large violation matching in the support of this distribution whose edges are blue. This large blue matching implies that the downwalk test succeeds in discovering a violation with probability $\approx \frac{1}{\sqrt{d}\log n}$. This concludes the proof sketch for the case when the good subgraph is a matching.

Thresholded Degrees, Peeling, and the $d^{o(1)}$ Loss: Another gnarly issue with hypergrids is the distinction between degree and "thresholded degree". The relevant "degree" of a vertex x (for the path tester analysis) in a violation subgraph is not the *number* of edges incident to it, but rather the number of different dimensions i so that there is an i-edge incident to it. This is the "thresholded degree" (coming from Definition 2.3.7), and it is between 0 and d, whereas the standard degree could be as large as nd. It is critical one uses thresholded degree for the path tester analysis, to avoid the linear dependence on n. Observe that for the matching case, these degrees are identical, making the analysis easier.

While the path tester analysis works with thresholded degree, the flow-based translation arguments alluded to above need to use normal degrees. In particular, one can use flow-arguments to relate the bound on the standard degree of the new violation subgraphs. However, one cannot a priori do so for the thresholded degree. To argue about the thresholded degree, we need a stronger notion of a good subgraph, satisfying specific conditions for both thresholded and standard degrees of the vertices. It is in the construction of this graph where we lose the $d^{o(1)}$ factor.

2.5 Open Questions in Monotonicity Testing

We conclude the chapter by discussing some of the open problems in monotonicity testing which we find most notable.

Adaptivity: The most obvious question is to resolve the gap of $\widetilde{\Omega}(d^{1/3})$ vs $\widetilde{O}(\sqrt{d})$ for adaptive testing.

Question 2.5.1. *What is the adaptive query complexity of monotonicity testing of functions* $f\colon \{0,1\}^d \to \{0,1\}$*?*

So far, essentially all upper bounds for monotonicity testing use non-adaptive testers. The state of the art non-adaptive testers of [KMS18, BCS23a] work by querying the endpoints of a directed random walk. At a qualitative level, the longer the random walk the better, since this increases the likelihood of the test leaving the set of 1's. An important limitation of this approach is that the walk length cannot be $\gg \sqrt{d}$ since otherwise one of the endpoints will be outside of the middle layers of the hypercube. For any function $f\colon \{0,1\}^d \to \{0,1\}$ one can consider the "truncated" version of f which sets $f(x) = 0$ whenever $|x| < \frac{d}{2} - C\sqrt{d\log(d/\varepsilon)}$ and $f(x) = 1$ whenever $|x| > \frac{d}{2} + C\sqrt{d\log(d/\varepsilon)}$. This truncation has a negligible effect on the distance to monotonicity, but now querying points in the outer layers gives no information.

Thus, to increase the walk length beyond $\sqrt{d}$, one has to consider *undirected* random walks: sample x uniformly and obtain y by incrementing or decrementing x on a random set of τ coordinates. The problem is now that x and y are not comparable. To deal with this, one may try to use adaptivity in various ways. For example, [CS19] gave an $O(I)\cdot\mathrm{poly}(\varepsilon^{-1}, \log d)$ query adaptive tester for functions with total influence $I_f \leq I$, which uses adaptivity to perform a binary search procedure which looks for violations on a long *undirected* random walk. Another novel use of adaptivity is described by [CWX17] (see Section 7) who gave an adaptive tester achieving $O(d^{1/3})$ queries against the hard distribution which they used to prove their $\widetilde{\Omega}(d^{1/3})$ lower bound. This tester uses adaptivity to search for coordinates which are "safe" to decrement.

Tolerant Testing: A *tolerant monotonicity tester* is given two proximity parameters $\varepsilon_1 < \varepsilon_2 \in (0,1)$ and should (a) accept with probability $\geq 2/3$ when $\varepsilon_f \leq \varepsilon_1$ and reject with probability $\geq 2/3$ when $\varepsilon_f \geq \varepsilon_2$. Tolerant monotonicity testing has seen some recent attention [PRW22, BKR23, CDL$^+$23]. Motivated by a direct connection with the problem of approximating the distance to monotonicity, [PRW22] and [BKR23] obtain $\mathrm{poly}(n, 1/\varepsilon_2)$ query tolerant testers for the "large-gap" regime when $\varepsilon_1 \leq \varepsilon_2/\Omega(\sqrt{d \log d})$ for Boolean functions, and real-valued functions, respectively, over the hypercube. In general, agnostic learning algorithms (e.g. [LV23]) imply $2^{\widetilde{O}(\sqrt{d/(\varepsilon_2-\varepsilon_1)})}$ query testers and recently [CDL$^+$23] proved a $2^{\Omega(d^{1/4}/\sqrt{\varepsilon_2-\varepsilon_1})}$ lower bound for non-adaptive testers. Resolving this gap is an outstanding open question.

Hypergrids: In light of our $d^{1/2+o(1)}$ tester for hypergrids, one may wonder if the $o(1)$ factor can be removed to obtain a bound more closely resembling the best upper of $\widetilde{O}(\sqrt{d})$ for hypercubes due to [KMS18].

Question 2.5.2. *Is there an $\widetilde{O}(\sqrt{d})$ query monotonicity tester for functions $f \colon [n]^d \to \{0,1\}$?*

A possible approach towards answering this question was proposed by [BCS23b, Section 8] who conjectured a stronger directed isoperimetric inequality than Theorem 2.3.10 holds, using a notion they called the *weighted influence* [BCS23b, Def. 8.1, 8.3]. In fact they prove that the non-robust version of the directed Talagrand inequality using weighted influence does hold [BCS23b, Theorem 8.6]. Interestingly, it follows quite easily from the robust version of the Talagrand inequality using *thresholded influence* (Theorem 2.3.10). However, to yield the desired monotonicity testing result we again require a robust version, which seems much harder to prove [BCS23b, Conjecture 8.5].

Monotonicity Testing on General Posets: Another natural question which has received little attention is that of monotonicity testing of Boolean functions over arbitrary partial orders (posets). For N-element posets, [FLN+02] proved an upper bound of $O(\sqrt{N/\varepsilon})$ and a lower bound of $N^{\Omega(\frac{1}{\log\log N})}$ for non-adaptive testers. It would be interesting to resolve this gap. Towards this, it would be interesting to understand what structural properties of a poset dictate the hardness of testing monotonicity. Given that directed isoperimetric theorems have been central in the study of monotonicity testing over hypercubes and hypergrids, it may be an interesting direction to explore the concept of directed isoperimetry for other partial orders.

Real-Valued Functions: Recently [Fer23] studied monotoncity testing of Lipshitz functions with continuous domain and range $f\colon [0,1]^d \to \mathbb{R}$ with respect to L^1 distance. They prove a version of the the directed Poincaré inequality (see [Fer23, Theorem 1.2]) for such functions and conjecture that a version of the directed Talagrand inequality should also hold (see [Fer23, Conjecture 1.8]).

Functions with Bounded Image Size: Monotonicity testing of functions whose image consists of at most r distinct elements $f\colon \{0,1\}^d \to [r]$ is studied in Part II of this thesis (originally published in [BKR23]) and also by [PRV18]. We prove a Boolean decomposition theorem (see Theorem 6.0.1 or [BKR23, Theorem 1.3]) for functions $f\colon D \to \mathbb{R}$ for any partial order D which implies that all the directed isoperimetric inequalities discussed in this chapter also hold for real-valued functions, even those over hypergrids (in particular, see Theorem 6.0.3). Still, many aspects of the path tester need to be generalized and the modified analysis incurs a loss of a factor of r. We generalize this analysis over the hypercube in Chapter 7 and show matching upper and lower bounds of $\widetilde{\Theta}(r\sqrt{d})$ for one-sided non-adaptive monotonicity testing of functions $f\colon \{0,1\}^d \to [r]$. A natural question is whether the tester analysis of [BCS23a] (Chapter 5 of this thesis) for $f\colon [n]^d \to \{0,1\}$ can be generalized to functions with image $[r]$. We believe this is possible using Theorem 6.0.3 (our robust directed Talagrand inequality for $f\colon [n]^d \to \mathbb{R}$), but may require carefully generalizing each step of the arguments in Chapter 5, which is quite involved and technical.

Conjecture 2.5.3. *There is a* $(r\sqrt{d})^{1/2+o(1)}\varepsilon^{-2}$ *query monotonicity tester for functions* $f\colon [n]^d \to [r]$.

Another natural open question is whether one can generalize the $\widetilde{\Omega}(\sqrt{d})$ lower bound for non-adaptive two-sided testers of [CWX17].

Conjecture 2.5.4. *Non-adaptive monotonicity testing of functions* $f\colon \{0,1\}^d \to [r]$ *requires* $\widetilde{\Omega}(r\sqrt{d})$ *queries.*

Part I

Monotonicity Testing of Boolean Functions over Hypergrids

CHAPTER 3

Domain Reduction

In this chapter we prove domain reduction theorems for d-dimensional hypergrids and continuous product spaces. These results were originally published in [BCS20]. We refer the reader to Section 2.2 for a discussion on domain reduction. Our main result is the following *domain reduction theorem for hypergrids*, which reduces monotonicity testing of Boolean functions over $[n]^d$ for arbitrary n to the same problem over domain $[k]^d$ where $k \leq \text{poly}(d/\varepsilon)$.

Theorem 3.0.1 (Domain Reduction Theorem for Hypergrids)**.** *Let $f \colon [n]^d \to \{0,1\}$ be any function and let $k \in \mathbb{Z}^+$ be a positive integer. If $\boldsymbol{T} = T_1 \times \cdots \times T_d$ is a randomly chosen sub-grid, where for each $i \in [d]$, T_i is a (multi)-set formed by taking k i.i.d. samples from the uniform distribution on $[n]$, then*

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

where $C > 0$ is a universal constant. In particular, if $k \geq \left(\frac{2Cd}{\varepsilon_f}\right)^7$, then $\mathbb{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f/2$.

The construction in Section 3.6 shows that such a theorem is impossible for $k = o(\sqrt{d})$, and thus, domain reduction requires k and d to be polynomially related.

Continuous Domains. The independence of n in Theorem 3.0.1 suggests the possibility of a domain reduction result for Boolean functions defined over $\mathbb{R}^d$. We show that this is indeed true if $f \colon \mathbb{R}^d \to \{0,1\}$ is measurable (formal definitions in Section 3.5) and defined with respect to a (Lebesgue integrable) product distribution.

Theorem 3.0.2 (Domain Reduction Theorem for $\mathbb{R}^d$)**.** *Let $f\colon \mathbb{R}^d \to \{0,1\}$ be any measurable function and let $k \in \mathbb{Z}^+$ be a positive integer. Let $\mathcal{D} = \prod_{i=1}^d \mathcal{D}_i$ be a (Lebesgue integrable) product distribution such that the distance to monotonicity of f w.r.t. $\mathcal{D}$ is ε_f. If $\boldsymbol{T} = T_1 \times \cdots \times T_d$ is a randomly chosen hypergrid, where for each $i \in [d]$, $T_i \subset \mathbb{R}$ is formed by taking k i.i.d. samples from $\mathcal{D}_i$, then $\mathbb{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f - \frac{C \cdot d}{k^{1/7}}$, where $C > 0$ is a universal constant. In particular, if $k \geq \left(\frac{2Cd}{\varepsilon_f}\right)^7$, then $\mathbb{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq \varepsilon_f/2$.*

The above theorem essentially reduces the continuous domain to a discrete hypergrid $[k]^d$ where k is at most some polynomial of the dimension d, and this enables one to obtain monotonicity testers for Boolean functions over $\mathbb{R}^d$ under product measures.

The main ingredient in the proof of Theorem 3.0.2 is a discretization lemma (Lemma 3.5.5). Using standard measure theory, one can show that for any measurable Boolean function over $\mathbb{R}^d$ and any $\delta > 0$, there exists a large enough natural number $N = N(f, \delta)$ with the following property. The domain $\mathbb{R}^d$ can be divided into an N^d sized d-dimensional grid, such that in at least a $(1-\delta)$-fraction of grid boxes, the function f has the same value. (In some sense, this is what it means for f to be measurable.) Ignoring the δ-fraction of "mixed" boxes, the function f can be thought of as a discrete function on $[N]^d$.

The only guarantee on N is that it is finite; as it depends on f, N could be extremely large compared to d. This is where Theorem 3.0.1 shows its power. The sampling parameter k is independent of N, and this establishes Theorem 3.0.2. We give a detailed proof in Section 3.5.2.

3.1 Domain Reduction Proof Overview

Theorem 3.0.1 is a direct corollary of the following lemma, applied to each dimension.

Lemma 3.1.1 (Domain Reduction Lemma)**.** *Let $f\colon [n] \times \left(\prod_{i=2}^d [n_i]\right) \to \{0,1\}$ be any function over a rectangular hypergrid for some $n, n_2, \ldots, n_d \in \mathbb{Z}^+$ and let $k \in \mathbb{Z}^+$. Choose T to* *be a (multi-) set formed by taking k i.i.d. samples from the uniform distribution on $[n]$ and let f_T denote f restricted to $T \times \left(\prod_{i=2}^{d}[n_i]\right)$. Then $\mathbb{E}_T\left[\varepsilon_f - \varepsilon_{f_T}\right] \leq \frac{C}{k^{1/7}}$ where $C > 0$ is a universal constant.*

This lemma is the heart of our results, and in this section we give an overview of its proof. Let us start with the simple case of $d = 1$ (the line). Monotonicity testers for the line immediately imply domain reduction for $d = 1$ [DGL+99, BRY14a]. A u.a.r. sample of $\widetilde{O}(1/\varepsilon_f)$ points in $[n]$ contains a monotonicity violation with large probability ($> 9/10$, say), and thus the restriction of f to this sample has distance $\widetilde{\Omega}(\varepsilon_f)$. However, $\Omega(\varepsilon_f)$ is weak for what we need since, even if one could generalize this argument to the setting of Lemma 3.1.1, we would need to apply it d times to get the full domain reduction (Theorem 3.0.1). This would imply a final lower bound of ε_f/C^d, for some constant C, which has little value towards proving a sublinear-in-d query tester.

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

Consider the general case of Lemma 3.1.1. For brevity, we let $D := [n] \times \left(\prod_{i=2}^{d}[n_i]\right)$ and $D_T := T \times \left(\prod_{i=2}^{d}[n_i]\right)$ denote the original and reduced domains, respectively. Note that $|D_T| = \frac{k}{n}|D|$.

The standard handle on the distance to monotonicity is the violation graph of f, arguably first formalized by Fischer et al. [FLN+02]. The graph has vertex set D and an edge (x, y) iff $x \prec y$ and $f(x) = 1, f(y) = 0$. A theorem of [FLN+02] states that any maximum cardinality matching M in the violation graph satisfies $|M| = \varepsilon_f|D|$. Fix such a matching M. For a fixed sample T, we let M_T denote a maximum cardinality matching in the violation graph of f_T. To argue about ε_{f_T}, we want to give a lower bound on the expected size $|M_T|$. To do so, we give a lower bound the expected number of endpoints of M that can still be matched (simultaneously) in the violation graph of f_T.

We use the following standard notions of lines and slices in D, with respect to the first dimension. Refer to Fig. 3.1 and Fig. 3.2 for visual examples in two dimensions. In these examples the rows represent the lines while the columns represent the slices. Below, for $x \in D$, the vector x_{-1} is used to denote $(x_2, x_3, \ldots, x_d)$.

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

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

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

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

- (The matching $M_T^{(\ell)}$) For each $\ell \in \mathcal{L}$, consider the collection of all maximum cardinality violation matchings w.r.t. f_T on the set of vertices that (a) are matched by $M^{(\ell)}$, and (b) lie in some slice S_i where $i \in T$. We let $M_T^{(\ell)}$ denote any such fixed matching.

We stress that $M_T^{(\ell)}$ is *not* a subset of $M^{(\ell)}$, but the *endpoints* of the pairs in $M_T^{(\ell)}$ are a subset of the *endpoints* of the pairs in $M^{(\ell)}$. Thus, by the above definition, the union $M_T := \cup_{\ell \in \mathcal{L}} M_T^{(\ell)}$ is a valid matching in the violation graph of f_T since $M^{(\ell)}$ and $M^{(\ell')}$ have disjoint endpoints for all $\ell \neq \ell' \in \mathcal{L}$. We will lower bound the size of this matching, $|M_T|$, by giving a lower bound on $|M_T^{(\ell)}|$ for each line ℓ.

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

Since all the x_i's lie on the same line ℓ, their 1-coordinates are distinct. Suppose that the 1-coordinates of all the y_i's were also distinct and distinct from those of the x_i's too. Under this assumption we can proceed with our analysis as if all the x_i's *and* y_i's lie on ℓ, and the analysis becomes identical to the one-dimensional case. We could thus apply [BRY14a, Theorem 3.1] to each $\ell \in \mathcal{L}$ to prove Lemma 3.1.1. However, the assumption that the y_i's have distinct 1-coordinates is far from the truth. As we explain below, there are examples where all the y_i's have the same 1-coordinate, thereby lying in the same slice S_a (for some $a \in [n]$). In this case, with probability $(1 - k/n)$ we would have the size of $M_T^{(\ell)}$ be 0 (if $a \notin T$), implying that $\mathbb{E}_T\left[|M_T^{(\ell)}|\right]$ could be as small as $(k/n)^2 \cdot |M^{(\ell)}|$. Thus, if there existed a function f such that a "collision of y's 1-coordinates" could not be avoided for a large number of lines, then this would preclude such a line-by-line approach to proving Lemma 3.1.1. Unfortunately, there are examples of violation matchings where this happens. Consider Ex. 3.1.4, and the left part of Fig. 3.2, shown at the end of this section. For the lowest line, all the corresponding y's in $M^{(\ell)}$ have the same 1-coordinate.

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

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

We call $|M^{(\ell,S)}|$ the "*size* of the stack (ℓ, S)". To summarize the above discussion, small stacks are good news while big stacks are bad news. This is formalized in Lemma 3.1.3.

If there is a maximum cardinality matching M in the violation graph of f such that all stacks have size at most 1, then the one-dimensional domain reduction can be directly applied. Unfortunately, this is not possible. We give an example in Fig. 3.1 of a function where stacks of size at least 2 are unavoidable[1]. One reason for this difficulty may be that there can be various maximum cardinality matchings in the violation graph that have vastly different stack sizes (shapes); again consider Ex. 3.1.4. Nevertheless, we prove that there is a matching M such that for every positive integer λ, the total number of pairs belonging to stacks of size at least λ is at most $|D|/\text{poly}(\lambda)$.

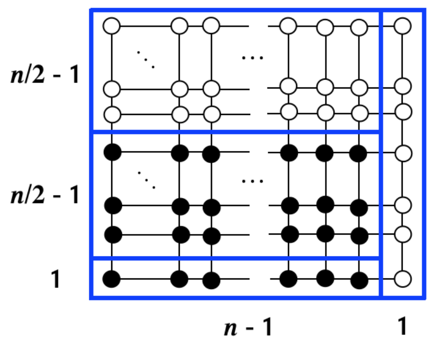


Figure 3.1: *An example of a function $f\colon [n] \times [n-1] \to \{0,1\}$ where stacks of size ≥ 2 are unavoidable. Black (white, resp.) circles represent vertices where $f = 1$ ($f = 0$, resp.). First observe that there exists a perfect violation matching as follows: perfectly match the two blocks of size $(n-1)(n/2-1)$ and then perfectly match the bottom line of 1's to the right-most slice of 0's. Thus, any maximum cardinality violation matching, M, will match all of the $(n-1)$ 0's in the right-most slice. There are only $n/2$ lines containing 1's and so by the pigeonhole principle M contains at least $n/2-1$ pairs belonging to stacks of size ≥ 2.*

Lemma 3.1.2 (Stack Bound)**.** *There exists a maximum cardinality matching M in the violation graph of f such that for every $\lambda \in \mathbb{Z}^+$, M satisfies $\sum_{(\ell,S):|M^{(\ell,S)}|\geq\lambda} |M^{(\ell,S)}| \leq \frac{5}{\sqrt{\lambda}} \cdot |D|$.*

The main creativity to prove this lemma lies in the choice of M. Given a matching, we define the vector $\Lambda(M)$ that enumerates all the stack sizes in non-decreasing order. We show that the maximum cardinality matching M with the lexicographically largest $\Lambda(M)$ serves our purpose. That is, we choose M that maximizes the minimum stack size, and then

[1]Interestingly, we don't know of a function where stacks of size strictly larger than 2 can't be avoided. In fact, we can prove that for the grid (the $d = 2$ case) one can always find a maximum cardinality violation matching M where $|M^{(\ell,S)}| \leq 3$ for all (ℓ, S). The proof is cumbersome and so we exclude it since it is not relevant to our main result.

subject to this maximizes the second minimum, and so on. It may seem counter-intuitive that we want a matching with small stack sizes, and yet our potential function maximize the minimum. The intuitive explanation is that the sum of the stack sizes is $|M|$, which is fixed, and so in a sense maximizing the minimum also balances out the $\Lambda(M)$ vector. The proof uses a matching rewiring argument to show that any large stack must be "adjacent" to many moderate size stacks. If two stacks are appropriately "aligned", one could change the matching to move points from one stack to the other. Large stacks cannot be aligned with small stacks, since one could rewire the matching to increase the potential. But since the function is Boolean one can show that there are many opportunities for rewiring the violation matching. Thus, there isn't enough "room" for many large stacks. We then apply some technical charging arguments to bound the total number of points in large stacks. The full proof is given in Section 3.3.

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

Lemma 3.1.3 (Line Sampling)**.** *Suppose that M is a matching in the violation graph of f, such that for some $\lambda \in \mathbb{Z}^+$, $|M^{(\ell,S)}| \leq \lambda$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$. Then, for any $\ell \in \mathcal{L}$,*

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

The proof is a fairly straightforward generalization of the arguments in [BRY14a] for the $\lambda = 1$ case. The idea is to control the size of the maximum cardinality matching $M_T^{(\ell)}$ by analyzing the discrepancy of a random subsequence of a sequence of 1s and 0s. For the sake of simplicity, we give a proof that achieves a weaker dependence on ε_f than in [BRY14a]. Our proof of Lemma 3.1.3 is given in Section 3.4. We note that BRY give a stronger lower bound (without the $\sqrt{\ln k}$) and also bound the variance for the $\lambda = 1$ case. A more careful generalization of BRY which removes the $\sqrt{\ln k}$ would yield an improved loss of $C/k^{1/6}$ instead of $C/k^{1/7}$ in Lemma 3.1.1, but we prefer to give the simpler $C/k^{1/7}$ exposition for the purpose of ease of reading.

Example 3.1.4 (A Two Dimensional Example)**.** Consider the anti-majority function on two dimensions. More precisely, let $f\colon [n]^2 \to \{0,1\}$ be defined as $f(x,y) = 1$ if $x+y \leq n$, and $f(x,y) = 0$ otherwise. We describe two maximum cardinality matchings with vastly different stack sizes. The first matching R matches a point (x,y) with $x+y \leq n$ to the point $(n-y+1, n-x+1)$. For an illustration, see the left matching in Fig. 3.2 for the case $n=5$. Observe that whenever $x+y \leq n$, we have $(n-y+1)+(n-x+1) > n$. The second matching B matches a point (x,y) with $x+y \leq n$ to the point $(x+y, n-x+1)$. Again, observe that $(x+y)+(n-x+1) > n$. For an illustration, see the right blue matching in Fig. 3.2 for the case $n=5$. Note that the stack sizes for the matching R are large; in particular, they are $n-1, n-2, \ldots, 2, 1$ for $n-1$ stacks and 0 for the rest. On the other hand, any stack in B is of size ≤ 1.

Figure 3.2: *Accompanying illustration for Ex. 3.1.4 showing two different maximum cardinality violation matchings for the anti-majority function $f\colon [5]^2 \to \{0,1\}$ which have very different stack sizes. Black (white, resp.) circles represent vertices where $f=1$ ($f=0$, resp.) and connecting lines represent pairs of the matching. Observe that for the left matching, the bottom line and the right-most slice form a stack of size 4 while the right matching has stack sizes all ≤ 1.*

3.2 Domain Reduction: Proof of Lemma 3.1.1

In this section, we use Lemma 3.1.2 and Lemma 3.1.3 to prove Lemma 3.1.1. Recall that $D := [n] \times \left(\prod_{i=2}^{d} [n_i]\right)$ and $D_T := T \times \left(\prod_{i=2}^{d} [n_i]\right)$ denote the original and reduced domains, respectively. Note that $|D_T| = \frac{k}{n}|D|$. Let M be the matching given by Lemma 3.1.2 and consider $\lambda = \lceil 25k^{2/7} \rceil$. Clearly, $\lambda \in [25k^{2/7}, 26k^{2/7}]$.

Thus, by Lemma 3.1.2, we have $\left|\bigcup_{(\ell,S):|M^{(\ell,S)}|\geq 26k^{2/7}} M^{(\ell,S)}\right| \leq \frac{5}{\sqrt{25k^{2/7}}} \cdot |D| = \frac{|D|}{k^{1/7}}$. Let

$$\widehat{M} := M \setminus \left(\bigcup_{(\ell,S):|M^{(\ell,S)}|\geq 26k^{2/7}} M^{(\ell,S)} \right)$$

denote the set of pairs in M which do not belong to stacks larger than $26k^{2/7}$; we therefore have

$$\sum_{\ell \in \mathcal{L}} |\widehat{M}^{(\ell)}| = |\widehat{M}| \geq |M| - \frac{|D|}{k^{1/7}}. \tag{3.1}$$

In this proof, our goal is to *construct* a matching M_T in the violation graph of f_T whose cardinality is sufficiently large. We measure $\mathbb{E}_T\left[|M_T|\right]$ by summing over all lines in $\mathcal{L}$ and applying Lemma 3.1.3 to each. Notice that $\widehat{M}$ is a matching in the violation graph of f which satisfies $|\widehat{M}^{(\ell,S)}| \leq 26k^{2/7}$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$. Thus by Lemma 3.1.3, for every $\ell \in \mathcal{L}$,

$$\mathbb{E}_T\left[|M_T^{(\ell)}|\right] \geq \frac{k}{n} \cdot |\widehat{M}^{(\ell)}| - 3 \cdot (26k^{2/7}) \cdot \sqrt{k \ln k} \geq \frac{k}{n} \cdot |\widehat{M}^{(\ell)}| - 78k^{5/6} \tag{3.2}$$

where we have used $\sqrt{\ln k} < k^{1/3-2/7}$. Now, using eq. (3.1) and eq. (3.2), we can calculate $\mathbb{E}_T\left[|M_T|\right]$. We use the fact that $\{\widehat{M}^{(\ell)}\}_{\ell \in \mathcal{L}}$ is a partition of $\widehat{M}$, apply linearity of expectation and use Lemma 3.1.3 to measure $\mathbb{E}_T[|M_T^{(\ell)}|]$ for each ℓ. Also note that the number of lines is

$|\mathcal{L}| = |D|/n$.

$$
\begin{aligned}
\mathbb{E}_T\left[|M_T|\right] &= \mathbb{E}_T\left[\sum_{\ell\in\mathcal{L}} |M_T^{(\ell)}|\right] = \sum_{\ell\in\mathcal{L}} \mathbb{E}_T\left[|M_T^{(\ell)}|\right] \geq \sum_{\ell\in\mathcal{L}} \left(\frac{k}{n}\cdot|\widehat{M}^{(\ell)}| - 78k^{5/6}\right) \text{ (by eq. (3.2))} \\
&= \left(\frac{k}{n}\cdot\sum_{\ell\in\mathcal{L}}|\widehat{M}^{(\ell)}|\right) - \left(78k^{5/6}\cdot\frac{|D|}{n}\right) \\
&\geq \frac{k}{n}\cdot\left(|M| - \frac{|D|}{k^{1/7}}\right) - \left(78k^{5/6}\cdot\frac{|D|}{n}\right) \text{ (by eq. (3.1))} \\
&= \frac{k}{n}\cdot\left(|M| - \frac{|D|}{k^{1/7}} - \frac{78|D|}{k^{1/6}}\right) \geq \frac{k}{n}\cdot\left(|M| - \frac{C\cdot|D|}{k^{1/7}}\right)
\end{aligned}
\tag{3.3}
$$

for a constant $C > 0$, since $\frac{1}{k^{1/7}}$ dominates $\frac{1}{k^{1/6}}$. eq. (3.3) gives the expected cardinality of our matching after sampling. To recover the distance to monotonicity we simply normalize by the size of the domain. Dividing by $|D_T| = \frac{k}{n}|D|$, we get $\mathbb{E}_T\left[\varepsilon_{f_T}\right] \geq \frac{|M|}{|D|} - \frac{C}{k^{1/7}} = \varepsilon_f - \frac{C}{k^{1/7}}$. This completes the proof of Lemma 3.1.1. □

3.3 Stack Bound: Proof of Lemma 3.1.2

We are given a Boolean function $f\colon D \to \{0,1\}$ where $D = [n]\times\left(\prod_{i=2}^{d}[n_i]\right)$ is a rectangular hypergrid for some $n, n_2, \ldots, n_d \in \mathbb{Z}^+$. Lemma 3.1.2 asserts there is a maximum cardinality matching M such that $\sum_{(\ell,S):|M^{(\ell,S)}|\geq\lambda} |M^{(\ell,S)}| \leq \frac{5}{\sqrt{\lambda}}\cdot|D|$ for all $\lambda\in\mathbb{Z}^+$.

Given a matching M, we consider the vector (or technically, the list) $\Lambda(M)$ indexed by stacks (ℓ, S) with $\Lambda_{\ell,S} := |M^{(\ell,S)}|$, and list these in *non-decreasing* order. Consider the maximum cardinality matching M in the violation graph of f which has the lexicographically largest $\Lambda(M)$. That is, the minimum entry of $\Lambda(M)$ is maximized, and subject to that the second-minimum is maximized and so on. We fix this matching M and claim that it satisfies $\sum_{(\ell,S):|M^{(\ell,S)}|\geq\lambda} |M^{(\ell,S)}| \leq \frac{5}{\sqrt{\lambda}}\cdot|D|$ for all $\lambda\in\mathbb{Z}^+$. Note that the inequality is trivial for $\lambda \leq 100$, since M itself is of size at most $\varepsilon_f|D| \leq \frac{1}{2}|D|$. Thus, in what follows we prove that the inequality is true for an arbitrary, fixed $\lambda > 100$. We first introduce the following notation.

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

Let $V(H)$ denote the set of vertices matched by $\bigcup_{(\ell,S)\in H} M^{(\ell,S)}$. Let B (for blue) be the set of points in $V(H)$ with function value 0, and R (for red) be the set of points in $V(H)$ with function value 1. M induces a perfect matching between B and R, and we wish to prove $|B| = |R| \le \frac{5}{\sqrt{\lambda}} \cdot |D|$. Indeed, define δ to be such that $|B| = \delta|D|$. In the remainder of the proof, we will show that $\delta \le \frac{5}{\sqrt{\lambda}}$.

We make a simple observation that for any fixed line ℓ, there cannot be too many non-low stacks (ℓ, S).

Claim 3.3.1. *For any line ℓ, the number of non-low stacks ℓ participates in is at most $\frac{n}{\lambda-1}$.*

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

We show that if the number of blue points $|B|$ is large ($> 5|D|/\sqrt{\lambda}$), then we will find a line participating in more than $n/(\lambda - 1)$ non-low stacks. To do so, we need to "find" these non-low stacks. We need some more notation to proceed. For a vertex z, we let ℓ_z (S_z, resp.) denote the unique line (slice, resp.) containing z. For each blue point $y \in B$, we define the following interval

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

Note that $\mathcal{I}_y$ is the interval of ℓ_y whose endpoints are given by the projection of (x, y) onto ℓ_y. Armed with this notation, we can find our non-low stacks. Our next claim, which is the heart of the proof and uses the potential function, shows that for every high stack (ℓ, S), we get a bunch of other "non-low" stacks participating with the line ℓ. Refer to Fig. 3.3 for an accompanying illustration of the proof.

Claim 3.3.2. *Given $y \in B$, let $x := M^{-1}(y)$ and suppose $(\ell, S) \in H$ is such that $(x, y) \in M^{(\ell,S)}$ (note that this stack, (ℓ, S), exists by definition of B). Then, for any $z \in \mathcal{I}_y \cap B$, $(\ell, S_z) \notin L$.*

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

Since $z \in B$, it is matched to some $w \in R$. Note $w \prec z \prec y$. Furthermore, the stack $(\ell_w, S_z) \in H$ (by definition of B). Thus, note that if $\ell_w = \ell$ (i.e., $w \in \ell$), then we're done and so in what follows we assume $\ell_w \neq \ell$. By assumption of the claim, $(\ell, S) \in H$. In particular, $x, w, z, y \in V(H)$. Now consider the new matching N which deletes (x, y) and (w, z) and adds (x, z) and (w, y). Note that the cardinality remains the same, i.e. $|N| = |M|$.

We now show that $\Lambda(N)$ is lexicographically bigger than $\Lambda(M)$. To see this, consider the stacks whose sizes have changed from M to N. There are four of them (since we swap two pairs), namely the stacks $(\ell, S), (\ell_w, S_z), (\ell, S_z)$, and (ℓ_w, S). For brevity's sake, let us denote their sizes in M as $\lambda_1, \lambda_2, \lambda_3$, and λ_4, respectively. In N, their sizes are $\lambda_1 - 1, \lambda_2 - 1, \lambda_3 + 1$, and $\lambda_4 + 1$. Note that $\lambda_3 \leq \lambda - 2$ and both λ_1 and λ_2 are $\geq \lambda$. In particular, the "new" size of stack (ℓ, S_z) is still smaller than the "new" sizes of stacks (ℓ, S) and (ℓ_w, S_z). That is, the vector $\Lambda(N)$, even without the increase in λ_4, is lexicographically larger than $\Lambda(M)$. Since increasing the smallest coordinate (among some coordinates) increases the lexicographic order, we get a contradiction to the lexicographic maximality of $\Lambda(M)$. $\square$

The rest of the proof is a (slightly technical) averaging argument to prove that $|B|$ is small. We introduce some more notation to carry this through. For a blue point $y \in B$, let $\beta_y := \frac{|\mathcal{I}_y \cap B|}{|\mathcal{I}_y|}$ denote the fraction of blue points in $\mathcal{I}_y$. For $\alpha \in (0, 1)$, we say that $y \in B$ is

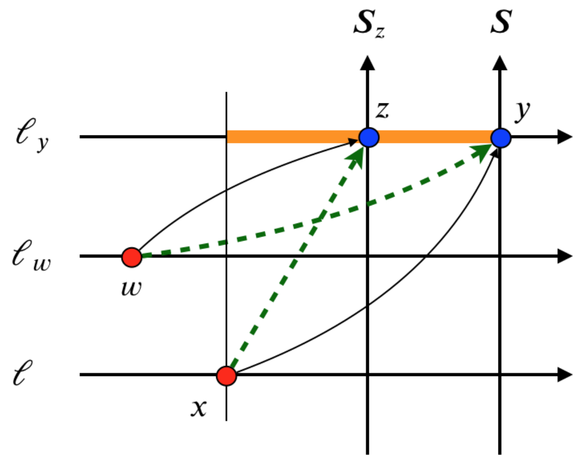


Figure 3.3: *Accompanying illustration for the proof of Claim 3.3.2. The black connecting arrows represent the matching, M, while the dashed green arrows represent the new matching, N. The bold orange segment of ℓ_y is the interval $\mathcal{I}_y$.*

α-rich if $\beta_y \geq \alpha$. A point $x \in R$ is α-rich if its blue partner $y \in B$ (i.e. $(x, y) \in M$) is α-rich. We also call the pair (x, y) an α-rich pair. For what follows, recall that $\delta \in (0, 1)$ is defined such that $|B| = \delta|D|$.

Claim 3.3.3. *At least $\delta|D|/2$ of the points in B are $\delta/4$-rich.*

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

$$\sum_{y \in B^{(\mathsf{poor})}_{\mathsf{min}}} |\mathcal{I}_y| \leq 2 \cdot \left| \bigcup_{y \in B^{(\mathsf{poor})}_{\mathsf{min}}} \mathcal{I}_y \right|. \tag{3.4}$$

Therefore,

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

The first equality follows from the definition of $B^{(\mathsf{poor})}_{\mathsf{min}}$ (taking intersection with B), and the third (strict) inequality follows from the fact that none of these points are $\delta/4$-rich. The fourth inequality is eq. (3.4). This completes the proof. □

A corollary of Claim 3.3.3 is that there are at least $\delta|D|/2$ red points which are $\delta/4$-rich. In particular, there must exist some line ℓ that contains $\ge \delta n/2$ red points in it which are $\delta/4$-rich. Let this line be ℓ and let $R_\ell \subseteq \ell$ be the set of rich red points. Let B_ℓ be their partners in M. Let $\mathcal{S}^\ell = \left\{S \in \mathcal{S} : \exists z \in S \cap \left(\cup_{y\in B^\ell}\mathcal{I}_y \cap B\right)\right\}$ denote the set of slices containing blue points from the collection of rich intervals, $\{\mathcal{I}_y : y \in B^\ell\}$. By Claim 3.3.2, we know that all these stacks are non-low, that is, $(\ell, S) \notin L$ for all $S \in \mathcal{S}^\ell$. We now *lower bound* the cardinality of this set.

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

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

Claim 3.3.4. $\left|\bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y\cap B)\right| \geq \frac{\delta}{8}\left|\bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y)\right|$.

Proof. As in the proof of Claim 3.3.2, let $B^\ell_{\mathsf{min}} \subseteq B^\ell$ be a minimal cardinality subset of B^ℓ such that $\bigcup_{y\in B^\ell_{\mathsf{min}}}\mathbf{o}(\mathcal{I}_y) = \bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y)$. For any $y \in B$, y belongs to *at most* two intervals from B^ℓ_{min}.

$$\begin{aligned}\left|\bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y\cap B)\right| &= \left|\bigcup_{y\in B^\ell_{\mathsf{min}}}\mathbf{o}(\mathcal{I}_y\cap B)\right| \geq \frac{1}{2}\sum_{y\in B^\ell_{\mathsf{min}}}|\mathbf{o}(\mathcal{I}_y\cap B)| \\ &\geq \frac{\delta}{8}\sum_{y\in B^\ell_{\mathsf{min}}}|\mathbf{o}(\mathcal{I}_y)| \geq \frac{\delta}{8}\left|\bigcup_{y\in B^\ell_{\mathsf{min}}}\mathbf{o}(\mathcal{I}_y)\right| = \frac{\delta}{8}\left|\bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y)\right|. \qquad \square\end{aligned}$$

Now importantly, $|\mathbf{o}(R^\ell)| = |R^\ell| \geq \frac{\delta}{2}\cdot n$ since the 1-coordinates of elements of R^ℓ are distinct (since R^ℓ is contained on a single line). Moreover, by definition of $\mathcal{I}_y$, $\mathbf{o}(R^\ell) \subseteq \bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y)$ and so $\left|\bigcup_{y\in B^\ell}\mathbf{o}(\mathcal{I}_y)\right| \geq |\mathbf{o}(R^\ell)| \geq \frac{\delta}{2}\cdot n$. Finally, combining this with Claim 3.3.4, we get

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

Therefore, ℓ participates in at least $\frac{\delta^2}{16}\cdot n$ non-low stacks. Thus, by Claim 3.3.1, $\frac{\delta^2}{16}\cdot n \leq \frac{n}{\lambda-1}$ and so $\delta \leq \frac{4}{\sqrt{\lambda-1}}$. Since $\lambda > 100$, we conclude that $\delta \leq \frac{5}{\sqrt{\lambda}}$. This concludes the proof of Lemma 3.1.2. $\square$

3.4 Line Sampling: Proof of Lemma 3.1.3

We recall the lemma for ease of reading. Given a line $\ell \in \mathcal{L}$, we have defined $M^{(\ell)} := \{(x,y) \in M : x \in \ell\}$. Given a stack S, we have defined $M^{(\ell,S)} := \{(x,y) \in M^{(\ell)} : y \in S\}$. Given a multi-set $T \subseteq [n]$, recall $M_T^{(\ell)}$ is a maximum cardinality matching of violations (x, y) such that (a) x and y are both matched by $M^{(\ell)}$, and (b) x_1 and y_1 both lie in T. Given $\lambda \in \mathbb{Z}^+$ such that $|M^{(\ell,S)}| \leq \lambda$ for all $\ell \in \mathcal{L}$ and $S \in \mathcal{S}$, the line sampling lemma (Lemma 3.1.3) states

$$\mathbb{E}_T\left[|M_T^{(\ell)}|\right] \geq \frac{k}{n} \cdot |M^{(\ell)}| - 3\lambda\sqrt{k \ln k}. \tag{3.5}$$

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

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

Claim 3.4.1. *We make a few observations.*

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

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

Define $V^+ := \{i : w^+(i) > 0\}$ and $V^- := \{j : w^-(j) > 0\}$. Given a multiset $T \subseteq [n]$, denote $V_T^+ := V^+ \cap T$ and $V_T^- := V^- \cap T$. Also, define the bipartite graph $G_T := (V_T^+, V_T^-, E_T)$ where $(i, j) \in E_T$ iff $i \leq j$. A w-matching A in G_T is a subset of edges of E_T such that every vertex $i \in V_T^+$ has at most $w^+(i)$ edges of A incident on it, and every vertex $j \in V_T^-$ has at most $w^-(j)$ edges of A incident on it. Let $\nu(G_T)$ denote the size of the largest w-matching in G_T.

Lemma 3.4.2. *For any multiset $T \subseteq [n]$ and any w-matching $A \subseteq E_T$ in G_T, we have $|M_T^{(\ell)}| \geq |A|$. In particular, $\mathbb{E}_T\left[|M_T^{(\ell)}|\right] \geq \mathbb{E}_T\left[\nu(G_T)\right]$.*

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

Lemma 3.4.3. *For any $T \subseteq [n]$, we have*

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

Proof. By Hall's theorem, the maximum w-matching in G_T is given by the total weight on the V_T^- side, that is, $\sum_{j \in T} w^-(j)$, minus the total *deficit*

$$\delta(T) := \max_{S \subseteq V_T^-} \left(\sum_{s \in S} w^-(s) - \sum_{s \in \Gamma_T(S)} w^+(s) \right)$$

where for $S \subseteq V_T^-$, $\Gamma_T(S) \subseteq V_T^+$ is the neighborhood of S in G_T. Consider such a maximizer S, and let t be the largest index present in S. Then note that $\sum_{s \in \Gamma_T(S)} w^+(s)$ is precisely $\sum_{s \in T : s \leq t} w^+(s)$. Furthermore note that adding any $s \leq t$ from V_T^- won't increase $|\Gamma_T(S)|$.

Thus, given that the largest index present in S is t, we get that $\delta(T)$ is precisely the summation in the second term of the RHS. $\delta(T)$ is maximized by choosing the t which maximizes the summation. $\square$

Next, we bound the expectation of the RHS in Lemma 3.4.3. Recall that $T := \{s_1, \ldots, s_k\}$ is a multiset where each s_i is u.a.r. picked from $[n]$. For the first term, we have

$$\mathbb{E}_T\left[\sum_{j\in T} w^-(j)\right] = \sum_{i=1}^{k}\sum_{j=1}^{n} \mathbb{P}[s_i = j]\cdot w^-(j) = \frac{k}{n}\cdot\sum_{j=1}^{n} w^-(j) = \frac{k}{n}\cdot|M^{(\ell)}|. \tag{3.6}$$

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

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

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

$$\mathbb{P}\left[Z_t > \mathbb{E}[Z_t] + a\right] \le 2\exp\left(\frac{-a^2}{2k\lambda^2}\right). \tag{3.7}$$

Now we use Claim 3.4.1, part (3) to deduce that

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

since $\mathbb{P}[s_i = s] = 1/n$. Therefore, the RHS of eq. (3.7) is an upper-bound on $\mathbb{P}[Z_t \ge a]$. In particular, invoking $a := 2\lambda\sqrt{k \ln k}$ and applying a union bound, we get

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

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

$$\mathbb{E}_T\left[\max_{t\in T} \sum_{s\in T: s\leq t} \left(w^-(s) - w^+(s)\right)\right] \leq \lambda k \cdot \mathbb{P}\left[\max_{t\in T} Z_t > a\right] + a \leq \lambda + a \leq 3\lambda\sqrt{k\ln k}. \quad (3.8)$$

Lemma 3.1.3 follows from Lemma 3.4.2, Lemma 3.4.3, eq. (3.6), and eq. (3.8).

3.5 The Continuous Domain

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

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

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

Definition 3.5.1. *Given a product measure $\mu = \prod_i \mu_i$ over $\mathbb{R}^d$, we say $E \subseteq \mathbb{R}^d$ is Lebesgue-measurable with respect to μ if for any $\varepsilon > 0$, there exists an open set $U \supseteq E$ such that $\mu_*(U \setminus E) < \varepsilon$. If this holds, then the μ-measure of E is defined as $\mu(E) := \mu_*(E)$.*

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

We are now ready to define the notion of distance between two functions. In Section 3.5.3, we prove that all monotone Boolean functions are measurable (Theorem 3.5.6) with respect to μ. Also, measurability is closed under basic set operations and thus the following notion of distance to monotonicity is well-defined. Thus, the monotonicity testing problem for functions $f \colon \mathbb{R}^d \to \{0,1\}$ is also well-defined.

Definition 3.5.2 (Distance to Monotonicity)**.** *Fix a product measure $\mathcal{D}$ on $\mathbb{R}^d$. We define the distance between two measurable functions $f, g \colon \mathbb{R}^d \to \{0,1\}$ with respect to μ, as*

$$\mathsf{dist}_{\mathcal{D}}(f,g) := \mu\left(\left\{z \in \mathbb{R}^d \colon f(z) \neq g(z)\right\}\right) = \mu\left(f \Delta g\right). \tag{3.9}$$

The distance to monotonicity of f w.r.t. $\mathcal{D}$ is defined as

$$\varepsilon_{f,\mathcal{D}} := \inf_{g \in \mathcal{M}} \mathsf{dist}_{\mathcal{D}}(f,g) = \inf_{g \in \mathcal{M}} \mu\left(f \Delta g\right) \tag{3.10}$$

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

3.5.1 Approximating Measurable Sets by Grids

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

Lemma 3.5.3. *Given any probability measure $\mathcal{D}$ over $\mathbb{R}$, and any $N \in \mathbb{N}$, there exists an almost partition of $\mathbb{R}$ into N intervals $\boldsymbol{I}_N = \{\mathcal{I}_1, \ldots, \mathcal{I}_N\}$ of equal μ-measure. That is, for each $j \in [N]$, $\mathbb{P}_{x \sim \mathcal{D}}[x \in \mathcal{I}_j] = \frac{1}{N}$. Furthermore, for any $k \in \mathbb{N}$, $\boldsymbol{I}_{kN}$ is a refinement of $\boldsymbol{I}_N$.*

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

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

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

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

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

Lemma 3.5.4. *For any measurable set E and any $\alpha > 0$, there exists $N = N(E, \alpha) \in \mathbb{N}$ such that there is a collection $\boldsymbol{B} \subseteq \boldsymbol{G}_N$ satisfying $\mu(E \ \Delta \ \bigcup_{B \in \boldsymbol{B}} B) \leq \alpha$.*

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

$$\mu\left(E \Delta \bigcup_{B \in \boldsymbol{B}} B\right) \leq \mu\left(E \Delta \bigcup_{r=1}^{m} B_r\right) + \mu\left(\bigcup_{r=1}^{m} B_r \setminus \bigcup_{B \in \boldsymbol{B}} B\right) \leq \alpha/2 + \sum_{r=1}^{m} \mu\left(B_r \setminus \bigcup_{B \in \boldsymbol{B}} B\right) \tag{3.11}$$

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

Let $B_r := \prod_{i=1}^{d} [a_i, b_i]$ denote an arbitrary box from $\{B_1, \ldots, B_m\}$ and let $\delta_i := \mu_i([a_i, b_i])$. Observe that the interval $[a_i, b_i]$ contains exactly $\lfloor \delta_i N \rfloor$ contiguous intervals from the almost partition $\{\mathcal{I}_j^{(i)} : j \in [N]\}$ of $\mathbb{R}$. Let $\boldsymbol{I}_i$ denote the set of such intervals. Thus, $\mu_i([a_i, b_i] \setminus \bigcup_{I \in \boldsymbol{I}_i} I) \leq \delta_i - (1/N)(\lfloor \delta_i N \rfloor) \leq \delta_i - (1/N)(\delta_i N - 1) = 1/N$. Thus, the total measure of B_r we discard is $\mu(B_r \setminus \bigcup_{B \in \boldsymbol{B}} B) \leq \prod_i \delta_i - \prod_i (\delta_i - 1/N)$. This quantity is maximized when the δ_i's are maximized; since $\delta_i \leq 1$ (each μ_i is a probability measure), we get that $\mu(B_r \setminus \bigcup_{B \in \boldsymbol{B}} B) \leq 1 - (1 - 1/N)^d \leq \frac{d}{N}$.

Finally, plugging this into eq. (3.11), we get $\mu(E \Delta \bigcup_{B \in \boldsymbol{B}} B) \leq \alpha/2 + m \cdot \frac{d}{N} \leq \alpha$, since $N \geq 2md/\alpha$. □

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

Lemma 3.5.5 (Discretization Lemma)**.** *Given a measurable function $f \colon \mathbb{R}^d \to \{0, 1\}$ and $\delta > 0$, there exists $N := N(f, \delta) \in \mathbb{N}$, and a function $f^{\mathsf{disc}} : [N]^d \to \{0, 1\}$, such that*

$\mathbb{P}_{x\sim\mathcal{D}}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] \leq \delta$.

Proof. By assumption, f and $\overline{f}$ are measurable sets. By Lemma 3.5.4, there exists some N_1 and a collection of boxes $\boldsymbol{Z}_1 \subseteq \boldsymbol{G}_{N_1}$ such that $\mu(f \Delta \bigcup_{B\in\boldsymbol{Z}_1} B) \leq \delta/6$. (An analogous statement holds for $\overline{f}$, with some N_0 and a collection $\boldsymbol{Z}_0$.) Since Lemma 3.5.4 also holds for any refinement of the relevant grid, let us set $N = N_0 N_1$. Abusing notation, we have two collections $\boldsymbol{Z}_0, \boldsymbol{Z}_1 \subseteq \boldsymbol{G}_N$ such that $\mu(f\Delta \bigcup_{B\in\boldsymbol{Z}_1} B) \leq \delta/6$ and $\mu(\overline{f}\Delta \bigcup_{B\in\boldsymbol{Z}_0} B) \leq \delta/6$.

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

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

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

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

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

3.5.2 Domain Reduction for Continuous Space: Proof of Theorem 3.0.2

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

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

for some universal constant $C' > 0$. Set $\delta \leq k^{-d} \cdot \frac{C \cdot d}{k^{1/7}}$, where C is the universal constant in Theorem 3.0.1. Applying Lemma 3.5.5 to f with this δ, we know there exists $N > 0$ and $f^{\mathsf{disc}} \colon [N]^d \to \{0, 1\}$, such that $\mathbb{P}_{x\sim\mathcal{D}}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] \leq \delta$. Given a random $\boldsymbol{T}$ sampled as described above, define $\widehat{\boldsymbol{T}} := \{\mathsf{box}_N(x) \in [N]^d : x \in \boldsymbol{T}\}$. Observe that (a) $\widehat{\boldsymbol{T}}$ is a $[k]^d$ sub-hypergrid in $[N]^d$ which (b) can be equivalently defined as $\widehat{\boldsymbol{T}} = \widehat{T}_1 \times \cdots \times \widehat{T}_d$ where each $\widehat{T}_i$ is formed by taking k i.i.d. uniform samples from $[N]$. This is by construction of the partition $\{\mathsf{box}_z : z \in [N]^d\}$ and by definition of $\mathsf{box}_N(x)$. Theorem 3.0.1 and the observations above imply

$$\mathbb{E}_{\widehat{\boldsymbol{T}}}\left[\varepsilon_{f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}}\right] \geq \varepsilon_{f^{\mathsf{disc}}} - \frac{C \cdot d}{k^{1/7}} \tag{3.12}$$

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

By a union bound over the k^d samples,

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

since each $x \in \boldsymbol{T}$ has the same distribution as $x \sim \mathcal{D}$, and $\mathbb{P}_{x\sim\mathcal{D}}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] \leq \delta$. Thus, we get $\mathbb{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq (1 - \delta')\mathbb{E}_{\widehat{\boldsymbol{T}}}\left[\varepsilon_{f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}}\right] - \delta'$, since in the case $f_{\boldsymbol{T}} \neq f^{\mathsf{disc}}_{\widehat{\boldsymbol{T}}}$, the difference in their distance to monotonicity is at most 1. Substituting in eq. (3.12), we get

$$\mathbb{E}_{\boldsymbol{T}}\left[\varepsilon_{f_{\boldsymbol{T}}}\right] \geq (1 - \delta') \cdot \left(\varepsilon_{f^{\mathsf{disc}}} - \frac{C \cdot d}{k^{1/7}}\right) - \delta' \geq \varepsilon_{f^{\mathsf{disc}}} - \frac{3C \cdot d}{k^{1/7}} \tag{3.13}$$

by definition of δ'.

Now, let $g\colon [N]^d \to \{0,1\}$ be any monotone function satisfying $d(f^{\mathsf{disc}}, g) = \varepsilon_{f^{\mathsf{disc}}}$. Define the monotone function $\hat{f}(x) = g(\mathsf{box}_N(x))$ for all $x \in \mathbb{R}^d$. Note that $\varepsilon_f \leq \mathsf{dist}(f, \hat{f}) \leq \mathbb{P}_{x\sim\mathcal{D}}[f(x) \neq f^{\mathsf{disc}}(\mathsf{box}_N(x))] + \mathsf{dist}(f^{\mathsf{disc}}, g) \leq \delta + \varepsilon_{f^{\mathsf{disc}}}$. This, in turn, implies $\varepsilon_{f^{\mathsf{disc}}} \geq \varepsilon_f - \delta \geq \varepsilon_f - \frac{C\cdot d}{k^{1/7}}$. Substituting in eq. (3.13), we get

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

which proves the theorem. $\square$

3.5.3 Measurability of Monotone Functions

Theorem 3.5.6. *Monotone functions $f\colon \mathbb{R}^d \to \{0,1\}$ are measurable w.r.t. product measures $\mu = \prod_{i=1}^d \mu_i$.*

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

Now for the induction. Choose any $\varepsilon > 0$. We will construct an open set $\mathcal{O}$ such that $\mu_*(\mathcal{O} \setminus f) \leq 8\varepsilon$. Consider the first dimension, and the corresponding measure μ_1. We use μ_{-1} for the $(d-1)$-dimensional product measure in the remaining dimensions. (We use $\mu_{-1,*}$ for the $(d-1)$-dimensional exterior measure.) As shown in Lemma 3.5.3, there is an almost partition of $\mathbb{R}$ into $N = \lceil 1/\varepsilon^2 \rceil$ closed intervals such that each interval has μ_1-measure at most ε^2. Let these intervals be $I_1, I_2, I_3, \ldots, I_N$. We will consider the set of intervals $\boldsymbol{I} = \{I_1 \cup I_2, I_2 \cup I_3, \ldots, I_{N-1} \cup I_N\}$ (let us treat these as open intervals). Observe that $\cup_{I\in\boldsymbol{I}} I = \mathbb{R}$, and $\mu_1(I) \leq 2\varepsilon^2$ for all $I \in \boldsymbol{I}$.

For any $x \in \mathbb{R}$, let S_x be the subset of f with first coordinate x. We will treat S_x as a subset of $\mathbb{R}^{d-1}$ and use $\{x\} \times S_x$ to denote the corresponding subset of $\mathbb{R}^d$. By monotonicity, $\forall x < y$, $S_x \subseteq S_y$. By induction, each set S_x is measurable in $\mathbb{R}^{d-1}$ and thus there exists an open set $\mathcal{O}_x \subseteq \mathbb{R}^{d-1}$ such that $\mu_{-1,*}(\mathcal{O}_x \setminus S_x) \leq \varepsilon$. Define the function $h\colon \mathbb{R} \to [0,1]$ such that $h(x)$ is the measure of S_x (in $\mathbb{R}^{d-1}$). Crucially, h is monotone because f is monotone.

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

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

$$\begin{aligned}
\mu_*(\mathcal{O} \setminus f) &\leq \mu_*\left(\bigcup_{J\in\boldsymbol{J}} J \times \mathbb{R}^{d-1}\right) + \mu_*\left(\bigcup_{I\in\boldsymbol{I}\setminus\boldsymbol{J}} \mathcal{O}_I \setminus f\right) \\
&\leq \sum_{J\in\boldsymbol{J}} \mu_1(J) + \sum_{I\in\boldsymbol{I}\setminus\boldsymbol{J}} \mu_*(\mathcal{O}_I \setminus f) \leq 2\varepsilon^2|\boldsymbol{J}| + \sum_{I\in\boldsymbol{I}\setminus\boldsymbol{J}} \mu_*(\mathcal{O}_I \setminus f). \qquad (3.14)
\end{aligned}$$

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

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

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

All in all, we can upper bound the expression in eq. (3.14) by $2\varepsilon^2(2/\varepsilon) + 4\varepsilon = 8\varepsilon$. $\square$

3.6 Lower Bound for Domain Reduction

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

Theorem 3.6.1 (Lower Bound for Domain Reduction)**.** *There exists a function $f\colon [n]^d \to \{0,1\}$ with distance to monotonicity $\varepsilon_f = \Omega(1)$, for which $\mathbb{E}_{\boldsymbol{T}}[\varepsilon_{f_{\boldsymbol{T}}}] \leq O(k^2/d)$. In particular, $k = \Omega(\sqrt{d})$ samples in each dimension is necessary to preserve distance to monotonicity.*

3.6.1 Proof of Theorem 3.6.1

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

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

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

Claim 3.6.2. *The distance to monotonicity of* $\mathsf{Centrist}$ *is* $\Omega(1)$.

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

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

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

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

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

Lemma 3.6.3. *Let $k \in \mathbb{Z}^+$ be any positive integer. If $\boldsymbol{T} := T_1 \times \cdots \times T_d$ is a randomly chosen hypergrid, where for each $i \in [d]$, T_i is a set formed by taking k i.i.d. samples from the uniform distribution on $[0,1]$, then with probability $> 1 - 4k^2/d$, $\mathsf{Centrist}_{\boldsymbol{T}}$ is a monotone function.*

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

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

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

- The vote does not change. Then the function value does not change.

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

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

Theorem 3.6.1 follows from Claim 3.6.2 and Lemma 3.6.3.

CHAPTER 4

The Directed Talagrand Inequality for Hypergrids

In this chapter we prove a generalization of the celebrated directed Talagrand inequality of [KMS18] to the hypergrid. The results in this chapter were originally published in [BCS23b]. We refer the reader to Section 2.3 for a discussion on isoperimetric inequalities over hypergrids. Our results use the following notion of influence for a function $f\colon [n]^d \to \{0,1\}$.

Definition 4.0.1 (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 [KMS18] result.

Theorem 4.0.2 (Directed Talagrand Inequality on the Hypergrid)**.** *Let $f : [n]^d \to \{0,1\}$ be ε-far from monotone.*

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

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 4.0.3 (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 4.0.4 (Robust Directed Talagrand Inequality on the Hypergrid)**.** *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.*

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

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 [KMS18] 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 4.0.2) 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.

4.1 Challenges

We explain the challenges faced in proving Theorem 4.0.4. The KMS proof of the directed Talagrand inequality for the hypercube is a tour-de-force [KMS18], 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 $T(f) = \mathbb{E}_{\mathbf{x}}[\sqrt{I_f^-(\mathbf{x})}]$ denote the hypercube directed Talagrand objective for a $f : \{0,1\}^d \to \{0,1\}$. To lower bound $T(f)$, [KMS18] 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 [KMS18] 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 [KMS18] (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 $T(g) = \Omega(\varepsilon_g)$ (the distance of g to monotonicity). But how does one handle ε_g, or g more generally? This is done by relating splitting to the 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. [KMS18] 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 [KMS18] 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 this via what we call "tracker functions" which are n^d different Boolean functions tracking the changes in f. We discuss this more in Section 4.2.

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 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 the 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. 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.

4.2 Main Ideas

We sketch some key ideas needed to prove Theorem 4.0.4 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 [KMS18], 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.

4.2.1 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 f on 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 4.2.1 (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*

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

We provide an illustration of this definition in Figure 4.1.

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 4.0.4. 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 to robust undirected Talagrand inequalities on hypercubes. This is the main technical contribution of

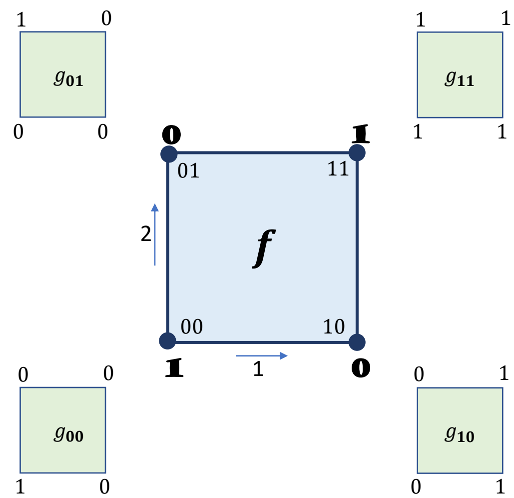


Figure 4.1: *The blue function $f : [n]^d \to \{0,1\}$ is defined in the middle using bold, gothic characters. We have $d = 2$ and $n = 2$. For each of the 4 points of this square, we have four different $g_{\mathbf{x}} : \{0,1\}^2 \to \{0,1\}$ and they are described in the four green squares. For any $S \subseteq \{1,2\}$, 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.*

our paper. Let us define the (colored) influences of these $g_{\mathbf{x}}$ functions.

Definition 4.2.2 (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 jth coordinate is defined as*

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

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 $I_{g_{\mathbf{x}}}^{=j}(S)$ and $I_{g_{\mathbf{x}}}^{=j}(S \oplus j)$. We define a robust, colored version of these influences.

Definition 4.2.3 (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 jth coordinate is defined as*

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

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 <u>exists</u> a coloring $\xi_{\mathbf{x}} : E(2^{[d]}) \to \{0,1\}$ of the Boolean *hypercube* edges, such that

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

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

Why is a statement like eq. (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 4.3.8) to get an upper bound on the variance.

Corollary 4.2.4 (Corollary of Theorem 1.8 in [KMS18])**.** *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*

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

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.

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

We now see that eq. (H1), Corollary 4.2.4, and eq. (H2) together implies Theorem 4.0.4 (indeed without the $\log n$).

4.2.2 High level description of our approaches

Addressing the $\approx$ in eq. (H1) via semisorting. As stated, we do not know if eq. (H1) is true. However, we establish eq. (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 4.0.4 on semi-sorted functions is already a generalization of the [KMS18] result. Theorem 4.4.2 is the formal restatement of eq. (H1).

We reduce Theorem 4.0.4 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 4.0.4 on the semisorted version of f implies Theorem 4.0.4 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 4.4.1. 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 4.0.4 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 4.0.4 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 4.4.

The real work happens in proving Theorem 4.4.3, that is, eq. (H1) for semisorted functions.

Approach to proving eq. (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 $\mathbb{E}_{\mathbf{x}\in[n]^d}[\sqrt{\Phi_{f,\chi}(\mathbf{x})}]$ which is the LHS of eq. (H1). At the end of round d it takes the value $\mathbb{E}_{\mathbf{x}\in[n]^d}\mathbb{E}_{S\subseteq[d]}[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}]$ which is the RHS of eq. (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

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

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 $\Lambda_i := \mathbb{E}_{S \subseteq [i]}[R_i(S)]$.

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 eq. (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 eq. (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, $\Lambda_i = (\mathbb{E}_{S \subseteq [i-1]}[R_i(S) + R_i(S + i)])/2$, while $\Lambda_{i-1} = \mathbb{E}_{S \subseteq [i-1]}[R_{i-1}(S)]$. So, if we prove that $R_{i-1}(S)$ is at least 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 eq. (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, $R_{i-1}(S) \geq R_i(S)$.

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, the 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, eq. (Hybrid) is replaced by the actual quantity eq. (Colorful Hybrid) described in Section 4.6. 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 4.4.1 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 4.6. The most technical part of the paper is in Section 4.7, which proves that the potential decreases in each round.

Addressing the $\approx$ in eq. (H2) via random sorts. To finally complete the argument, we need eq. (H2) that relates the average variance of the $g_{\mathbf{x}}$ functions to the distance to monotonicity of f. As discussed earlier, eq. (H2) is false, even for the case of hypercubes. Nevertheless, one can use eq. (H1) and Corollary 4.2.4 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 4.5.2) is that $\mathbb{E}_{\mathbf{x}\in[n]^d}[\mathrm{var}(g_{\mathbf{x}})]$ is roughly $\mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f]$ 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 the 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 eq. (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 eq. (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 4.0.4. We re-assert that the main ideas are already present in [KMS18, PRW22]. However, we require a more general presentation to make things work for hypergrids. These details can be found in Section 4.5.

4.3 Preliminaries

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

Definition 4.3.1. *Consider a Boolean function on the line* $h : [n] \to \{0, 1\}$*. The sort operator* $\texttt{sort}()$ *is defined as follows.*

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

Thus, the sort operator "moves" the values on a line to ensure that it is sorted. Note that $\texttt{sort}(h)$ 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 4.3.2. *Let* i *be a dimension and* $f : [n]^d \to \{0, 1\}$*. The sort operator for dimension* i*,* $\texttt{sort}_i()$*, is defined as follows. For every* i*-line* ℓ*,* $\texttt{sort}_i(f)|_\ell = \texttt{sort}(f|_\ell)$*.*

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

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

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 4.3.3. *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 4.5, where we apply our main isoperimetric theorem on random restrictions.We provide a proof for completeness.

Claim 4.3.4. *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

$\|\texttt{sort}(h)\|_1 = \|h\|_1$.

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

□

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 [KMS18] (Lemma 3.5). The proofs goes through word-for-word applied to hypergrids.

Claim 4.3.5. *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 4.3.3. Let h be the monotone function closest to f, that is, $\varepsilon_f = \Delta(f, h)$. So,

$$\Delta(f, [d]\circ f) \underbrace{\leq}_{\text{triangle ineq}} \Delta(f, h) + \Delta([d]\circ f, h) \underbrace{=}_{\text{since } h=[d]\circ h} \Delta(f, h) + \underbrace{\Delta([d] \circ f, [d] \circ h)}_{\leq \Delta(f,h) \text{ by Claim 4.3.4}} \leq 2\Delta(f, h) = 2\varepsilon_f$$

□

We provide one more simple claim about the sort operator that will be used throughout

Section 4.7. Given $h, h' \colon [n] \to \{0,1\}$, define

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

Claim 4.3.6. *Let $h, h' \colon [n] \to \{0,1\}$ be any two functions. Then, $\Delta^-(\texttt{sort}(h), \texttt{sort}(h')) \leq \Delta^-(h,h')$.*

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

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

□

4.3.1 Colorful Influences and the Talagrand Objective

We will need undirected, colorful Talagrand inequalities for proving Theorem 4.0.4. For the sake of completeness, we explicitly define the undirected colored influence.

Definition 4.3.7. *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 $\mathbb{E}_{\mathbf{z}}[\sqrt{I_g(\mathbf{z})}] = \Omega(\mathrm{var}(g))$ [Tal93]. The robust/colored version proven by KMS asserts this to be true for arbitrary colored influences.

Theorem 4.3.8 (Paraphrasing Theorem 1.8 of [KMS18]). *(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,*

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

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

Definition 4.3.9 (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*

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

where, $\Phi_{f,\chi}$ *is defined in Definition 4.0.3.*

4.3.2 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:

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

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 $i \in [d]$ and 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) eq. (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 $1 \leq k \leq t$, $\sum_{i \leq k} (\mathbf{a})^{\downarrow}_i \geq \sum_{i \leq k} (\mathbf{b})^{\downarrow}_i$.

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 4.3.10 (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 4.3.11. *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 $\vec{U}$, and let them comprise $T \subseteq [t]$ where $|T| = \tau$. Consider the τ-dimensional vectors $\mathbf{w}_i[T]$ where we restrict our attention to only these coordinates. Let $\vec{S'}$ be the τ-dimensional vector formed by the sum of the sorted versions $(\mathbf{w}_i[T])^{\downarrow}$. 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 $\sum_i(j\text{th max of } \mathbf{w}_i[T])$ and $\vec{S}_j$ equals $\sum_i(j\text{th max of } \mathbf{w}_i)$. Thus, $\vec{S}_j \geq \vec{S'}_j$, proving that $\sum_{j=1}^{\tau} \vec{S}_j \geq \sum_{j=1}^{\tau} \vec{V}_j$. $\square$

4.4 Semisorting and Reduction to Semisorted Functions

As we mentioned earlier when we stated eq. (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 4.0.4 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 4.4.1.

Lemma 4.4.1 (Semisorting only decreases T_Φ.). *Let f be any hypergrid function and let χ be any bicoloring 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 eq. (H1) holds. In particular, we prove this theorem.

Theorem 4.4.2 (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*

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

We can use the above theorem to get set the intuition behind eq. (H2) correct, and prove Theorem 4.0.4 for semisorted functions. We state this below, but we defer the proof of this to Section 4.5. 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 [KMS18, PRW22] for the hypercube case.

Theorem 4.4.3 (Theorem 4.0.4 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*

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

Lemma 4.4.1 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 4.4.3 to prove Theorem 4.0.4.

Sequence of Semisorted Functions and Reduction to the Semisorted Case. We now describe a semi-sorting 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 $(2^i)^d$ 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.

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

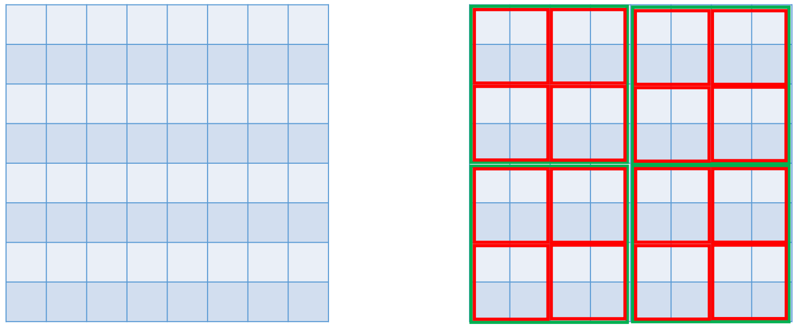


Figure 4.2: *In the figure, we see an example with $d = 2$ and $n = 8 = 2^3$. There are 2^2 many 4×4 green (hyper)-grids, and 4^2 many 2×2 red squares.*

singleton set containing only one hypergrid, $[n]^d$. Define the function $f_1 : [n]^d \to \{0, 1\}$ as follows: consider every hypergrid[1] 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 4.3 is an illustration for $d = 2$ and $k = 3$, i.e. $n = 8$.

Claim 4.4.4. *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 $\Delta(f_0, f_k) \geq \varepsilon_f$. $\square$

Proof of Theorem 4.0.4. We now show how Theorem 4.0.4 follows from Lemma 4.4.1 and Theorem 4.4.3 via an averaging argument. We fix the j as in Claim 4.4.4. By Lemma 4.4.1 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 4.4.3 (on the coloring χ') we know that for all $H \in \mathcal{H}_{k-j-1}$,

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

[1]these will be hypercubes

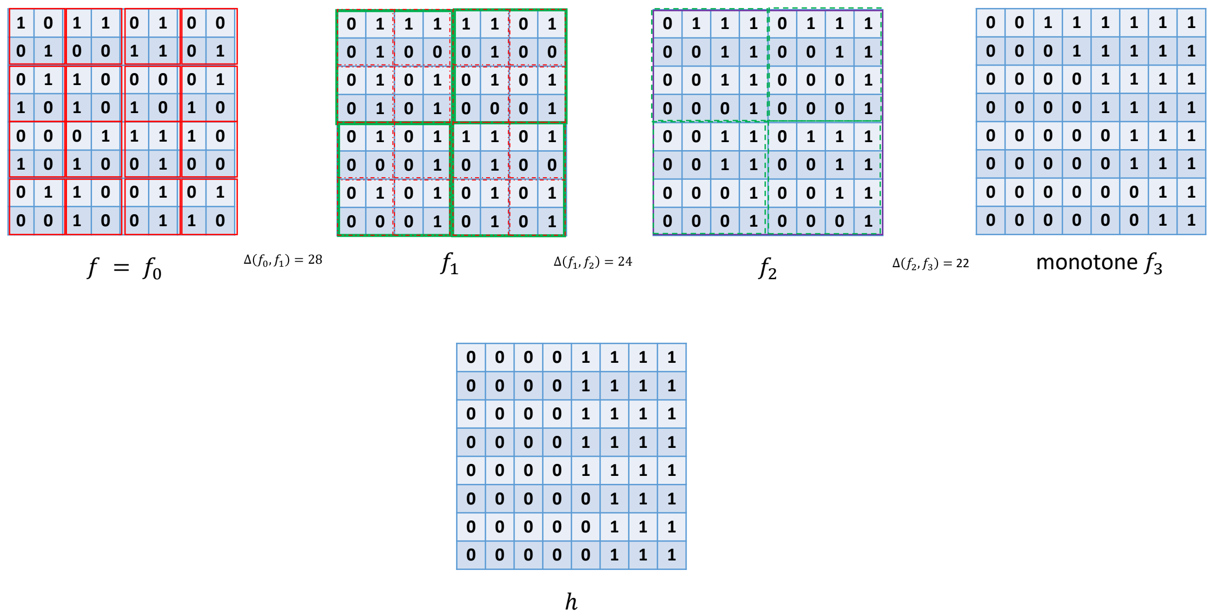


Figure 4.3: *The function $f = f_0$ is described to the left, and then one obtains f_1, f_2 and f_3. The function h which is obtained doing sort on the whole of f is described below. Note $h \neq f_3$.*

By Claim 4.3.5, 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 4.0.4. $\square$

4.4.1 Semisorting only decreases the Talagrand objective: Proof of Lemma 4.4.1

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 4.4 for an illustration.

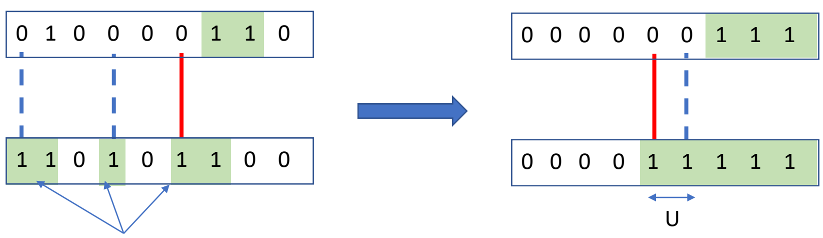


Figure 4.4: *We are considering only the interval I. The line below is ℓ and the line above is ℓ'. The green shaded zones correspond to where the function evaluates to 1s. The situation to the right is after sorting. Only the violating edges are marked. On the left, the red solid edges are colored $\chi(e) = 1$ 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.*

- 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 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 4.5 for an illustration.

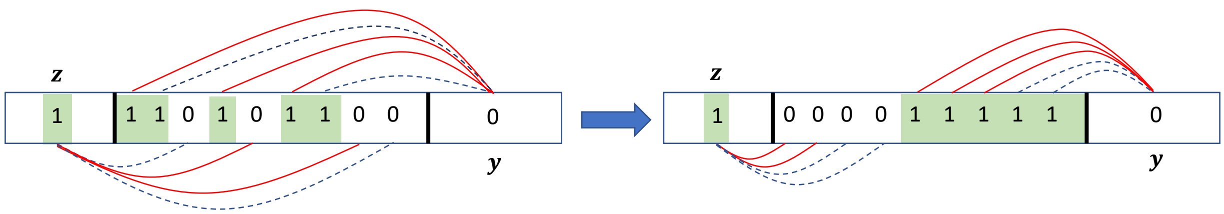


Figure 4.5: *The two vertical black lines demarcate I. The green shaded zones correspond to where the function evaluates to 1s. The situation to the right is after sorting. $\mathbf{y}$ is a point with $f(\mathbf{y}) = 0$ to the right of I; $\mathbf{z}$ is a point with $f(\mathbf{z}) = 1$ to the left 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

$$\left\|\overrightarrow{\Phi_\chi(f)}\right\|_{1/2} \geq \left\|\overrightarrow{\Phi_{\chi'}(h)}\right\|_{1/2} \tag{4.1}$$

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

$$\left\|\overrightarrow{\Phi^{(1)}_\chi(f)}\right\|_{1/2} \geq \left\|\overrightarrow{\Phi^{(1)}_{\chi'}(h)}\right\|_{1/2} \quad \text{and} \quad \left\|\overrightarrow{\Phi^{(0)}_\chi(f)}\right\|_{1/2} \geq \left\|\overrightarrow{\Phi^{(0)}_{\chi'}(h)}\right\|_{1/2} \tag{4.2}$$

and this will prove eq. (4.1). 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 4.3.11.

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 4.4.5. *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 the following.

Observation 4.4.6. *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 4.4.7. *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'. $\square$

What we need is the following corollary.

$$\text{For any } \ell' = \ell + a\mathbf{e}_j, \quad \mathbf{v}_{\ell'} \ \leq_{\text{coor}} (\mathbf{u}_{\ell'})^{\downarrow} \tag{4.3}$$

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

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

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

Proof. Follows from eq. (4.3), and the defintions of $\mathbf{z}_j$ and $\mathbf{w}_j$ as described in definition 4.4.5 and definition 4.4.8. □

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 4.4.6, note that the following holds.

Observation 4.4.10. *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 4.4.11. $\overrightarrow{B^{\parallel}} \leq_{\text{coor}} \left(\overrightarrow{A^{\parallel}}\right)^{\downarrow}$

Proof. Similar to Claim 4.4.7, this follows from the following claim.

Claim 4.4.12. *The number of* 1*s in* $\overrightarrow{B^{\|}}$ *is at most that in* $\overrightarrow{A^{\|}}$*, and* $\overrightarrow{B^{\|}}$ *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 4.4.6 and definition 4.4.5,

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

that is, we have written the LHS as a sum of Boolean vectors. And, we have from Observation 4.4.10 and definition 4.4.8, followed by Claim 4.4.9 and claim 4.4.11 that

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

Trivially, we have $\left\|\overrightarrow{\Phi_{\chi'}(h)}\right\|_{1/2} \leq \left\|\overrightarrow{s\Phi}\right\|_{1/2}$, and from Lemma 4.3.11, 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 eq. (4.2).

4.5 Connecting with the Distance to Monotonicity: Proof of Theorem 4.4.3

In this section, we set the intuition behind eq. (H2) straight. We show how the isoperimetric theorem Theorem 4.4.2 on semisorted functions can be used to prove Theorem 4.4.3. We begin by recalling the corollary of the undirected, colored Talagrand objective on the hypercube.

Corollary 4.5.1 (Corollary of Theorem 1.8 in [KMS18])**.** *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*

$$\mathbb{E}_{S \subseteq [d]} \; [\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}] = \Omega(\text{var}(g_{\mathbf{x}}))$$

As mentioned earlier, one can't show eq. (H2), that is, $\mathbb{E}_{\mathbf{x}}[\text{var}(g_{\mathbf{x}})] = \Omega(\varepsilon_f)$. 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 4.4.2 to random restrictions of f. One can show that there is some restriction where the corresponding $\mathbb{E}_{\mathbf{x}}[\text{var}(g_{\mathbf{x}})]$ 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 4.4.2 to a more convenient form, using the undirected Theorem 4.3.8.

Theorem 4.5.2. *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*

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

Proof. By Theorem 4.4.2, there exists some colorings $\xi_{\mathbf{x}}$ such that $T_{\Phi_\chi}(f) \geq \mathbb{E}_{\mathbf{x}}\mathbb{E}_S[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}]$. By the undirected Talagrand bound Theorem 4.3.8, $\mathbb{E}_S[\sqrt{I_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}] \geq C \cdot \mathrm{var}(g_{\mathbf{x}})$.

$$\begin{aligned}
\mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)] &= \mathbb{E}_S\mathbb{E}_{\mathbf{x}}[\mathbb{1}((S \circ f)(\mathbf{x}) \neq (\overline{S} \circ f)(\mathbf{x}))] \\
&= \mathbb{E}_S\mathbb{E}_{\mathbf{x}}[\mathbb{1}(g_{\mathbf{x}}(S) \neq g_{\mathbf{x}}(\overline{S}))] \\
&= \mathbb{E}_{\mathbf{x}}\mathbb{E}_S[\mathbb{1}(g_{\mathbf{x}}(S) \neq g_{\mathbf{x}}(\overline{S}))] \leq 4\mathbb{E}_{\mathbf{x}}[\mathrm{var}(g_{\mathbf{x}})] \qquad (4.4)
\end{aligned}$$

(The final inequality uses Claim 4.5.3, stated below.) Hence, $\mathbb{E}_{\mathbf{x}}\mathbb{E}_S[\sqrt{I_{g_{\mathbf{x}},\xi}(S)}] \geq (C/4)\mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$.

Claim 4.5.3. *For any Boolean function $h : \{0,1\}^d \to \{0,1\}$, $\mathbb{P}_S[h(S) \neq h(\overline{S})] \leq 4\mathrm{var}(h)$.*

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

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, $\mathbb{P}_S[h(S) \neq h(\overline{S})] \leq 2\min_{b\in\{0,1\}} \mathbb{P}_S[h(S) = b]$. Combining with the bound from the previous paragraph, $\mathbb{P}_S[h(S) \neq h(\overline{S})] \leq 4\mathrm{var}(h)$. □

□

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

Definition 4.5.4. *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 4.4.2 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 4.5.5. *For any function $h : [n]^k \to \{0, 1\}$, define $\delta(h)$ to be $\mathbb{E}_\pi[\Delta(h, \pi([k]) \circ h)]$.*

By Claim 4.3.3, 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 $\mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$. 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 4.5.6. $\delta(f) \leq \mathbb{E}_S \mathbb{E}_{h \sim \mathcal{R}(S)}[\delta(h)] + \mathbb{E}_\pi \mathbb{E}_S[\Delta(\pi(S) \circ f, \pi(\overline{S}) \circ f)]$

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 sorting only reduces the Hamming distance between functions (Claim 4.3.4).

Plugging this bound in and taking expectations over ordered subset S of dimensions:

$$\mathbb{E}_S[\Delta(f, S \circ \overline{S} \circ f)] \leq \mathbb{E}_S[\Delta(f, S \circ f)] + \mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)] \tag{4.5}$$

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 $\mathbb{E}_S[\Delta(f, S \circ f)] = \mathbb{E}_{h \sim \mathcal{R}(S)}[\Delta(h, S \circ h)]$. Roughly speaking, the quantity $\Delta(f, S \circ \overline{S} \circ f)$ is $\varepsilon(f)$ and $\mathbb{E}_{h \sim \mathcal{R}(S)}[\Delta(h, S \circ h)]$ is $\mathbb{E}_{h \sim \mathcal{R}S}\varepsilon(h)$. So we would hope that eq. (4.5) implies $\varepsilon(f) \leq \varepsilon(h) + \mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$.

Unfortunately, the quantities are only constant factor approximations of $\varepsilon(f), \varepsilon(h)$. So by converting eq. (4.5) in terms of $\varepsilon(f)$, we would potentially lose a constant factor in eq. (4.5).

To avoid this problem, we deal with $\delta(f)$ instead. By randomly permuting S and taking expectations, the quantities in eq. (4.5) can be replaced by $\delta(\cdot)$ terms. Taking expectations over a uar π, eq. (4.5) implies

$$\mathbb{E}_\pi \mathbb{E}_S[\Delta(f, \pi(S) \circ \pi(\overline{S}) \circ f)] \leq \mathbb{E}_\pi \mathbb{E}_S[\Delta(f, \pi(S) \circ f)] + \mathbb{E}_\pi \mathbb{E}_S[\Delta(\pi(S) \circ f, \pi(\overline{S}) \circ f)] \tag{4.6}$$

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

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

Combining all our bounds, we get that

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

□

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

Claim 4.5.7. *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, $T_{\Phi_\chi}(f) \geq (1/\sqrt{p}) \cdot \mathbb{E}_{S \sim \mathcal{H}(p)} \mathbb{E}_{h \sim \mathcal{R}_S}[T_{\Phi_\chi}(h)]$.*

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 $\mathbb{E}_{h \sim \mathcal{R}_S}[T_{\Phi_\chi}(h)] = \mathbb{E}_{h \sim \mathcal{R}_S}\mathbb{E}_{\mathbf{z}}[\sqrt{\Phi_{h,\chi}(\mathbf{z})}]$. 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,

$$\mathbb{E}_{h \sim \mathcal{R}_S}\mathbb{E}_{\mathbf{z}}[\sqrt{\Phi_{h,\chi}(\mathbf{z})}] = \mathbb{E}_{\mathbf{x}}[\sqrt{\Phi_{f,\chi}(\mathbf{x};S)}] \tag{4.7}$$

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

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

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 4.4.3 whice we restate below for convenience.

Theorem 4.4.3 (Theorem 4.0.4 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*

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

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 4.5.2.

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

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

(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 $\mathbb{E}_c[\Phi_{f|_\ell,\chi}(c)] \geq \varepsilon(f|_\ell)$, and $\sum_{i=1}^{d} \mathbb{E}_{\ell \text{ uar } i\text{-line}} \varepsilon(f|_\ell) = \Omega(\varepsilon(f))$.

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.

<u>Case 1, $\mathbb{E}_\pi \mathbb{E}_S[\Delta(\pi(S) \circ f, \pi(\overline{S}) \circ f)] \geq \delta(f)/10$:</u> By Theorem 4.5.2, $T_{\Phi_\chi}(f) \geq c \cdot \mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$ (for any ordering of coordinates). So $T_{\Phi_\chi}(f) \geq c \cdot \mathbb{E}_\pi \mathbb{E}_S[\Delta(\pi(S) \circ f, \pi(\overline{S}) \circ f)] \geq (c/10) \cdot \delta(f)$.

<u>Case 2, $\mathbb{E}_\pi \mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)] < \delta(f)/10$:</u> By Claim 4.5.6, $\mathbb{E}_S \mathbb{E}_{h \sim \mathcal{R}(S)}[\delta(h)] \geq \delta(f) - \mathbb{E}_\pi \mathbb{E}_S[\Delta(S \circ f, \overline{S} \circ f)]$. In this case, we can lower bound $\mathbb{E}_S \mathbb{E}_{h \sim \mathcal{R}(S)}[\delta(h)] \geq (9/10)\delta(f)$. 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,

$$\begin{aligned}
\mathbb{E}_{S\sim\mathcal{H}(1/2)}\mathbb{E}_{h\sim\mathcal{R}(S)}[T_{\Phi_\chi}(h)] &\geq 2^{-d}\sum_{S\neq[d]}\mathbb{E}_{h\sim\mathcal{R}(S)}[T_{\Phi_\chi}(h)] \geq 2^{-d}\cdot(c/10)\cdot\sum_{S\neq[d]}\mathbb{E}_{h\sim\mathcal{R}(S)}[\delta(h)] \\
&= 2^{-d}\cdot(c/10)\cdot\Big(\sum_{S\subseteq[d]}\mathbb{E}_{h\sim\mathcal{R}(S)}[\delta(h)] - \mathbb{E}_{h\sim\mathcal{R}([d])}[\delta(h)]\Big) \\
&= (c/10)\Big(\mathbb{E}_S\mathbb{E}_{h\sim\mathcal{R}(S)}[\delta(h)] - 2^{-d}\delta(f)\Big) \quad (h\sim\mathcal{R}([d]) \text{ is } f) \\
&\geq (c/10)\cdot(9/10)\cdot\delta(f) - 2^{-d}\cdot(c/10)\cdot\delta(f) \quad \text{(by case condition)} \\
&= (9/10-2^{-d})\cdot(c/10)\cdot\delta(f) \geq (4/5)\cdot(c/10)\cdot\delta(f) \qquad (4.8)
\end{aligned}$$

By Claim 4.5.7, $T_{\Phi_\chi}(f) \geq \sqrt{2}\cdot\mathbb{E}_{S\sim\mathcal{H}(1/2)}\mathbb{E}_{h\sim\mathcal{R}(S)}[T_{\Phi_\chi}(h)]$. Combining with the inequality of section 4.5, $T_{\Phi_\chi}(f) \geq (\sqrt{2}\cdot 4/5)\cdot(c/10)\cdot\delta(f) \geq (c/10)\cdot\delta(f)$. □

4.6 Connecting Talagrand Objectives of f and the Tracker Functions

In this section and the next, we establish our main technical result Theorem 4.4.2 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 4.4.2 (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*

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

To prove Theorem 4.4.2 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 $\xi_{\mathbf{x}}^{(i)}$ of the hypercube which assigns a $\{0,1\}$ value to every hypercube edge of the form $(T, T \oplus j)$ for all $j \le i$, and for all $T \subseteq [i]$. The process will begin with the null coloring, $\xi_{\mathbf{x}}^{(0)}$, and end with a complete coloring, $\xi_{\mathbf{x}} := \xi_{\mathbf{x}}^{(d)}$, 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 $\chi_S^{(i)}$ of the edges of the augmented hypergrid. We start with $\chi_\emptyset^{(0)} := \chi$ 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*

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

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 $\xi_{\mathbf{x}}^{(i)}$ for each $\mathbf{x} \in [n]^d$ as defined above. Also observe that since $\chi_\emptyset^{(0)} := \chi$, the arbitrary coloring specified in the theorem statement, we have that $R_0(\emptyset)$ is precisely the LHS in the statement of Theorem 4.4.2, that is, $R_0(\emptyset) = \mathbb{E}_{\mathbf{x}\in[n]^d}\left[\sqrt{\Phi_{f,\chi}(\mathbf{x})}\right]$. Additionally, since we use $\xi_{\mathbf{x}} := \xi_{\mathbf{x}}^{(d)}$, observe that $\mathbb{E}_{S\subseteq[d]}[R_d(S)]$ is precisely the RHS in the statement of Theorem 4.4.2.

With the above setup in mind, we show that the following Lemma 4.6.1 suffices to prove Theorem 4.4.2.

Lemma 4.6.1 (Potential Drop Lemma). *Fix $i \in \{1,\ldots,d\}$, $\xi_{\mathbf{x}}^{(i-1)}$ for all $\mathbf{x} \in [n]^d$, and $\chi_S^{(i-1)}$ for every $S \subseteq [i-1]$, which all satisfy the specifications described in the previous paragraph. There exists a choice of $\xi_{\mathbf{x}}^{(i)}$ for every $\mathbf{x} \in [n]^d$ and $\chi_S^{(i)}$, $\chi_{S+i}^{(i)}$ 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) $R_{i-1}(S) \ge R_i(S)$ and (b) $R_{i-1}(S) \ge R_i(S+i)$.*

Proof of Theorem 4.4.2. 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:

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

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

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

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

and this establishes the claim after exchanging the expectations. $\square$

4.7 Proof of Potential Drop Lemma 4.6.1

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 $j \le i-1$ 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 $j \le i$.

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

$$\vec{L}_\ell := \left(\underbrace{\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}},\xi_{\mathbf{x}}}(S)}_{\overrightarrow{L^{(1)}}_\ell} \;+\; \underbrace{\Phi_{h,\chi}(\mathbf{x};i)}_{\overrightarrow{L^{(2)}}_\ell} \;+\; \underbrace{\sum_{j=i+1}^{d} \Phi_{h,\chi}(\mathbf{x};j)}_{\overrightarrow{L^{(3)}}_\ell} \quad : \; \mathbf{x} \in \ell \right) \tag{4.9}$$

Observe that

$$R_{i-1}(S) = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \vec{L}_\ell \right\|_{1/2} = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{L^{(1)}}_\ell + \overrightarrow{L^{(2)}}_\ell + \overrightarrow{L^{(3)}}_\ell \right\|_{1/2} \tag{4.10}$$

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

Define

$$\vec{R}_\ell := \left(\underbrace{\sum_{j=1}^{i-1} I_{g_{\mathbf{x}},\color{red}{\xi'_{\mathbf{x}}}}{}^{=j}(S)}_{\overrightarrow{R^{(1)}}_\ell} \;+\; \underbrace{I^{=i}_{g_{\mathbf{x}},\color{red}{\xi'_{\mathbf{x}}}}(S)}_{\overrightarrow{R^{(2)}}_\ell} \;+\; \underbrace{\sum_{j=i+1}^{d} \Phi_{h,\color{red}{\chi'}}(\mathbf{x};j)}_{\overrightarrow{R^{(3)}}_\ell} \quad : \; \mathbf{x} \in \ell \right) \tag{4.11}$$

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

$$R_i(S) := \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \vec{R}_\ell \right\|_{1/2} = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{R^{(1)}}_\ell + \overrightarrow{R^{(2)}}_\ell + \overrightarrow{R^{(3)}}_\ell \right\|_{1/2} \tag{4.12}$$

Similarly, define

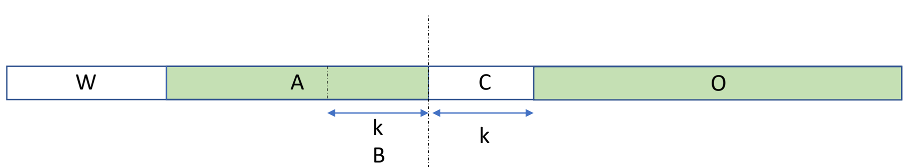
$$\vec{M}_\ell := \left(\underbrace{\sum_{j=1}^{i-1} I^{=j}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S+i)}_{\overrightarrow{M^{(1)}}_\ell} + \underbrace{I^{=i}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S+i)}_{\overrightarrow{M^{(2)}}_\ell} + \underbrace{\sum_{j=i+1}^{d} \Phi_{i \circ h, \chi''}(\mathbf{x}; j)}_{\overrightarrow{M^{(3)}}_\ell} \quad : \ \mathbf{x} \in \ell \right) \tag{4.13}$$

and notice that the "second" RHS term is

$$R_i(S+i) := \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \vec{M}_\ell \right\|_{1/2} = \frac{1}{n^d} \sum_{i\text{-lines } \ell} \left\| \overrightarrow{M^{(1)}}_\ell + \overrightarrow{M^{(2)}}_\ell + \overrightarrow{M^{(3)}}_\ell \right\|_{1/2} \tag{4.14}$$

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 4.2. 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 4.3.11 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.

4.7.1 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 4.3.11, 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 4.3.11 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 4.3.11 implies $\|L_O\|_{1/2} \geq \|R_O\|_{1/2}$.

4.7.1.1 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.

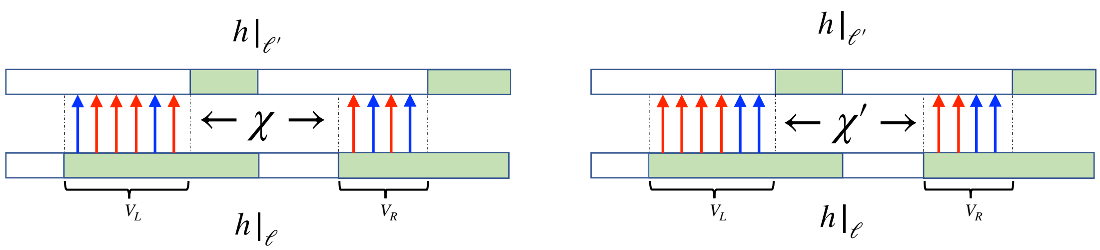
Given a pair of i-lines ℓ and $\ell' = \ell + a\mathbf{e}_j$ for $j \geq i + 1$ and $a > 0$, we consider the set of violations from ℓ to ℓ' in h:

$$V := \{(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) \colon \mathbf{x} \in \ell,\ h(\mathbf{x}) = 1, \text{ and } h(\mathbf{x} + a\mathbf{e}_j) = 0\}. \tag{4.15}$$

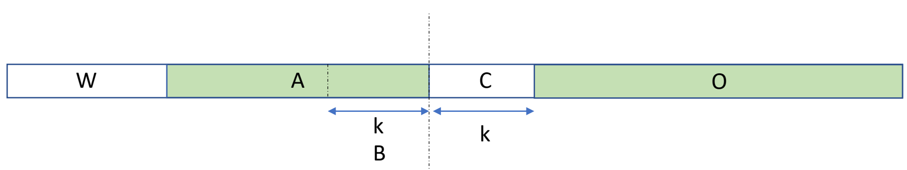
Since h is semi-sorted, it's clear that we can write $V = V_L \cup V_R$ as a union of two intervals, in the sense that $\{\mathbf{x} \colon (\mathbf{x}, \mathbf{x}+a\mathbf{e}_j) \in V_L\}$ is an interval in the lower half of ℓ and $\{\mathbf{x} \colon (\mathbf{x}, \mathbf{x}+a\mathbf{e}_j) \in V_R\}$ 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 4.7.1. *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

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

and

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

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

Definition 4.7.2. *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*

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

and

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

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

$$\Phi_{h,\chi'}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^R_{j,a}(\mathbf{x})\right) \quad \text{and} \quad \Phi_{h,\chi}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right) \tag{4.16}$$

Claim 4.7.3. *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 $\chi'(\mathbf{x}, \mathbf{x}+a\mathbf{e}_j) = 1$. $\square$

Claim 4.7.4. *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. $\square$

In conclusion, using the observation eq. (4.16), we conclude that we can write

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

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

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

is the same weighted sum of the *sorted decreasing* orders of those Boolean vectors. Therefore, we can conclude using Lemma 4.3.11,

$$\overrightarrow{R^{(3)}_A} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_A} \right)^{\downarrow} \tag{4.17}$$

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

$$\overrightarrow{R^{(3)}_O} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_O} \right)^{\downarrow} \tag{4.18}$$

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 4.7.5. *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 $h(\mathbf{x}' - a\mathbf{e}_j) = 1$. $\square$

Again, observe that our definition of χ' gives us

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

and

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

Definition 4.7.6. *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*

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

and

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

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

$$\Phi_{h,\chi'}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^R_{j,a}(\mathbf{x})\right) \quad \text{and} \quad \Phi_{h,\chi}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right) \tag{4.19}$$

Claim 4.7.7. *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 $\chi'(\mathbf{x}, \mathbf{x} + a\mathbf{e}_j) = 0$. $\square$

Claim 4.7.8. *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 $\overrightarrow{L^{(3)}_C}$ is a sum of Boolean vectors, and $\overrightarrow{R^{(3)}_C}$ is the sum of the *sorted increasing* orders of those Boolean vectors. Using Lemma 4.3.11, we can conclude

$$\overrightarrow{R^{(3)}_C} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_C}\right)^{\uparrow} \tag{4.20}$$

And an absolutely analogous argument gives

$$\overrightarrow{R^{(3)}_W} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_W}\right)^{\uparrow} \tag{4.21}$$

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

4.7.1.2 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 $\xi'_{\mathbf{x}} := \xi^{(i)}_{\mathbf{x}}$ 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 $\xi_{\mathbf{x}} := \xi^{(i-1)}_{\mathbf{x}}$ 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 $g_{\mathbf{x}}$:

$$V := \{\mathbf{x} \in \ell \colon g_{\mathbf{x}}(S) = 1 \text{ and } g_{\mathbf{x}}(S \oplus j) = 0\}\,. \tag{4.22}$$

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:

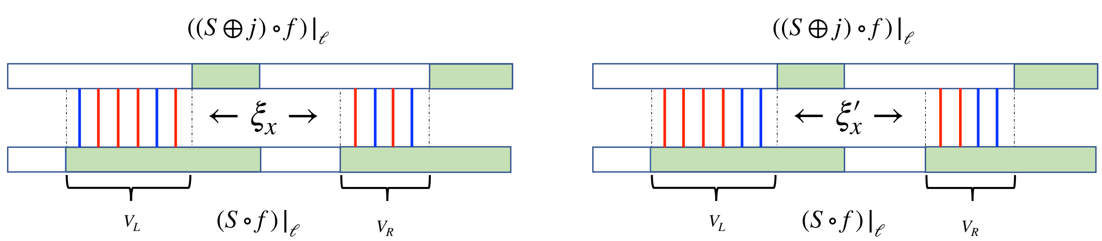
$$(\xi'_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_L) = (\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_L)^{\downarrow}$$

and similarly

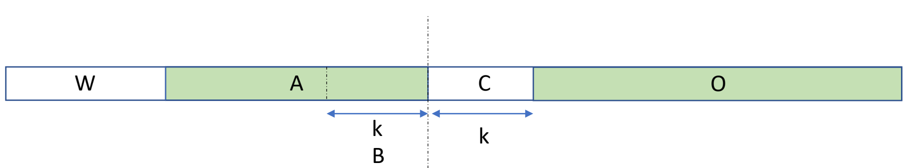
$$(\xi'_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_R) = (\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V_R)^{\downarrow}.$$

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 4.7.1.1, 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

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

and similarly

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

Claim 4.7.9. *$\left(I^{=j}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S) \;:\; \mathbf{x} \in A\right)$ 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 4.3.11 and the claim above, we get

$$\overrightarrow{R_A^{(1)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_A^{(1)}}\right)^{\downarrow} \tag{4.23}$$

Absolutely analogously, we get

$$\overrightarrow{R_O^{(1)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_O^{(1)}}\right)^{\downarrow} \tag{4.24}$$

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

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

and similarly

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

Analogous to Claim 4.7.9, we have the following claim.

Claim 4.7.10. $\left(I^{=j}_{g_{\mathbf{x}}, \xi'_{\mathbf{x}}}(S) \quad : \quad \mathbf{x} \in W\right)$ *is a sorted increasing vector, and is a permutation of* $\left(I^{=j}_{g_{\mathbf{x}}, \xi_{\mathbf{x}}}(S) \quad : \quad \mathbf{x} \in W\right)$.

Arguing similarly to the proof of eq. (4.23) we get

$$\overrightarrow{R_W^{(1)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_W^{(1)}}\right)^{\uparrow} \tag{4.25}$$

and absolutely analogously, we get

$$\overrightarrow{R_C^{(1)}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L_C^{(1)}}\right)^{\uparrow} \tag{4.26}$$

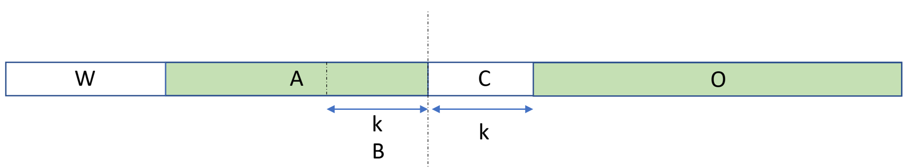
eq. (4.17), eq. (4.18), eq. (4.20), eq. (4.21), and eq. (4.23), eq. (4.24), eq. (4.25), eq. (4.26) establish (a) and (c).

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

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



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 4.7.11. *No matter how χ is defined, either $\overrightarrow{L_A^{(2)}}$ is the all 1s vector, or $\overrightarrow{L_C^{(2)}}$ 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. $\square$

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.

$$\text{If } \overrightarrow{L^{(2)}_A} \equiv \mathbf{1}, \quad \text{then } \xi'_{\mathbf{x}}(S, S+i) = 1 \ \ \forall \mathbf{x} \in A \cup C \tag{4.27}$$

otherwise,

$$\text{we have } \overrightarrow{L^{(2)}_C} \equiv \mathbf{1}, \quad \text{and so we define } \xi'_{\mathbf{x}}(S, S+i) = 0 \ \ \forall \mathbf{x} \in A \cup C \tag{4.28}$$

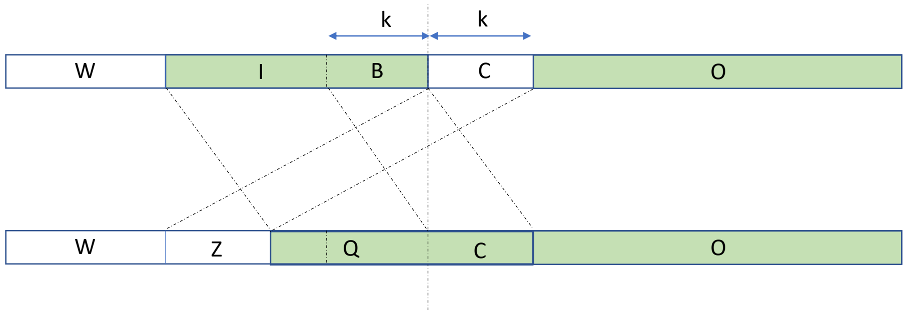
In the former case, we have $\overrightarrow{R^{(2)}_A} = (\underbrace{111\cdots1}_{k \text{ many}}0000)$ and $\overrightarrow{L^{(2)}_A} \equiv \mathbf{1}$ and so we pick $\overrightarrow{L'^{(2)}_A} = \overrightarrow{L^{(2)}_A}$. Also note that we have $\overrightarrow{R^{(2)}_C}$ as the all zeros 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 eq. (4.27) or eq. (4.28).

4.7.2 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 $\overrightarrow{L}$ as

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

and the vector $\overrightarrow{M}$ as

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

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 4.3.11, 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, $\|L_{WZ}\|_{1/2} \geq \|M_{WZ}\|_{1/2}$.

4.7.2.1 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 $\chi''(e) = \chi(e)$.

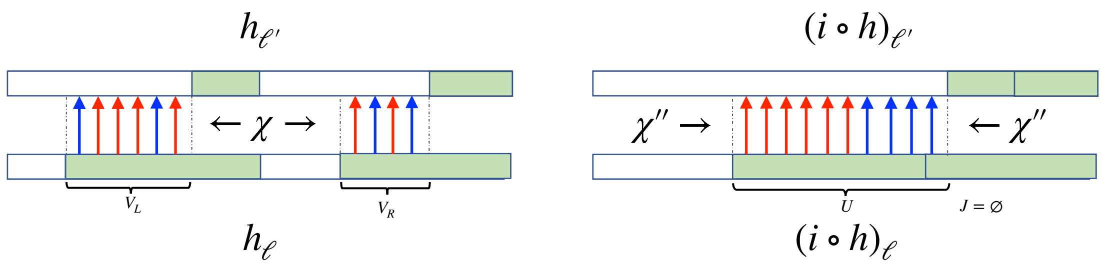
Given a pair of i-lines ℓ and $\ell' = \ell + a\mathbf{e}_j$ for $j \geq i+1$ and $a > 0$ we consider the set of violations from ℓ to ℓ' in h and in $i \circ h$. As before, the violations in h form two a union of two intervals $V = V_L \cup V_R$. Recall the definition of V in eq. (4.15). Since $(i \circ h)$ is sorted in dimension i, the violations from ℓ to ℓ' in $(i \circ h)$ form a single interval which we will call U:

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

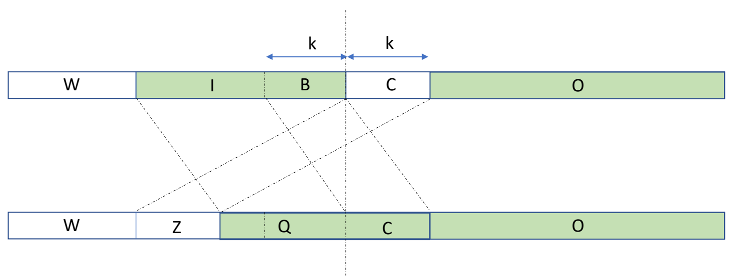
Since the sort operator can only reduce the number of violations in a dimension, we have $|U| \leq |V|$ (Claim 4.3.6 applied to $h|_\ell$ and $h|_{\ell'}$). We define J to be the interval of $|V| - |U|$ points directly to the right of U so that $U \cup J$ is an interval of size $|V|$. We then define

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

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 4.3.6 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

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

Let's see what this leads to.

Definition 4.7.12. *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:*

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

and

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

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

$$\Phi_{i\circ h,\chi''}(\mathbf{x};j) = \min\left(1, \sum_a \mathbf{v}^M_{j,a}(\mathbf{x})\right) \tag{4.29}$$

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

$$\Phi_{h,\chi}(\mathbf{x};j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right). \tag{4.30}$$

Claim 4.7.13. *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 $\chi''(\mathbf{x},\mathbf{x}+a\mathbf{e}_j)=1$. $\square$

Claim 4.7.14. *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 4.3.6 applied to χ restricted to the edges going from ℓ to ℓ'). $\square$

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

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

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 4.3.11, we establish part (e) for $q = 3$. Namely, we get

$$\exists \overrightarrow{M'^{(3)}_{QCO}} : \quad \overrightarrow{M'^{(3)}_{QCO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_{IBO}} \right)^{\downarrow} \quad \text{and} \quad \overrightarrow{M'^{(3)}_{QCO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(3)}_{QCO}} \tag{4.31}$$

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 $\ell'' = \ell - a e_j$. 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 4.3.6 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

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

Let's see what this leads to.

Definition 4.7.15. *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:*

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

and

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

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

$$\Phi_{i\circ h,\chi''}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^M_{j,a}(\mathbf{x})\right) \tag{4.32}$$

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

$$\Phi_{h,\chi}(\mathbf{x}; j) = \min\left(1, \sum_a \mathbf{v}^L_{j,a}(\mathbf{x})\right). \tag{4.33}$$

Claim 4.7.16. *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 $\chi''(\mathbf{x}' - a\mathbf{e}_j, \mathbf{x}') = 0$. $\square$

Claim 4.7.17. *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 4.3.6 applied to χ restricted to the edges going from ℓ'' to ℓ). $\square$

In conclusion, we can write

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

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

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

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 4.3.11, we establish part (g) for $q = 3$. Namely, we get

$$\exists \overrightarrow{M'^{(3)}_{MZ}} : \quad \overrightarrow{M'^{(3)}_{MZ}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(3)}_{WC}}\right)^{\uparrow} \quad \text{and} \quad \overrightarrow{M'^{(3)}_{WZ}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(3)}_{WZ}} \tag{4.34}$$

4.7.2.2 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 eq. (4.22). 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:

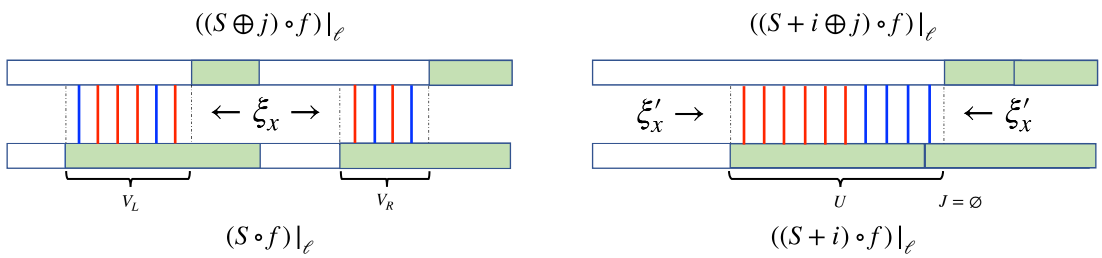
$$U := \{\mathbf{x} \in \ell \colon g_{\mathbf{x}}(S + i) = 1 \text{ and } g_{\mathbf{x}}(S + i \oplus j) = 0\}\,.$$

Again, we have $|U| \leq |V|$ (Claim 4.3.6 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

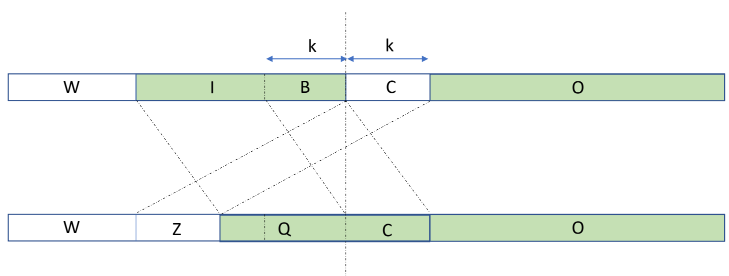
$$(\xi'_{\mathbf{x}}(S + i, S + i \oplus j) \colon \mathbf{x} \in U \cup J) = (\xi_{\mathbf{x}}(S, S \oplus j) \colon \mathbf{x} \in V)^{\downarrow}\,.$$

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 4.7.2.1 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 4.3.6 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

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

We now get the following claim.

Claim 4.7.18. $\left(I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i)\ :\ \mathbf{x}\in Q\cup C\cup O\right)$ *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

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

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 4.3.11, we establish part (e) for $q=3$. Namely, we get

$$\exists \overrightarrow{M'^{(1)}_{QCO}} :\ \ \overrightarrow{M'^{(1)}_{QCO}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_{IBO}}\right)^{\downarrow} \text{ and } \overrightarrow{M'^{(1)}_{QCO}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(1)}_{QCO}}. \tag{4.35}$$

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 4.3.6 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

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

Claim 4.7.19. *$\left(I^{=j}_{g_{\mathbf{x}},\xi'_{\mathbf{x}}}(S+i) \;:\; \mathbf{x} \in W \cup Z\right)$ 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,

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

and

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

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). $\square$

Observing that

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

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 4.3.11, we establish part (e) for $q = 3$. Namely,

$$\exists \overrightarrow{M'^{(1)}_{WZ}} : \quad \overrightarrow{M'^{(1)}_{WZ}} \succeq_{\mathsf{maj}} \left(\overrightarrow{L^{(1)}_{WC}}\right)^{\downarrow} \quad \text{and} \quad \overrightarrow{M'^{(1)}_{WZ}} \succeq_{\mathsf{coor}} \overrightarrow{M^{(1)}_{WZ}}. \tag{4.36}$$

4.7.2.3 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 eq. (4.27) and eq. (4.28) 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:

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

otherwise,

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

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 $\overrightarrow{M_{QCO}^{(2)}}$ is the zero vector, and thus we can choose $\overrightarrow{L'^{(2)}_{IBO}}$ to be $\left(\overrightarrow{L_{IBO}^{(2)}}\right)^{\downarrow}$.

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

CHAPTER 5

A $d^{1/2+o(1)}$ Query Tester

In this chapter we obtain a nearly optimal non-adaptive monotonicity tester for Boolean functions over d-dimensional hypergrids, which also yields the same result for continuous product spaces. These results were originally published in [BCS23a]. We refer the reader to Section 2.4 for a discussion of the main proof techniques and challenges.

Theorem 5.0.1. *Consider Boolean functions over the hypergrid, $f : [n]^d \to \{0,1\}$. There is a one-sided, non-adaptive tester for monotonicity that makes $\varepsilon^{-2} d^{1/2+O(1/\log\log d)}$ queries.*

Let $\mu = \prod_{i=1}^{d} \mu_i$ be a product Lebesgue measure over $\mathbb{R}^d$. A function $f \colon \mathbb{R}^d \to \{0,1\}$ is measurable if the set $f^{-1}(1)$ is Lebesgue-measurable with respect to μ. The μ-distance of f to monotonicity is defined as $\inf_{g \in \mathcal{M}} \mu(\Delta(f,g))$, where $\mathcal{M}$ is the family of measurable monotone functions and Δ is the symmetric difference operator. (Refer to Section 3.5 for more details.) Domain reduction results (Theorem 3.0.1, Theorem 3.0.2, or alternatively the techniques by [HY22]) show that monotonicity testing over general hypergrids and continuous product spaces can be reduced to testing over $[k]^d$ where $k = \mathrm{poly}(\varepsilon^{-1} d)$ via sampling. Thus, a direct consequence of Theorem 5.0.1 is the following theorem for continuous monotonicity testing.

Theorem 5.0.2. *Consider Boolean functions $f : \mathbb{R}^d \to \{0,1\}$, with an associated product measure μ. There is a one-sided, non-adaptive tester for monotonicity that makes $\varepsilon^{-2} d^{1/2+O(1/\log\log d)}$ queries.*

5.1 Random Walks and the Monotonicity Tester

Without loss of generality[1] we assume that n is a power of 2. We use $x \in_R S$ to denote choosing a uniform random element x from the set S. Abusing notation, we define intervals in $\mathbb{Z}_n$ by wrapping around. So, if $1 \leq i \leq n < j$, then the interval $[i, j]$ in $\mathbb{Z}_n$ is the set $[i, n] \cup [1, (j-1) \pmod{n}]$.

The directed (lazy) random walk distribution in $[n]^d$ that we consider is defined as follows. The distribution induced by this directed walk has multiple equivalent formulations, which are discussed in Section 5.2.2.

Definition 5.1.1 (Hypergrid Walk Distribution)**.** *For a point* $\mathbf{x} \in [n]^d$ *and walk length* τ*, the distribution* $\mathcal{U}_\tau(\mathbf{x})$ *over* $\mathbf{y} \in [n]^d$ *reached by an* upward lazy random walk from $\mathbf{x}$ of τ-steps *is defined as follows.*

1. *Pick a uniform random subset* $R \subseteq [d]$ *of* τ *coordinates.*
2. *For each* $r \in R$*:*
   (a) *Choose* $q_r \in_R \{1, 2, \ldots, \log n\}$ *uniformly at random.*
   (b) *Choose a uniform random interval* I_r *in* $\mathbb{Z}_n$ *of size* 2^{q_r} *such that* $\mathbf{x}_r \in I_r$*.*
   (c) *Choose a uniform random* $c_r \in_R I_r \setminus \{\mathbf{x}_r\}$*.*
3. *Generate* $\mathbf{y}$ *as follows. For every* $r \in [d]$*, if* $r \in R$ and $c_r > \mathbf{x}_r$*, set* $\mathbf{y}_r = c_r$*. Else, set* $\mathbf{y}_r = \mathbf{x}_r$*.*

Analogously, let $\mathcal{D}_\tau(\mathbf{x})$ *be the distribution defined precisely as above, but the* $>$*-sign is replaced by the* $<$*-sign in step 3. This is the distribution of the endpoint of a* downward lazy random walk from $\mathbf{x}$ of τ-steps.

[1]See Theorem A.1 of [BCS18]. Note this assumption is not crucial, but we choose to use it for the sake of a cleaner presentation.

A crucial step of our algorithm involves performing the exact same random walk, but originating from two different points. We can express our random walk distribution in terms of shifts (rather than destinations) as follows.

Definition 5.1.2 (Shift Distributions)**.** *The up-shift distribution from* $\mathbf{x}$*, denoted* $\mathcal{US}_\tau(\mathbf{x})$ *is the distribution of* $\mathbf{x}' - \mathbf{x}$*, where* $\mathbf{x}' \sim \mathcal{U}_\tau(\mathbf{x})$*. The down-shift distribution from* $\mathbf{x}$*, denoted* $\mathcal{DS}_\tau(\mathbf{x})$ *is the distribution of* $\mathbf{x} - \mathbf{x}'$*, where* $\mathbf{x}' \sim \mathcal{D}_\tau(\mathbf{x})$*.*

Note that $\mathcal{U}_\tau(\mathbf{x})$ is equivalent to the distribution of $\mathbf{x} + \mathbf{s}$, where $\mathbf{s} \sim \mathcal{US}_\tau(\mathbf{x})$. Similarly, $\mathcal{D}_\tau(\mathbf{x})$ is equivalent to the distribution of $\mathbf{x} - \mathbf{s}$, where $\mathbf{s} \sim \mathcal{DS}_\tau(\mathbf{x})$. Using Definition 5.1.1 and Definition 5.1.2, our tester is defined in Alg. 1.

Algorithm 1 Monotonicity tester for Boolean functions on $[n]^d$

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

1. Choose $p \in_R \{0, 1, 2, \ldots, \lceil \log d \rceil\}$ uniformly at random and set $\tau := 2^p$.
2. Run the *upward path test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{x} \in_R [n]^d$ and sample $\mathbf{y}$ from $\mathcal{U}_\ell(\mathbf{x})$.
   (b) If $f(\mathbf{x}) > f(\mathbf{y})$, then reject.
3. Run the *downward path test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{y} \in_R [n]^d$ and sample $\mathbf{x}$ from $\mathcal{D}_\ell(\mathbf{y})$.
   (b) If $f(\mathbf{x}) > f(\mathbf{y})$, then reject.
4. Run the *upward path + downward shift test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{x} \in_R [n]^d$, sample $\mathbf{y}$ from $\mathcal{U}_\ell(\mathbf{x})$, and sample $\mathbf{s}$ from $\mathcal{DS}_{\tau-1}(\mathbf{x})$.
   (b) If $f(\mathbf{x} - \mathbf{s}) > f(\mathbf{y} - \mathbf{s})$, then reject.
5. Run the *downward path + upward shift test* with walk length $\ell = \tau - 1$ and $\ell = \tau$:
   (a) Choose $\mathbf{y} \in_R [n]^d$, sample $\mathbf{x}$ from $\mathcal{D}_\ell(\mathbf{y})$, and sample $\mathbf{s}$ from $\mathcal{US}_{\tau-1}(\mathbf{y})$.
   (b) If $f(\mathbf{x} + \mathbf{s}) > f(\mathbf{y} + \mathbf{s})$, then reject.

Remark 5.1.3. *Given a function* $f : [n]^d \to \{0,1\}$*, consider the doubly-flipped function* $g : [n]^d \to \{0,1\}$ *defined as* $g(\mathbf{x}) := 1 - f(\bar{\mathbf{x}})$ *where* $\bar{\mathbf{x}}_i := n - \mathbf{x}_i$*. That is, we swap all the zeros and ones in* f*, and then reverse the hypergrid (the all zeros point becomes the all* n*'s point and vice-versa). The distance to monotonicity of both* f *and* g *are the same: a pair*

$(\mathbf{x}, \mathbf{y})$ *is violating in* f *if and only if* $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ *is violating in* g. *In Alg. 1, Step 2 on* f *is the same as Step 3 on* g, *and Step 4 on* f *is the same as Step 5 on* g. *In our analysis, we will construct a violation subgraph between vertex sets* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$. *Points in* $\boldsymbol{X}$ *are* 1*-valued and points in* $\boldsymbol{Y}$ *are* 0*-valued. If* $|\boldsymbol{X}| \leq |\boldsymbol{Y}|$, *then the steps 2, 3, and 4 suffice for the analysis. If* $|\boldsymbol{Y}| \leq |\boldsymbol{X}|$, *then (by the same analysis) we run steps 2,3, and 4 on the function* g. *This is equivalent to running steps 2, 3, and 5 on the function* f. *So, the tester covers both situations, and we can assume wlog that* $|\boldsymbol{X}| \leq |\boldsymbol{Y}|$. *This discussion happens in Section 5.5.1.1.*

Our main result is the following lower bound on the rejection probability of Alg. 1.

Theorem 5.1.4 (Main Theorem)**.** *Let* $n, \varepsilon^{-1} \leq \text{poly}(d)$. *If* $f\colon [n]^d \to \{0,1\}$ *is* ε*-far from being monotone, then for any* $\delta > (\log\log nd)^{-1}$, *Alg. 1 rejects* f *with probability at least* $\varepsilon^2 \cdot d^{-(1/2+O(\delta))}$.

Theorem 5.1.4 is proved in Section 5.4. We first use Theorem 5.1.4 to prove our main testing results, Theorem 5.0.1 and Theorem 5.0.2.

5.1.1 Proof of Theorem 5.0.1 and Theorem 5.0.2

To prove Theorem 5.0.1, we use the domain reduction Theorem 1.3 of [BCS20], which we state here for ease of reading.

Theorem 5.1.5 (Domain Reduction Theorem 1.3, [BCS20])**.** *Suppose* $f : [n]^d \to \{0,1\}$ *is* ε*-far from being monotone. Let* $k = (\varepsilon^{-1}d)^8$. *If* $\mathbf{T} = T_1 \times \cdots \times T_d$ *is a randomly chosen sub-grid, where for each* $i \in [d]$, T_i *is a (multi)-set formed by taking* k *independent, uniform samples from* $[n]$, *then* $\mathbb{E}_{\mathbf{T}}[\varepsilon_{f|_{\mathbf{T}}}] \geq \varepsilon/2$.

Remark 5.1.6. *We note that [HY22] obtain a more efficient domain reduction result. However, the domain reduction from [BCS20] can be used in a black-box fashion, resulting in a simpler tester.*

For ease of reading, we give a simplified proof of a weaker version of Theorem 5.0.1. This proof obtains a tester with an ε^{-3} *dependence, instead of the stated* ε^{-2}. *A more nuanced* *argument yields the improved $\varepsilon^{-2}\log(1/\varepsilon)$ dependence, which proves Theorem 5.0.1 as stated*[2]*. For details, we refer the reader to Section 7 of [BCS20]. In particular, we run Algorithm 1 in Section 7 of [BCS20] with the sub-routine in line 5 replaced by Alg. 1.*

Proof. of Theorem 5.0.1: Consider the tester which does the following, given $f\colon [n]^d \to \{0,1\}$ and $\varepsilon \in (0,1)$.

1. If $\varepsilon < d^{-1/2}$, then run the $\widetilde{O}(\varepsilon^{-1}d)$ query non-adaptive and 1-sided tester of [DGL+99] or [BRY14a].
2. If $\varepsilon \geq d^{-1/2}$, then set $k = (\varepsilon^{-1}d)^8 \leq d^{12}$ and repeat the following $8\varepsilon^{-1}$ times.
   (a) Sample a $[k]^d$ sub-grid $\mathbf{T} \subseteq [n]^d$ according to the distribution described in Theorem 5.1.5.
   (b) Run $32 \cdot \varepsilon^{-2} \cdot d^{1/2+O(\delta)}$ iterations of the tester described in Alg. 1 on the restricted function $f|_{\mathbf{T}}$.
3. Accept.

If $\varepsilon < d^{-1/2}$, then the number of queries is $O(\varepsilon^{-1}d) = O(\varepsilon^{-2}d^{1/2})$. We are done in this case.

Assume $\varepsilon \geq d^{-1/2}$. The total number of queries made by this tester is at most $\varepsilon^{-3} \cdot d^{1/2+O(\delta)}$. Clearly, if f is monotone, then the tester will accept, so suppose $\varepsilon_f \geq \varepsilon$. By the domain reduction Theorem 5.1.5, we have $\mathbb{E}_{\mathbf{T}}[\varepsilon_{f|_{\mathbf{T}}}] \geq \varepsilon/2$. So, $\mathbb{E}_{\mathbf{T}}[1-\varepsilon_{f|_{\mathbf{T}}}] \leq 1-\varepsilon/2$ and thus by Markov's inequality,

$$\mathbb{P}_{\mathbf{T}}\left[1-\varepsilon_{f|_{\mathbf{T}}} \geq 1-\varepsilon/4\right] \leq \frac{1-\varepsilon/2}{1-\varepsilon/4} = \frac{1-\varepsilon/4-\varepsilon/4}{1-\varepsilon/4} \leq 1-\varepsilon/4.$$

[2]When we invoke Alg. 1, we assume that $\varepsilon \geq d^{-1/2}$ and so $\log 1/\varepsilon = d^{\frac{\log\log 1/\varepsilon}{\log d}} \ll d^{O(1/\log\log d)}$. The factor of $\log 1/\varepsilon$ is absorbed by $d^{O(1/\log\log d)}$ in the query complexity.

Thus, $\mathbb{P}_{\mathbf{T}}[\varepsilon_{f|_{\mathbf{T}}} \geq \varepsilon/4] \geq \varepsilon/4$. Thus, with probability at least $1-(1-\varepsilon/4)^{8/\varepsilon} \geq 1-e^{-2}$, some iteration of step (2a) will produce $\mathbf{T}$ such that $\varepsilon_{f|_{\mathbf{T}}} \geq \varepsilon/4$. When this happens, some iteration of step (2b) will reject with probability at least $1-e^{-2}$, by Theorem 5.1.4. Thus, the tester rejects f with probability at least $(1-e^{-2})^2 \geq 2/3$. □

The proof of Theorem 5.0.2 for testing on $\mathbb{R}^d$ follows the exact same argument, using the corresponding domain reduction Theorem 1.4 of [BCS20] for functions over $\mathbb{R}^d$. We omit the proof.

5.2 Technical Preliminaries

In this section, we list out preliminary definitions and notations. Throughout the section, we fix a function $f : [n]^d \to \{0,1\}$ that is ε-far from monotone. For ease of readability, most proofs of this section are in the appendix.

5.2.1 Violation Subgraphs and Isoperimetry

The *fully augmented hypergrid* is a graph whose vertex set is $[n]^d$ where edges connect all pairs that differ in exactly one coordinate. We direct all edges from lower to higher endpoint. The edge $(\mathbf{x}, \mathbf{y})$ is called an *i-edge* for $i \in [d]$ if $\mathbf{x}$ and $\mathbf{y}$ differ in the ith coordinate. We use $I(\mathbf{x}, \mathbf{y}) = \{\mathbf{z} \colon \mathbf{x} \preceq \mathbf{z} \preceq \mathbf{y}\}$ to denote the points $\mathbf{z}$ in the segment $[\mathbf{x}, \mathbf{y}]$, that is, they are the points which differ from $\mathbf{x}$ and $\mathbf{y}$ only in the ith coordinate, and $\mathbf{x}_i \leq \mathbf{z}_i \leq \mathbf{y}_i$. Given a function $f : [n]^d \to \{0,1\}$ the edge $(\mathbf{x}, \mathbf{y})$ of the fully augmented hypergrid is a *violating/violated edge* if $f(\mathbf{x}) = 1$ and $f(\mathbf{y}) = 0$.

Definition 5.2.1. *A* violation subgraph *is a subgraph of the fully augmented hypergrid all of whose edges are violations.*

Note that any violation subgraph is a bipartite subgraph, where the bipartition is given by the 1-valued and 0-valued points. We henceforth always express a violation subgraph as

$G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ such that $\forall \mathbf{x} \in \boldsymbol{X}$, $f(\mathbf{x}) = 1$ and $\forall \mathbf{y} \in \boldsymbol{Y}$, $f(\mathbf{y}) = 0$. There are a number of relevant parameters of violation subgraphs that play a role in our analysis.

Definition 5.2.2. *Fix a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ and a point $\mathbf{x} \in \boldsymbol{X}$.*

- *The* degree *of $\mathbf{x}$ in G is the number of edges in E incident to $\mathbf{x}$ and is denoted as $D_G(\mathbf{x})$.*
- *For any coordinate $i \in [d]$, the i-*degree *of $\mathbf{x}$ in G is the total number of i-edges in E incident to $\mathbf{x}$ and is denoted as $\Gamma_{G,i}(\mathbf{x})$. Note $D_G(\mathbf{x}) = \sum_{i=1}^{d} \Gamma_{G,i}(\mathbf{x})$.*
- *The* thresholded degree *of $\mathbf{x}$ in G is the number of coordinates $i \in [d]$ with $\Gamma_{G,i}(\mathbf{x}) > 0$ and is denoted as $\Phi_G(\mathbf{x})$.*

Whenever G is clear from context, for brevity, we remove it from the subscript.

Note that $\Phi(\mathbf{x})$ is an integer between 0 and d, $\Gamma_i(\mathbf{x})$ is an integer between 0 and $(n-1)$, and $D(\mathbf{x})$ is an integer between 0 and $(n-1)d$. We next define the following parameters of a violation subgraph G.

Definition 5.2.3. *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$.*

- *$D(\boldsymbol{X})$ is the maximum degree of a vertex in $\boldsymbol{X}$, that is, $D(\boldsymbol{X}) = \max_{\mathbf{x} \in \boldsymbol{X}} D(\mathbf{x})$.*
- *For $i \in [d]$, $\Gamma_i(\boldsymbol{X})$ is the maximum i-degree in $\boldsymbol{X}$, that is, $\Gamma_i(\boldsymbol{X}) = \max_{\mathbf{x} \in \boldsymbol{X}} \Gamma_i(\mathbf{x})$.*
- *$\Gamma(\boldsymbol{X})$ is the maximum value of $\Gamma_i(\boldsymbol{X})$, that is, $\Gamma(\boldsymbol{X}) = \max_{i=1}^{d} \Gamma_i(\boldsymbol{X})$.*
- *$\Phi(\boldsymbol{X})$ is the maximum* thresholded *degree in $\boldsymbol{X}$, that is, $\Phi(\boldsymbol{X}) = \max_{\mathbf{x} \in \boldsymbol{X}} \Phi(\mathbf{x})$.*
- *$m(G)$ is the number of edges in G.*

(We analogously define these parameters for $\boldsymbol{Y}$.)

We recall the notion of *thresholded influence* of a function $f : [n]^d \to \{0, 1\}$ as defined in [BCS23b, BKKM23]. For any $\mathbf{x} \in [n]^d$ and $i \in [d]$, $\Phi_f(\mathbf{x}; i)$ is the indicator for the existence of a violating i-edge incident to $\mathbf{x}$. The thresholded influence of f at $\mathbf{x}$ is $\Phi_f(\mathbf{x}) = \sum_{i=1}^{d} \Phi_f(\mathbf{x}; i)$. We use the same Greek letter Φ both for thresholded influence and thresholded degree. In the graph $G_0 = (\boldsymbol{X}_0, \boldsymbol{Y}_0, E)$ consisting of *all* violating edges of the fully augmented hypergrid, $\Phi_f(\mathbf{x})$ is indeed $\Phi_{G_0}(\mathbf{x})$.

For applications to monotonicity testing, we require *colored/robust* versions of the thresholded influence. For hypercubes this was suggested by [KMS18], and for hypergrids this was generalized by [BCS23b]. Let $\chi : E \to \{0,1\}$ be an *arbitrary* coloring of all the edges of the fully augmented hypergrid to 0 or 1. Given a point $\mathbf{x}$ and $i \in [d]$, $\Phi_{f,\chi}(\mathbf{x}; i)$ is the indicator of a violating i-edge e incident to $\mathbf{x}$ with $\chi(e) = f(\mathbf{x})$. The colored thresholded influence of $\mathbf{x}$ wrt χ is simply $\Phi_{f,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Phi_{f,\chi}(\mathbf{x}; i)$. The Talagrand objective of f is defined as

$$T_{\Phi_\chi}(f) := \min_{\chi: E \to \{0,1\}} \sum_{\mathbf{x} \in [n]^d} \sqrt{\Phi_{f,\chi}(\mathbf{x})}.$$

The main result of [BCS23b] is the following.

Theorem 5.2.4 (Theorem 1.4, [BCS23b])**.** *If $f : [n]^d \to \{0,1\}$ is ε-far from monotone, then $T_{\Phi_\chi}(f) = \Omega(\frac{\varepsilon n^d}{\log n})$.*

We stress that the RHS above only loses a $\log n$ factor, which allows for domain reduction (setting $n = \mathrm{poly}(d)$). This is what yields the nearly optimal $\sqrt{d}$ dependence and independence on n in the tester query complexity.

We extend the definition of $T_{\Phi_\chi}(f)$ to arbitrary violation subgraphs as follows. Given a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ and a bicoloring $\chi : E \to \{0,1\}$ of its edges, for $\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}$ and $i \in [d]$ let $\Phi_{G,\chi}(\mathbf{z}; i) = 1$ if there is a violating i-edge $e \in E(G)$ incident to $\mathbf{z}$ such that $\chi(e) = f(\mathbf{z})$, and $\Phi_{G,\chi}(\mathbf{z}; i) = 0$ otherwise. Define $\Phi_{G,\chi}(\mathbf{x}) = \sum_{i=1}^{d} \Phi_{G,\chi}(\mathbf{x}; i)$. Note, if $\chi \equiv 1$, that is every edge is colored 1, then $\Phi_{G,\chi}(\mathbf{x}) = \Phi_G(\mathbf{x})$ for $\mathbf{x} \in \boldsymbol{X}$ and $\Phi_{G,\chi}(\mathbf{y}) = 0$ for all $\mathbf{y} \in \boldsymbol{Y}$. Similarly, if $\chi \equiv 0$, then $\Phi_{G,\chi}(\mathbf{y}) = \Phi_G(\mathbf{y})$ for $\mathbf{y} \in \boldsymbol{Y}$ and $\Phi_{G,\chi}(\mathbf{x}) = 0$ for $\mathbf{x} \in \boldsymbol{X}$.

Definition 5.2.5. *Given a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$, we define*

$$T_{\Phi_\chi}(G) := \min_{\chi} \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \left[\sqrt{\Phi_{G,\chi}(\mathbf{z})} \right],$$

where the min is taken over all edge bicolorings $\chi : E(G) \to \{0,1\}$.

If G_0 is the subgraph of all violations in the fully augmented hypergrid, then Theorem 5.2.4 states $T_{\Phi_\chi}(G_0) = \Omega(\varepsilon n^d / \log n)$. We make a couple of observations.

Observation 5.2.6. *For any violation subgraph* $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$,

- $D(\boldsymbol{X}) \leq \Gamma(\boldsymbol{X})\Phi(\boldsymbol{X})$ *and* $D(\boldsymbol{Y}) \leq \Gamma(\boldsymbol{Y})\Phi(\boldsymbol{Y})$.
- $m(G) \geq T_{\Phi_\chi}(G)$.

Proof. For any $\mathbf{x} \in \boldsymbol{X}$, we have $D(\mathbf{x}) = \sum_{i=1}^{d} \Gamma_i(\mathbf{x}) = \sum_{i:\Gamma_i(\mathbf{x})>0} \Gamma_i(\mathbf{x}) \leq (\max_i \Gamma_i(\mathbf{x})) \cdot \Phi(\mathbf{x}) \leq \Gamma(\boldsymbol{X})\Phi(\boldsymbol{X})$. The proof is analogous for $\boldsymbol{Y}$. For the second bullet, observe that $m(G) = \sum_{\mathbf{x}\in\boldsymbol{X}} D(\mathbf{x}) \geq \sum_{\mathbf{x}\in\boldsymbol{X}} \Phi(\mathbf{x}) \geq \sum_{\mathbf{x}\in\boldsymbol{X}} \sqrt{\Phi(\mathbf{x})} = \sum_{\mathbf{z}\in\boldsymbol{X}\cup\boldsymbol{Y}} \sqrt{\Phi_{G,\chi\equiv 1}(\mathbf{z})} \geq T_{\Phi_\chi}(G)$. □

Remark 5.2.7. *Throughout the remainder of the paper, we consider* d *to be at least a large constant and fix* $\delta > \frac{1}{\log\log nd}$. *As a result, we use bounds such as "*$d^\delta \geq \operatorname{poly}\log d$*" or "*$d - C\sqrt{d} \geq d/3$*" without explicitly reminding the reader that* d *is large. We use* $O(\delta)$ *to denote* $C \cdot \delta$ *for some unspecified, but fixed constant* C.

5.2.2 Equivalent Formulations of the Random Walk Distribution

Recall the random walk distribution described in Definition 5.1.1. It is useful to think of this walk as first sampling a random hypercube and then taking a random walk on that hypercube. The following definition describes the appropriate distribution over sub-hypercubes in $[n]^d$.

Definition 5.2.8 (Hypercube Distribution)**.** *We define the following distribution* $\mathbb{H}_{n,d}$ *over sub-hypercubes in* $[n]^d$. *For each coordinate* $i \in [d]$:

1. *Choose* $q_i \in_R \{1, 2, \ldots, \log n\}$ *uniformly at random.*

2. Choose a uniform random interval I_i of size 2^{q_i} in $\mathbb{Z}_n$.

3. Choose a uniform random pair $a_i < b_i$ from I_i.

Output $\mathbf{H} = \prod_{i=1}^{d}\{a_i, b_i\}$. When n and d are clear from context, we abbreviate $\mathbb{H} = \mathbb{H}_{n,d}$.

It will also be useful for us to think of our random walk distribution as first sampling $\mathbf{x} \in_R [n]^d$, then sampling a random hypercube which contains $\mathbf{x}$, and then taking a random walk from $\mathbf{x}$ in that hypercube. The appropriate distribution over hypercubes containing a point $\mathbf{x}$ is defined as follows.

Definition 5.2.9 (Conditioned Hypercube Distribution)**.** *Given $\mathbf{x} \in [n]^d$, we define the conditioned sub-hypercube distribution $\mathbb{H}_{n,d}(\mathbf{x})$ as follows. For each $i \in [d]$:*

1. Choose $q_i \in_R \{1, 2, \ldots, \log n\}$ uniformly at random.

2. Choose a uniform random interval I_i in $\mathbb{Z}_n$ of size 2^{q_i} such that $\mathbf{x}_i \in I_i$.

3. Choose a uniform random $c_i \in_R I_i \setminus \{\mathbf{x}_i\}$.

4. Set $a_i = \min(\mathbf{x}_i, c_i)$ and $b_i = \max(\mathbf{x}_i, c_i)$.

Output $\mathbf{H} = \prod_{i=1}^{d}\{a_i, b_i\}$. When n and d are clear from context we will abbreviate $\mathbb{H}(\mathbf{x}) = \mathbb{H}_{n,d}(\mathbf{x})$.

The random walk distribution in a hypercube $\mathbf{H}$ is defined as follows.

Definition 5.2.10 (Hypercube Walk Distribution)**.** *For a hypercube $\mathbf{H} = \prod_{i=1}^{d}\{a_i, b_i\}$, a point $\mathbf{x} \in \mathbf{H}$, and a walk length τ, we define the upward random walk distribution $\mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})$ over points $\mathbf{y} \in \mathbf{H}$ as follows.*

1. Pick a uniform random subset $R \subseteq [d]$ of τ coordinates.

2. Generate $\mathbf{y}$ as follows. For every $r \in [d]$, if $r \in R$ and $\mathbf{x}_r = a_r$, set $\mathbf{y}_r = b_r$. Else, set $\mathbf{y}_r = \mathbf{x}_r$.

Analogously, the downward random walk distribution $\mathcal{D}_{\mathbf{H},\tau}(\mathbf{x})$ is defined precisely as above, but instead in step 2 if $r \in R$ and $\mathbf{x}_r = b_r$, we set $\mathbf{y}_r = a_r$, and otherwise $\mathbf{y}_r = \mathbf{x}_r$.

We observe that the following walk distributions are equivalent and defer the proof to the appendix Section 5.8.1.

Fact 5.2.11. *The following three distributions over pairs $(\mathbf{x}, \mathbf{y}) \in [n]^d \times [n]^d$ are all equivalent.*

1. $\mathbf{x} \in_R [n]^d$, $\mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})$.
2. $\mathbf{H} \sim \mathbb{H}$, $\mathbf{x} \in_R \mathbf{H}$, $\mathbf{y} \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})$.
3. $\mathbf{x} \in_R [n]^d$, $\mathbf{H} \sim \mathbb{H}(\mathbf{x})$, $\mathbf{y} \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})$.

The analogous three distributions defined using downward random walks are also equivalent.

It is also convenient to define the shift distribution for hypercubes.

Definition 5.2.12 (Shift Distributions for Hypercube Walks)**.** *Given a hypercube $\mathbf{H}$, the up-shift distribution from $\mathbf{x} \in \mathbf{H}$, denoted $\mathcal{US}_{\mathbf{H},\tau}(\mathbf{x})$ is the distribution of $\mathbf{x}' - \mathbf{x}$, where $\mathbf{x}' \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})$. The down-shift distribution from $\mathbf{y} \in \mathbf{H}$, denoted $\mathcal{DS}_{\mathbf{H},\tau}(\mathbf{y})$ is the distribution of $\mathbf{y} - \mathbf{y}'$, where $\mathbf{y}' \sim \mathcal{DS}_{\mathbf{H},\tau}(\mathbf{y})$.*

5.2.3 Influence and Persistence

We define the following notion of influence for our random walk distribution Definition 5.1.1.

Definition 5.2.13. *The total and negative influences of $f\colon [n]^d \to \{0,1\}$ are defined as follows.*

- $\widetilde{I}_f = \mathbb{E}_{\mathbf{x}\in[n]^d}\left[d \cdot \mathbb{P}_{\mathbf{y}\sim\mathcal{U}_1(\mathbf{x})}[f(\mathbf{x}) \neq f(\mathbf{y})]\right]$
- $\widetilde{I}_f^- = \mathbb{E}_{\mathbf{x}\in[n]^d}\left[d \cdot \mathbb{P}_{\mathbf{y}\sim\mathcal{U}_1(\mathbf{x})}[f(\mathbf{x}) > f(\mathbf{y})]\right]$

The probability of the tester (Alg. 1) finding a violation in step (2b) when $\tau = 1$ is precisely $\widetilde{I}_f^-/d$. Recall the definition of the distribution $\mathbb{H}$ in Definition 5.2.8. For brevity, for a hypercube $\mathbf{H} = \prod_{i=1}^d \{a_i, b_i\}$ sampled from $\mathbb{H}$, we abbreviate $I_{\mathbf{H}} := I_{f|_{\mathbf{H}}}$ and $I_{\mathbf{H}}^- := I_{f|_{\mathbf{H}}}^-$. That is, if $f(\mathbf{x}) = 1$, then $I_{\mathbf{H}}(\mathbf{x})$ is the number of coordinates i for which $\mathbf{x}_i = a_i$, and $f(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, b_i, \mathbf{x}_{i+1}, \dots, \mathbf{x}_d) = 0$, and if $f(\mathbf{x}) = 0$, then $I_{\mathbf{H}}(\mathbf{x}) = 0$. Then, $I_{\mathbf{H}} = \mathbb{E}_{\mathbf{x} \in \mathbf{H}}[I_{\mathbf{H}}(\mathbf{x})]$. The definition is analogous for $I_{\mathbf{H}}^-$.

Claim 5.2.14. *$\widetilde{I}_f = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}}[I_{\mathbf{H}}]$ and $\widetilde{I}_f^- = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}}\left[I_{\mathbf{H}}^-\right]$.*

Proof. By Fact 5.2.11, the distribution $(\mathbf{x} \in_R [n]^d, \mathbf{y} \sim \mathcal{U}_1(\mathbf{x}))$ is equivalent to first sampling $\mathbf{H} \sim \mathbb{H}$, then sampling $(\mathbf{x} \in_R \mathbf{H}, \mathbf{y} \sim \mathcal{U}_{\mathbf{H},1}(\mathbf{x}))$. Recalling Definition 5.2.10, observe that $\mathbb{P}_{\mathbf{y} \sim \mathcal{U}_{\mathbf{H},1}(\mathbf{x})}[f(\mathbf{x}) \neq f(\mathbf{y})] = I_{\mathbf{H}}(\mathbf{x})/d$. Putting these observations together yields

$$\widetilde{I}_f = \mathbb{E}_{\mathbf{x} \in [n]^d}\left[d \cdot \mathbb{P}_{\mathbf{y} \sim \mathcal{U}_1(\mathbf{x})}[f(\mathbf{x}) \neq f(\mathbf{y})]\right] = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}} \mathbb{E}_{\mathbf{x} \in \mathbf{H}}\left[I_{\mathbf{H}}(\mathbf{x})\right] = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}}[I_{\mathbf{H}}]$$

An analogous argument proves the statement for negative influence. $\square$

The following claim states that if the normal influence is (very) large, then so is the negative influence. This is a simple generalization of, and indeed easily follows from, Theorem 9.1 in [KMS18]. The proof can be found in Section 5.8.2.

Claim 5.2.15. *If $\widetilde{I}_f > 9\sqrt{d}$, then $\widetilde{I}_f^- > \sqrt{d}$.*

Next, we define the notion of persistent points. This is similar to that in [KMS18] with a parameterization that we need for our purpose.

Definition 5.2.16. *Given a point $\mathbf{x} \in [n]^d$, a walk length τ, and a parameter $\beta \in (0, 1)$, we say that $\mathbf{x}$ is (τ, β)-up-persistent if*

$$\mathbb{P}_{\mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})}[f(\mathbf{y}) \neq f(\mathbf{x})] \leq \beta.$$

Similarly, $\mathbf{x}$ is called (τ, β)-down-persistent if the above bound holds when $\mathbf{y}$ is drawn from the downward walk distribution, $\mathcal{D}_\tau(\mathbf{x})$. If both bounds hold, then we call $\mathbf{x}$ (τ, β)-persistent.

The following claim upper bounds the fraction of non-persistent points. This is a generalization of Lemma 9.3 in [KMS18]. The proof is deferred to Section 5.8.2.

Claim 5.2.17. *If $\widetilde{I}_f \leq 9\sqrt{d}$, then the fraction of vertices that are not (τ, β)-persistent is at most $C_{per}\frac{\tau}{\beta\sqrt{d}}$ where C_{per} is a universal constant.*

5.2.4 The Middle Layers, Typical Points, and Walk Reversibility

All proofs in this section are deferred to Section 5.8.3.

Definition 5.2.18. *In a hypercube $\{0,1\}^d$, the c-*middle layers* consist of all points with Hamming weight in the range $[d/2 \pm \sqrt{4cd\log(d/\varepsilon)}]$. Given a d-dimensional hypercube $\mathbf{H}$, we let $\mathbf{H}_c \subseteq \mathbf{H}$ denote the c-middle layers of $\mathbf{H}$.*

We state a bound on the number of points in the hypercube which lie in the middle layers. This follows from a standard Chernoff bound argument.

Claim 5.2.19. *For a d-dimensional hypercube $\mathbf{H}$ and $c \geq 1$, we have $|\mathbf{H}_c| \geq (1-(\varepsilon/d)^c)\cdot 2^d$.*

We now define the notion of typical points in $[n]^d$. Recall the distribution $\mathbb{H}_{n,d}$ (Definition 5.2.8) over random sub-hypercubes in $[n]^d$ and the distribution $\mathbb{H}_{n,d}(\mathbf{x})$ (Definition 5.2.8) over random sub-hypercubes in $[n]^d$ that contain $\mathbf{x}$. A point $\mathbf{x}$ is *c-typical* if for most sub-hypercubes containing $\mathbf{x}$, the point $\mathbf{x}$ is present in their c-middle layers.

Definition 5.2.20 (Typical Points)**.** *Given $c \geq 1$, a point $\mathbf{x} \in [n]^d$ is called c-*typical *if*

$$\mathbb{P}_{\mathbf{H}\sim\mathbb{H}(\mathbf{x})}\left[\mathbf{x} \in \mathbf{H}_c\right] \geq 1-(\varepsilon/d)^5.$$

Claim 5.2.21 (Most Points are Typical)**.** *For any $\varepsilon \in (0,1)$ and $c \geq 6$,*

$$\mathbb{P}_{\mathbf{x}\in_R [n]^d}\left[\mathbf{x} \textit{ is } c\textit{-typical}\right] \geq 1-(\varepsilon/d)^{c-5}.$$

Intuitively, a short random walk from a typical point will always lead to point that is almost as typical. This is formalized as follows.

Claim 5.2.22 (Translations of Typical Points). *Suppose $\mathbf{x} \in [n]^d$ is c-typical. Then for a walk length $\tau \leq \sqrt{d}$, every point $\mathbf{x}' \in supp(\mathcal{U}_\tau(\mathbf{x})) \cup supp(\mathcal{D}_\tau(\mathbf{x}))$ is $(c + \frac{\tau}{\sqrt{d}})$-typical.*

Recall the three equivalent ways of expressing the walk distribution in Fact 5.2.11. We define the random walk probabilities only on points in the middle layers. This setup allows for the approximate reversibility of Lemma 5.2.24.

Definition 5.2.23. *Consider two vertices $\mathbf{x} \prec \mathbf{x}' \in [n]^d$ and a walk length τ. We define*

$$p_{\mathbf{x},\tau}(\mathbf{x}') = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}(\mathbf{x})} \left[\mathbf{1}\left(\mathbf{x} \in \mathbf{H}_{100} \ \wedge \ \mathbf{x}' \in \mathbf{H}_{100}\right) \cdot \mathbb{P}_{\mathbf{z} \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})}[\mathbf{z} = \mathbf{x}'] \right] \tag{5.1}$$

to be the probability of reaching $\mathbf{x}'$ by a random walk from $\mathbf{x}$, only counting the contribution when the random walk is taken on a hypercube that contains $\mathbf{x}$ and $\mathbf{x}'$ in the 100-middle layers. We analogously define $p_{\mathbf{x}',\tau}(\mathbf{x})$ using the downward random walk distribution in $\mathbf{H}$.

Consider $\mathbf{x} \prec \mathbf{x}'$ are two points in the middle layers. The following lemma asserts that the probability of reaching from $\mathbf{x}$ to $\mathbf{x}'$ via an upward walk of length $\ll \sqrt{d}$ is similar to the probability of reaching from $\mathbf{x}'$ to $\mathbf{x}$ via downward walk of the same length.

Lemma 5.2.24 (Reversibility Lemma). *For any points $\mathbf{x} \prec \mathbf{x}' \in [n]^d$ and walk length $\ell \leq \sqrt{d}/\log^5(d/\varepsilon)$, we have*

$$p_{\mathbf{x},\ell}(\mathbf{x}') = (1 \pm \log^{-3} d) p_{\mathbf{x}',\ell}(\mathbf{x}).$$

5.3 Red Edges, Blue Edges, and Nice Subgraphs

We now set the stage to prove Theorem 5.1.4. The first definition is that of *mostly-zero-below* points. These are points from which a downward random walk (Definition 5.1.1) leads to a point where the function evaluates to 0 with high probability.

Definition 5.3.1. *A point* $\mathbf{z}$ *is called* ℓ-mostly-zero-below*, or* ℓ-mzb*, if* $\mathbb{P}_{\mathbf{z}'\sim\mathcal{D}_\ell(\mathbf{z})}[f(\mathbf{z}') = 0] \geq 0.9$.

To appreciate the utility of ℓ-mzb points, consider the following scenario. Suppose $\mathbf{x}$ is a point with $f(\mathbf{x}) = 1$ and is (ℓ, β)-down-persistent (Definition 5.2.16) for some small β. Next suppose an upward random walk from $\mathbf{x}$ reaches an ℓ-mzb point $\mathbf{z}$. Then, we claim that Step 4 of Alg. 1 would succeed with constant probability in finding a violated edge. An ℓ-length downward walk from $\mathbf{x}$, due to down-persistence, would lead to a $\mathbf{x}'$ with $f(\mathbf{x}') = 1$ with probability at least $1-\beta$. The same ℓ-length downward walk from $\mathbf{z}$ would lead to a $\mathbf{z}'$ with $f(\mathbf{z}') = 0$ with ≥ 0.9 probability, since $\mathbf{z}$ is mostly-zero-below. Since $(\mathbf{x}, \mathbf{z})$ are comparable, so would be $(\mathbf{x}', \mathbf{z}')$. By a union bound, $(\mathbf{x}', \mathbf{z}')$ is a violation with probability at least $0.9-\beta$.

The next definition describes edges $(\mathbf{x}, \mathbf{y})$ of the violation subgraph most of whose internal vertices lead to mzb-points via an upward random walk. Uncreatively, we call such edges *red*. Recall that $I(\mathbf{x}, \mathbf{y}) = \{\mathbf{z} \colon \mathbf{x} \preceq \mathbf{z} \preceq \mathbf{y}\}$ denotes the closed interval of points from $\mathbf{x}$ to $\mathbf{y}$.

Definition 5.3.2. *A violated edge* $(\mathbf{x}, \mathbf{y})$ *is called/colored* red *for walk length* ℓ *if*

$$\mathbb{P}_{\mathbf{z}\in I(\mathbf{x},\mathbf{y})}\mathbb{P}_{\mathbf{z}'\sim\mathcal{U}_\ell(\mathbf{z})}[\mathbf{z}' \textit{ is } \ell\text{-}\mathsf{mzb}] \geq 0.01.$$

When ℓ *is clear by context, we call the edge red.*

There may be no ℓ-mzb points for the lengths we choose, that is, a downward walk from any point leads to a point where the function evaluates to 1. In that case, Step 3 of Alg. 1 is poised to succeed; for any violating edge $(\mathbf{x}, \mathbf{y})$, if we start from $\mathbf{y}$ then the downward walk should give a violation. This motivates the next definition which recognizes violated edges $(\mathbf{x}, \mathbf{y})$ most of whose internal vertices lead to points where the function evaluates to 1 via a downward random walk. We call such edges *blue*.

Definition 5.3.3. *A violated edge* $(\mathbf{x}, \mathbf{y})$ *is called/colored* blue *for walk length* ℓ *if*

$$\mathbb{P}_{\mathbf{z}\in I(\mathbf{x},\mathbf{y})}\mathbb{P}_{\mathbf{z}'\sim\mathcal{D}_\ell(\mathbf{z})}[f(\mathbf{z}') = 1] \geq 0.01.$$

When ℓ is clear by context, we simply call the edge blue.

We note that a violating edge $(\mathbf{x}, \mathbf{y})$ may be *both* red and blue, or perhaps more problematically, *neither* red nor blue. One of the key lemmas we prove is that we can get our hands on a violation subgraph with sufficiently many colored edges. If we have our hands on a large violation subgraph G with few red edges (but has some other properties), then we can find another comparable sized violation subgraph H all of whose edges are blue, and whose maximum degrees are bounded by those in G. The precise statement is given below. We defer the proof of this lemma to Section 5.6.

Lemma 5.3.4 (Red/Blue Lemma)**.** *Let $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ be a violation subgraph and $1 \leq \ell \leq \sqrt{d}/\log^5(d/\varepsilon)$ be a walk length such that the following hold.*

1. *At most half the edges are red for walk length ℓ.*
2. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are $(\ell, \log^{-5} d)$-up-persistent.*
3. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are 99-typical.*

Then there exists another violation subgraph $H(\boldsymbol{L}, \boldsymbol{R}, E')$ such that

1. *All edges are blue for walk length ℓ and $m(H) \geq m(G)/7$.*
2. *$\Gamma(\boldsymbol{L}) \leq \Gamma(\boldsymbol{X})$ and $\Gamma(\boldsymbol{R}) \leq \Gamma(\boldsymbol{Y})$.*
3. *$D(\boldsymbol{L}) \leq D(\boldsymbol{X})$ and $D(\boldsymbol{R}) \leq D(\boldsymbol{Y})$.*

The next two definitions capture certain "nice" violation subgraphs consisting of either red or blue edges. In Section 5.4, we show that if either of these subgraphs exist then we can prove the tester works with the desired probability. In Section 5.5 we show that one of these subgraphs must exist. Recall, $\Phi_H(\mathbf{x})$ is the *thresholded degree* of $\mathbf{x}$ in the subgraph H and $\delta > (\log\log nd)^{-1}$ is fixed (Remark 5.2.7).

Definition 5.3.5 ((σ, τ)-nice red violation subgraph)**.** *Given a parameter $\sigma \in (0, 1)$ and a walk length τ, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ is called a (σ, τ)-nice red violation subgraph if the following hold.*

(a) All edges in H are red for walk length $\tau - 1$.

(b) All vertices in $\boldsymbol{A}$ are $(\tau - 1, 0.6)$-down-persistent.

(c) $\sigma \Phi_H(\mathbf{x}) \le d^{1/2}$ for all $\mathbf{x} \in \boldsymbol{A}$.

(d) $\sigma \sum_{\mathbf{x} \in \boldsymbol{A}} \Phi_H(\mathbf{x}) \ge \varepsilon^2 \cdot n^d \cdot d^{-O(\delta)}$.

(e) $d^{1/2 - O(\delta)} \ge \tau \ge \sigma \cdot d^{1/2 - O(\delta)}$.

The first two conditions dictate that the subgraph is nice with respect to the length of the walk. In particular, the edges are red with respect to this length and furthermore the 1-vertices are down-persistent. As explained before the definition of red edges, this property is crucial for the success of Step 4 of Alg. 1. The fourth condition says that the total thresholded degree of the 1-vertices in H is large. I.e. for an average vertex $\mathbf{x} \in \boldsymbol{A}$, there will be many coordinates i for which there is an i-edge in H incident to $\mathbf{x}$. The third condition says that the max thresholded degree of vertices in $\boldsymbol{A}$ is not too large and so the total thresholded degree from the fourth condition must be somewhat spread amongst the vertices in $\boldsymbol{A}$. The final condition shows that the length of the walk is large compared to σ. Note, if $\sigma = \Theta(1)$ and the third bullet point's right hand side was 1 instead of $\sqrt{d}$, we would be in the case of a large matching of violated edges, which was the "simple case" discussed in Section 2.4.1.

The next definition is the analogous case of blue edges. When this type of subgraph exists we argue that Step 3 of Alg. 1 succeeds. Note that Step 3 is the downward path test (without a shift) and so we don't need a persistence property like condition (b) in the previous definition. This definition has the same conditions on the thresholded degree as the previous definition, but with respect to the 0-vertices of the subgraph.

Definition 5.3.6 ((σ, τ)-nice blue violation subgraph)**.** *Given a parameter $\sigma \in (0, 1)$ and a walk length τ, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ is called a (σ, τ)-nice blue violation subgraph if the following hold.*

(a) All edges in H are blue for walk length $\tau - 1$.

(b) $\sigma\Phi_H(\mathbf{y}) \leq d^{1/2}$ for all $\mathbf{y} \in \boldsymbol{B}$.

(c) $\sigma \sum_{\mathbf{y} \in \boldsymbol{B}} \Phi_H(\mathbf{y}) \geq \varepsilon^2 \cdot n^d \cdot d^{-O(\delta)}$.

(d) $d^{1/2-O(\delta)} \geq \tau \geq \sigma \cdot d^{1/2-O(\delta)}$.

The following lemma captures the utility of the above definitions. It's proof can be found in Section 5.4.

Lemma 5.3.7 (Nice Subgraphs and Random Walks)**.** *Suppose for a power of two $\tau \geq 2$, there exists a (σ, τ)-nice red subgraph or a (σ, τ)-nice blue subgraph. Then Alg. 1 finds a violating pair, and thus rejects f, with probability at least $\varepsilon^2 \cdot d^{-(1/2+O(\delta))}$.*

The following lemma shows that one of the two nice subgraphs always exists. It's proof can be found in Section 5.5.

Lemma 5.3.8 (Existence of nice subgraphs)**.** *Let $n, \varepsilon^{-1} \leq \mathrm{poly}(d)$. Suppose $f\colon [n]^d \to \{0,1\}$ is ε-far from monotone and $\widetilde{I}_f \leq 9\sqrt{d}$. Let $\delta > \frac{1}{\log\log nd}$ be a parameter. There exists $0 < \sigma_1 \leq \sigma_2 < 1$, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$, and a power of two $\tau \geq 2$, such that either H is a (σ_1, τ)-nice red subgraph or a (σ_2, τ)-nice blue subgraph.*

5.4 Tester Analysis

In this section we prove Theorem 5.1.4. First, in Section 5.4.1 we prove Lemma 5.3.7 which is the main tester analysis. Then in Section 5.4.2 we combine Lemma 5.3.7, Lemma 5.3.8 (which will be proven in Section 5.5), and Claim 5.2.15 to prove Theorem 5.1.4.

5.4.1 Main Analysis: Proof of Lemma 5.3.7

There are two cases depending on whether we have a nice red subgraph or a nice blue subgraph. In Case 1, Step 4 of Alg. 1 proves the lemma while in Case 2, Step 3 of Alg. 1

proves the lemma. The proofs are similar, but we provide both for completeness.

5.4.1.1 Case 1: H is a (σ, τ)-nice red subgraph

Since τ is a power of 2, the tester in Alg. 1 chooses it with probability $\log^{-1} d$. Thus, in the rest of the analysis we will condition on this event.

Given $\mathbf{x} \in \boldsymbol{A}$, let $C_{\mathbf{x}} \subseteq [d]$ denote the set of coordinates for which $\mathbf{x}$ has an outgoing edge in H. Note $|C_{\mathbf{x}}| = \Phi_H(\mathbf{x})$. Recall the upward path + downward shift test described in Step 4 of Alg. 1 and the walk distribution $\mathcal{U}_{\tau-1}(\mathbf{x})$ defined in Definition 5.1.1. We first lower bound the probability that $\mathbf{x} \in \boldsymbol{A}$ and $R \cap C_{\mathbf{x}} \neq \emptyset$ where $\mathbf{x}$ is chosen uniformly by the tester and $R \subseteq [d]$ is a random set of τ coordinates. Let $\mathcal{E}_1$ denote this event. The main calculation is to lower bound the probability of this event as follows.

$$\mathbb{P}[\mathcal{E}_1] = \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \mathbb{P}[R \cap C_{\mathbf{x}} \neq \emptyset] \geq \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \left[1 - \left(1 - \frac{|C_{\mathbf{x}}|}{d}\right)^{\tau}\right] \geq \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \left[1 - \exp\left(-\frac{\tau |C_{\mathbf{x}}|}{d}\right)\right]$$

The RHS can only decrease if we replace τ with its lower bound (Definition 5.3.5, (e)) of $\sigma \cdot d^{1/2 - O(\delta)}$. Also, observe that $\frac{\sigma d^{1/2-O(\delta)} |C_{\mathbf{x}}|}{d} = \frac{\sigma \Phi_H(\mathbf{x})}{d^{1/2+O(\delta)}} \leq 1$ using our upper bound, $\sigma \Phi_H(\mathbf{x}) \leq d^{1/2}$ (Definition 5.3.5, (c)). Now, using $e^{-x} \leq 1 - \frac{x}{2}$ for $x \leq 1$, the exponential term in the RHS is at most $1 - \frac{\sigma \Phi_H(\mathbf{x})}{2d^{1/2+O(\delta)}}$, yielding

$$\mathbb{P}[\mathcal{E}_1] \geq \frac{\sigma}{2d^{1/2+O(\delta)}} \cdot \frac{1}{n^d} \sum_{\mathbf{x} \in \boldsymbol{A}} \Phi_H(\mathbf{x}) \underbrace{\geq}_{\text{(Definition 5.3.5, (d))}} \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \tag{5.2}$$

The event $\mathcal{E}_1$ asserts that the tester has chosen a point $\mathbf{x} \in \boldsymbol{A}$ and there is at least one $r \in R$ for which there exists a red edge $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_r) \in E$ for some integer $a > 0$ in the subgraph H. Fix the smallest such $r \in R \cap C_{\mathbf{x}}$ and the corresponding edge in H.

Recall the random walk process in Definition 5.1.1. We define the following good events.

- $\mathcal{E}_2$: Step (2a) chooses q_r satisfying: if $a \leq n/4$, then $2^{q_r} \in [2a, 4a]$; if $a > n/4$, then $2^{q_r} = n$.

- $\mathcal{E}_3$: Step (2b) chooses the interval $I_r \supseteq [\mathbf{x}_r, \mathbf{x}_r + a]$.

- $\mathcal{E}_4$: Step (2c) chooses c_r uniformly[3] from $[\mathbf{x}_r, \mathbf{x}_r + a]$.

- $\mathcal{E}_5$: $\mathbf{y}$ is $(\tau - 1)$-mostly-zero-below as per Definition 5.3.1.

- $\mathcal{E}_6$: $f(\mathbf{y} - \mathbf{s}) = 0$ for $\mathbf{s}$ chosen in Step 4 of Alg. 1 from $\mathcal{DS}_{\tau-1}(\mathbf{x})$.

- $\mathcal{E}_7$: $f(\mathbf{x} - \mathbf{s}) = 1$ for $\mathbf{s}$ chosen in Step 4 of Alg. 1 from $\mathcal{DS}_{\tau-1}(\mathbf{x})$.

Firstly, note that $\mathbb{P}[\mathcal{E}_2] = \log^{-1} n$ for both cases of the edge length, a. Now, suppose $a \leq n/4$. Then, $\mathbb{P}[\mathcal{E}_3 \mid \mathcal{E}_2] \geq 1/2$ by the condition $q_r \geq 2a$ and $\mathbb{P}[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \geq 1/4$ by the condition $q_r \leq 4a$. If $a > n/4$, then $\mathbb{P}[\mathcal{E}_3 \mid \mathcal{E}_2] = 1$, since in this case $I_r = [n]$ and again $\mathbb{P}[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \geq 1/4$ since $[\mathbf{x}_r, \mathbf{x}_r + a]$ is at least a fourth of the entire line, $[n]$.

Now, since the edge $(\mathbf{x}, \mathbf{x} + a\mathbf{e}_r)$ is red (Definition 5.3.2) for walk length $\tau - 1$, we have $\mathbb{P}[\mathcal{E}_5 \mid \mathcal{E}_4] \geq 0.01$.

Since $\mathbf{y}$ is $(\tau - 1)$-mostly-zero-below, if we sample $\mathbf{s}'$ from $\mathcal{DS}_{\tau-1}(\mathbf{y})$ we get $f(\mathbf{y} - \mathbf{s}') = 0$ with probability ≥ 0.9. Now note that $\mathcal{DS}_{\tau-1}(\mathbf{y})$ and $\mathcal{DS}_{\tau-1}(\mathbf{x})$ differ only when the set $R \subseteq [d]$ chosen in Definition 5.1.1 contains a coordinate in $\mathrm{supp}(\mathbf{y} - \mathbf{x})$. Since $|\mathrm{supp}(\mathbf{y} - \mathbf{x})| \leq \tau$, $|R| \leq \tau$, and $\tau = o(\sqrt{d})$, we have $\mathbb{P}_R[R \cap \mathrm{supp}(\mathbf{y} - \mathbf{x}) \neq \emptyset] \leq \tau^2/d = o(1)$. Therefore, when $\mathbf{s}$ is drawn from $\mathcal{DS}_{\tau-1}(\mathbf{x})$, we get $f(\mathbf{y} - \mathbf{s}) = 0$ with probability $\geq 0.9(1 - o(1)) \geq 0.8$. That is, $\mathbb{P}[\mathcal{E}_6 \mid \mathcal{E}_5] \geq 0.8$.

Finally, all points in $\boldsymbol{A}$ are $(\tau - 1, 0.6)$-down-persistent (Definition 5.2.16) and so $\mathbb{P}[\mathcal{E}_7 \mid \mathbf{x} \in A] \geq 0.4$.

Now, let's put everything together. The final success probability of the tester is at least

[3] We point out the following minor technicality in our presentation. From Definition 5.1.1, note that c_r is chosen from $I_r \setminus \{\mathbf{x}_r\}$ and so technically we will never have $c_r = \mathbf{x}_r$. However, note that Step 4 of Alg. 1 also runs the upward path + downward shift tester using walk length $\tau - 1$ and this is equivalent to setting $c_r = \mathbf{x}_r$ in this analysis, so that the first step of the walk is of length 0. Thus, it is sound in this analysis to think of c_r as uniformly chosen from I_r and we make this simplification for ease of reading.

$\mathbb{P}[\mathcal{E}_6 \wedge \mathcal{E}_7]$, which by a union bound and the reasoning above, is at least

$$
\begin{aligned}
&(1 - \mathbb{P}[\neg\mathcal{E}_6 \mid \mathcal{E}_5] - \mathbb{P}[\neg\mathcal{E}_7 \mid \mathbf{x} \in A]) \cdot \mathbb{P}\left[\bigwedge_{i=1}^{5} \mathcal{E}_i\right] \\
&\geq (1 - 0.2 - 0.6) \cdot \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \cdot \frac{1}{\log n} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{100} \geq \frac{\varepsilon^2}{d^{1/2+O(\delta)}}
\end{aligned}
$$

where in the last inequality we used $n \leq \mathrm{poly}(d)$. This completes the proof when the nice subgraph is red.

5.4.1.2 Case 2: H is a (σ, τ)-nice blue subgraph

As in Case 1, since τ is a power of 2, the tester in Alg. 1 chooses it with probability $\log^{-1} d$. Thus, in the rest of the analysis we will condition on this event. Given $\mathbf{y} \in \boldsymbol{B}$, let $C_{\mathbf{y}} \subseteq [d]$ denote the set of coordinates for which $\mathbf{y}$ has an incoming edge in H. Note $|C_{\mathbf{y}}| = \Phi_H(\mathbf{y})$. Recall the downward path tester described in Step 3 of Alg. 1 and the walk distribution $\mathcal{D}_{\tau-1}(\mathbf{y})$ defined in Definition 5.1.1. We first lower bound the probability that $\mathbf{y} \in \boldsymbol{B}$ and $R \cap C_{\mathbf{y}} \neq \emptyset$ where $\mathbf{y}$ is chosen uniformly by the tester and $R \subseteq [d]$ is a random set of τ coordinates. Let $\mathcal{E}_1$ denote this event. The main calculation is to lower bound the probability of this event as follows.

$$
\mathbb{P}[\mathcal{E}_1] = \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \mathbb{P}[R \cap C_{\mathbf{y}} \neq \emptyset] \geq \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \left[1 - \left(1 - \frac{|C_{\mathbf{y}}|}{d}\right)^{\tau}\right] \geq \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \left[1 - \exp\left(-\frac{\tau |C_{\mathbf{y}}|}{d}\right)\right]
$$

As in Case 1, the RHS can only decrease if we replace τ with its lower bound (Definition 5.3.6, (d)) of $\sigma \cdot d^{1/2-O(\delta)}$, and a similar argument as in Case 1 gives

$$
\mathbb{P}[\mathcal{E}_1] \geq \frac{\sigma}{d^{1/2+O(\delta)}} \cdot \frac{1}{n^d} \sum_{\mathbf{y} \in \boldsymbol{B}} \Phi_H(\mathbf{y}) \underbrace{\geq}_{(\text{Definition 5.3.6, (c)})} \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \tag{5.3}
$$

As in Case 1, the event $\mathcal{E}_1$ says that the tester has chosen a point $\mathbf{y} \in \boldsymbol{B}$ and there exists $r \in R$ such that there exists an edge $(\mathbf{y} - a\mathbf{e}_r, \mathbf{y}) \in E$ in the subgraph H for some integer

$a > 0$. Fix the smallest $r \in R \cap C_{\mathbf{y}}$ and the corresponding edge in H. Now define the following good events for the remainder of the tester analysis.

- $\mathcal{E}_2$: Step (2a) chooses q_r satisfying: if $a \leq n/4$, then $2^{q_r} \in [2a, 4a]$; if $a > n/4$, then $2^{q_r} = n$.
- $\mathcal{E}_3$: Step (2b) chooses the interval $I_r \supseteq [\mathbf{y}_r - a, \mathbf{y}_r]$.
- $\mathcal{E}_4$: Step (2c) chooses c_r uniformly[4] from $[\mathbf{y}_r - a, \mathbf{y}_r]$.
- $\mathcal{E}_5$: $f(\mathbf{x}) = 1$.

The final success probability of the tester is at least $\mathbb{P}[\wedge_{i=1}^{5}\mathcal{E}_i]$. Firstly, note that $\mathbb{P}[\mathcal{E}_2] = \log^{-1} n$ for both cases of the edge length, a. Suppose $a \leq n/4$. Then, $\mathbb{P}[\mathcal{E}_3 \mid \mathcal{E}_2] \geq 1/2$ by the condition $q_r \geq 2a$ and $\mathbb{P}[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \geq 1/4$ by the condition $q_r \leq 4a$. If $a > n/4$, then $\mathbb{P}[\mathcal{E}_3 \mid \mathcal{E}_2] = 1$, since in this case $I_r = [n]$ and again $\mathbb{P}[\mathcal{E}_4 \mid \mathcal{E}_2, \mathcal{E}_3] \geq 1/4$.

Finally, since the edge $(\mathbf{y} - a\mathbf{e}_r, \mathbf{y})$ is blue for walk length $\tau - 1$, by definition (Definition 5.3.3) we have $\mathbb{P}[\mathcal{E}_5 \mid \mathcal{E}_4] \geq 0.01$. Putting everything together, we have

$$\mathbb{P}\left[\bigwedge_{i=1}^{5}\mathcal{E}_i\right] \geq \frac{\varepsilon^2}{d^{1/2+O(\delta)}} \cdot \frac{1}{\log n} \cdot \frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{100} \geq \frac{\varepsilon^2}{d^{1/2+O(\delta)}}$$

where in the last step we used $n \leq \mathrm{poly}(d)$ and this completes the proof when the nice subgraph is blue. Together, the cases complete the proof of Lemma 5.3.7.

5.4.2 Tying it Together: Proof of Theorem 5.1.4

Suppose $f : [n]^d \to \{0, 1\}$ is ε-far from being monotone. Recall the definitions of $\widetilde{I}_f, \widetilde{I}_f^-$ in Definition 5.2.13. By Claim 5.2.15, if $\widetilde{I}_f > 9\sqrt{d}$, then $\widetilde{I}_f^- > \sqrt{d}$ and so the tester (Alg. 1)

[4]The same minor technicality arises here as in the previous subsection. We will never have $c_r = \mathbf{y}_r$ as per Definition 5.1.1, but Step 3 of Alg. 1 also runs the downward path tester with walk length $\tau - 1$ and this is equivalent to setting $c_r = \mathbf{y}_r$ in this analysis. Thus, it is again sound in this analysis to think of c_r being uniformly chosen from I_r.

finds a violation in step (2) when $\tau = 1$ with probability $\Omega(d^{-1/2})$. Thus, we will assume $\widetilde{I}_f \le 9\sqrt{d}$ and so we may invoke Lemma 5.3.8 which gives us either a nice red subgraph or a nice blue subgraph. Lemma 5.3.7 then proves that Alg. 1 finds a violating pair and rejects with probability at least $\varepsilon^2 \cdot d^{-(1/2+O(\delta))}$. This proves Theorem 5.1.4.

5.5 Finding Nice Subgraphs

In this section we prove Lemma 5.3.8 which we restate below.

Lemma 5.5.1 (Existence of nice subgraphs)**.** *Let $n, \varepsilon^{-1} \le \mathrm{poly}(d)$. Suppose $f\colon [n]^d \to \{0,1\}$ is ε-far from monotone and $\widetilde{I}_f \le 9\sqrt{d}$. Let $\delta > \frac{1}{\log\log nd}$ be a parameter. There exists $0 < \sigma_1 \le \sigma_2 < 1$, a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$, and a power of two $\tau \ge 2$, such that either H is a (σ_1, τ)-nice red subgraph or a (σ_2, τ)-nice blue subgraph.*

The proof proceeds over multiple steps and constitutes a key technical contribution of the paper. We give a sketch of what is forthcoming.

- In Section 5.5.1 we describe the construction of a *seed* regular violation subgraph G. This uses the directed isoperimetric result Theorem 5.2.4 proved in [BCS23b] and a "peeling argument" not unlike that present in [KMS18]. At the end of this section, we will fix the parameters σ_1, σ_2 and the walk length τ. In particular, the length τ will be defined by the *larger* side of this violating bipartite graph.
- In Section 5.5.2, we obtain a regular violating graph H that has persistence properties with respect to the walk length τ. In [KMS18] and [BCS23b], one obtained this violating graph by simply deleting the non-persistent points from the seed violation subgraph. In our case, since we choose the walk length depending on the larger side, we need to be careful. We use the idea of "translating violation subgraphs" on G (repeatedly) to find a different violation subgraph H with the desired persistence properties.

- In Section 5.5.3, we use the graph H to obtain either a nice red subgraph H_1 or a nice blue subgraph H_2. If most of the edges in H were red, then a simple surgery on H itself gives us H_1. On the other hand, if H has few red edges (but has the persistence properties as guaranteed), then we apply the red/blue lemma (Lemma 5.3.4) to obtain the desired nice blue-subgraph H_2. The proof of the red/blue lemma, which is present in Section 5.6, uses the translating violation subgraphs idea as well.

Throughout, we assume $f : [n]^d \to \{0,1\}$ is a function which is ε-far from being monotone, $\widetilde{I}_f \le 9\sqrt{d}$ and $n, \varepsilon^{-1} \le \text{poly}(d)$. In particular, we fix a constant c so that $nd \le d^c$. We also fix a $\delta \approx \frac{1}{\log\log nd} = o(1)$.

5.5.1 Peeling Argument to Obtain Seed Regular Violation Subgraph

Recall the definition of the Talagrand objective (Definition 5.2.5) $T_{\Phi_\chi}(G)$ of a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$. Let G_0 denote the violation subgraph formed by all violating edges in the fully augmented hypergrid. Theorem 1.4 in [BCS23b] (paraphrased in this paper as Theorem 5.2.4) is that $T_{\Phi_\chi}(G_0) = \Omega(\varepsilon n^d / \log n)$. Also recall the definitions in Definition 5.2.2. The following lemma asserts that there exists a subgraph of G_0 whose Talagrand objective is not much lower, but satisfies certain regularity properties.

Lemma 5.5.2 (Seed Regular Violation Subgraph)**.** *There exists a violation subgraph $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ satisfying the following properties.*

(a) $T_{\Phi_\chi}(G) \ge \varepsilon \cdot d^{-c\delta} \cdot n^d$.

(b) $m(G) \ge d^{-3c\delta} \max(|\boldsymbol{X}|\Phi(\boldsymbol{X})\Gamma(\boldsymbol{X}), |\boldsymbol{Y}|\Phi(\boldsymbol{Y})\Gamma(\boldsymbol{Y}))$.

(c) All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are 98-typical.

(d) $|\boldsymbol{X}|, |\boldsymbol{Y}| \ge \frac{\varepsilon}{d^{1/2+c\delta}} \cdot n^d$.

Let us make a few comments before proving the above lemma. Condition (a) shows that the Talagrand objective degrades only by a $d^{o(1)}$ factor. Condition (b) shows that the graph is nearly regular since the RHS term without the $d^{-o(1)}$ term is the maximum value of $m(G)$. This is because $\Phi(\boldsymbol{X})\Gamma(\boldsymbol{X})$ is an upper bound on the maximum degree of any vertex $\mathbf{x} \in \boldsymbol{X}$. Indeed, if one can prove a stronger lemma which replaces the $d^{o(1)}$ terms in (a) and (b) by polylog(d)'s, then the remainder of our analysis could be easily modified to give a $\tilde{O}(\varepsilon^{-2}\sqrt{d})$ tester.

We need a few tools to prove this lemma. Our first claim is a consequence of the subadditivity of the square root function.

Claim 5.5.3. *Consider a partition of (the edges of) a violation subgraph G into $H_1, H_2, \ldots, H_k$. Then $\sum_{j \le k} T_{\Phi_\chi}(H_j) \ge T_{\Phi_\chi}(G)$.*

Proof. Let χ_j denote the coloring of the subgraph H_j that obtains the minimum $T_{\Phi_\chi}(H_j)$. Since the $H_1, \ldots, H_k$ form a partition, we can aggregate the colors to get a coloring χ of G.

Consider any $\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}$. Let $\Phi_{H_j,\chi_j}(\mathbf{z})$ be the thresholded degree of $\mathbf{z}$, restricted to the edges colored by χ_j. By the subadditivity of the square root function, $\sum_{j \le k} \sqrt{\Phi_{H_j,\chi_j}(\mathbf{z})} \ge \sqrt{\sum_{j \le k} \Phi_{H_j,\chi_j}(\mathbf{z})}$. Observe that thresholded degrees are also subadditive, so $\sum_{j \le k} \Phi_{H_j,\chi_j}(\mathbf{z}) \ge \Phi_{G,\chi}(\mathbf{z})$. Hence,

$$\begin{aligned} \sum_{j \le k} T_{\Phi_\chi}(H_j) &= \sum_{j \le k} \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sqrt{\Phi_{H_j,\chi_j}(\mathbf{z})} = \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sum_{j \le k} \sqrt{\Phi_{H_j,\chi_j}(\mathbf{z})} \\ &\ge \sum_{\mathbf{z} \in \boldsymbol{X} \cup \boldsymbol{Y}} \sqrt{\Phi_{G,\chi}(\mathbf{z})} \ge T_{\Phi_\chi}(G). \end{aligned} \tag{5.4}$$

□

Remark 5.5.4. *The proof of Claim 5.5.3 crucially uses the fact that in the definition of $T_{\Phi_\chi}()$, we minimize over all possible colorings χ's of the edges. In particular, if we had defined $T_{\Phi_\chi}(G)$ only with respect to the all ones or the all zeros coloring, then the above proof fails. In the remainder of the paper, we will only be using the $\chi \equiv 1$ or $\chi \equiv 0$ colorings, and the curious reader may wonder why we need the definition of $T_{\Phi_\chi}(G)$ to minimize over all colorings. This is exactly the point where we need it. We make this remark because the* *"uncolored" isoperimetric theorem is much easier to prove than the "colored" version, but the colored/robust version is essential for the tester analysis.*

Our next step is a simple bucketing argument.

Claim 5.5.5. *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$. Both of the following are true.*

1. *There exists a subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ of G such that $T_{\Phi_\chi}(G') \geq \delta^2 T_{\Phi_\chi}(G)$ and $m(G') \geq (nd)^{-\delta}|\boldsymbol{X}'|\Phi(\boldsymbol{X}')\Gamma(\boldsymbol{X}')$.*

2. *There exists a subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ of G such that $T_{\Phi_\chi}(G') \geq \delta^2 T_{\Phi_\chi}(G)$ and $m(G') \geq (nd)^{-\delta}|\boldsymbol{Y}'|\Phi(\boldsymbol{Y}')\Gamma(\boldsymbol{Y}')$.*

Proof. We prove item (1) and the proof of item (2) is analogous.

For convenience, we assume that δ is the reciprocal of a natural number. For each $\mathbf{x} \in \boldsymbol{X}$, we bucket the incident edges as follows. First, for each $a \in [1/\delta]$, let S_a be the set of dimensions i, such that the i-degree of $\mathbf{x}$ is in the range $[n^{(a-1)\delta}, n^{a\delta})$. Note that $S_1, \ldots, S_{1/\delta}$ forms a partition of the set of coordinates, $[d]$. Now, for each $a, b \in [1/\delta]$, let the (a, b) edge bucket of $\mathbf{x}$, denoted $E_{a,b,\mathbf{x}}$, be defined as follows. If $|S_a| \in [d^{(b-1)\delta}, d^{b\delta})$, then $E_{a,b,\mathbf{x}}$ is the set of all edges incident to $\mathbf{x}$ along dimensions in S_a. If $|S_a| \notin [d^{(b-1)\delta}, d^{b\delta})$, then $E_{a,b,\mathbf{x}} = \emptyset$. Observe $\{E_{a,b,\mathbf{x}} \colon a, b \in [1/\delta]\}$ partitions the edges incident to $\mathbf{x}$.

Now, let $G_{a,b}$ denote the subgraph formed by the edge set $\cup_{\mathbf{x} \in \boldsymbol{X}} E_{a,b,\mathbf{x}}$. Let $\boldsymbol{X}_{a,b}$ be the set of vertices in $\boldsymbol{X}$ with non-zero degree in $G_{a,b}$. Observe that $\Phi(\boldsymbol{X}_{a,b}) \leq d^{b\delta}$ and $\Gamma(\boldsymbol{X}_{a,b}) \leq n^{a\delta}$. Moreover, the degree of each $\mathbf{x} \in \boldsymbol{X}_{a,b}$ is at least $d^{(b-1)\delta} \times n^{(a-1)\delta} \geq (nd)^{-\delta}\Phi(\boldsymbol{X}_{a,b})\Gamma(\boldsymbol{X}_{a,b})$. Hence, $m(G_{a,b}) \geq (nd)^{-\delta}|\boldsymbol{X}_{a,b}|\Phi(\boldsymbol{X}_{a,b})\Gamma(\boldsymbol{X}_{a,b})$.

Finally, by construction, the $G_{a,b}$ subgraphs partition the edges of G. Hence, by Claim 5.5.3 we have $\sum_{a,b \in [1/\delta]} T_{\Phi_\chi}(G_{a,b}) \geq T_{\Phi_\chi}(G)$. By averaging, there exists some choice of a, b such that $T_{\Phi_\chi}(G_{a,b}) \geq \delta^2 T_{\Phi_\chi}(G)$. This gives the desired subgraph G'. □

Claim 5.5.5, part 1 above gives the regularity condition only with respect to $\boldsymbol{X}$, and part 2 gives the analogous guarantee with respect to $\boldsymbol{Y}$, but the trouble is in getting both simultaneously. We do an iterative construction using Claim 5.5.5 to get the simultaneous guarantee.

Proof. (part (a) and (b) of Lemma 5.5.2) By the robust directed Talagrand theorem for hypergrids (Theorem 5.2.4), there is a violation subgraph $G_0 = (\boldsymbol{X}_0, \boldsymbol{Y}_0, E_0)$ such that $T_{\Phi_\chi}(G_0) = \Omega(\varepsilon n^d / \log n)$. We construct a series of subgraphs $G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_r$ as follows.

Let $i \geq 1$. If i is odd, we apply item (1) of Claim 5.5.5 to G_{i-1} to get $G_i(\boldsymbol{X}_i, \boldsymbol{Y}_i, E_i)$ with the regularity condition on $\boldsymbol{X}_i$. If i is even, we apply item (2) of Claim 5.5.5 to G_{i-1} to get $G_i(\boldsymbol{X}_i, \boldsymbol{Y}_i, E_i)$ with the regularity condition on $\boldsymbol{Y}_i$. If $i > 1$ and $m(G_i) \geq (nd)^{-\delta} m(G_{i-1})$, then we terminate the series. By Claim 5.5.5, the series satisfies the following three properties for all $i \geq 1$.

- $T_{\Phi_\chi}(G_i) = \Omega(\delta^{2i} \varepsilon n^d / \log n)$.
- If i is odd, $m(G_i) \geq (nd)^{-\delta} |\boldsymbol{X}_i| \Phi(\boldsymbol{X}_i) \Gamma(\boldsymbol{X}_i)$. If i is even, $m(G_i) \geq (nd)^{-\delta} |\boldsymbol{Y}_i| \Phi(\boldsymbol{Y}_i) \Gamma(\boldsymbol{Y}_i)$.
- If the series has not terminated by step i, then $m(G_i) < (nd)^{-\delta} m(G_{i-1})$.

The first two statements hold by the guarantees of Claim 5.5.5 and the fact that $T_{\Phi_\chi}(G_0) = \Omega(\varepsilon n^d / \log n)$. The third statement holds simply by the termination condition for the sequence. The trivial bound on the number of edges is $m(G_0) \leq nd \cdot n^d$. The third bullet point yields $m(G_i) < (nd)^{-i\delta} \cdot nd \cdot n^d$, if the series has not terminated by step i.

Claim 5.5.6. *The series terminates in at most $3/\delta$ steps.*

Proof. Suppose not. Noting that $m(G_i) \geq T_{\Phi_\chi}(G_i)$ (Observation 5.2.6), we get the following chain of inequalities using the properties of our subgraph graph $G_{3/\delta}$.

$$(nd)^{-(3/\delta)\cdot\delta} \cdot nd \cdot n^d > m(G_i) \geq T_{\Phi_\chi}(G_i) = \Omega(\delta^{6/\delta} \varepsilon n^d / \log n) \quad \Longrightarrow \quad (nd)^{-2} = \Omega(\delta^{6/\delta} \varepsilon / \log n)$$

Note that we may assume $\varepsilon \geq 1/d$ and so $C\varepsilon/\log n \geq (nd)^{-1}$ for any constant C. Thus we have $(nd)^{-1} \geq \delta^{6/\delta}$. Given that $\delta > 1/\log\log nd$, this inequality is a contradiction. $\square$

By the previous claim the series terminates in some $r \leq 3/\delta$ steps, producing $G_r(\boldsymbol{X}_r, \boldsymbol{Y}_r, E_r)$, which we claim has the desired properties to prove conditions (a) and (b) of Lemma 5.5.2. Since $r \leq 3/\delta$, $T_{\Phi_\chi}(G_r) = \Omega(\delta^{6/\delta}\varepsilon n^d/\log n)$. Note that since $\delta > 1/\log\log nd$, we have

$$\delta^{6/\delta} > (\log\log nd)^{-\frac{6}{\delta}} = (nd)^{-\frac{6}{\delta}\cdot\frac{\log\log\log nd}{\log nd}} > (nd)^{-\delta^2} > (nd)^{-\delta}\cdot\log n > d^{-c\delta}\log n$$

where the second to last step holds because $\frac{6\log\log\log nd}{\log d} \ll \left(\frac{1}{\log\log nd}\right)^3 < \delta^3$. The last inequality used $nd \leq d^c$. This proves condition (a). Towards proving condition (b), note that $C\delta^{6/\delta}/\log n \geq (nd)^{-\delta}$ for any constant C.

Let's assume without loss of generality that r is even. Thus we have $m(G_r) \geq (nd)^{-\delta}|\boldsymbol{Y}_r|\Phi(\boldsymbol{Y}_r)\Gamma(\boldsymbol{Y}_r)$ by the second bullet point above. Next, since the series terminated at step r, we have

$$m(G_r) \geq (nd)^{-\delta}m(G_{r-1}) \geq (nd)^{-2\delta}|\boldsymbol{X}_{r-1}|\Phi(\boldsymbol{X}_{r-1})\Gamma(\boldsymbol{X}_{r-1}) \geq (nd)^{-2\delta}|\boldsymbol{X}_r|\Phi(\boldsymbol{X}_r)\Gamma(\boldsymbol{X}_r)$$

where the second inequality is again by the second bullet point above and the fact that $i-1$ is odd and the third inequality is simply because G_r is a subgraph of G_{r-1}. Again using $nd \leq d^c$, we have $(nd)^{-\delta} \geq d^{-c\delta}$ and so we get that G_r satisfies conditions (a) and (b) of Lemma 5.5.2. $\square$

Proof. (Conditions (c) and (d) Lemma 5.5.2) To obtain condition (c), we simply remove the non-typical points. Recall the definition of c-typical points (Definition 5.2.20). By Claim 5.2.21, the number of points that are not 98-typical is at most $(\varepsilon/d)^{93}n^d$. Thus, removing all such vertices can decrease $T_{\Phi_\chi}(G)$ by at most $(\varepsilon/d)^{93}n^d\cdot\sqrt{d}$ which is negligible compared to the RHS in condition (a). Thus, we remove all such vertices from G and henceforth assume that all of $\boldsymbol{X}\cup\boldsymbol{Y}$ is 98-typical.

Condition (d) follows from condition (a). Consider the constant coloring $\chi \equiv 1$ and observe that

$$|\boldsymbol{X}|\sqrt{d} \geq \mathrm{Tal}_{\chi\equiv 1}(G) \geq \mathrm{Tal}(G) \geq \varepsilon \cdot d^{-c\delta} \cdot n^d.$$

where the first inequality follows from the trivial observation that the maximum $\Phi_G(\mathbf{x})$ can be is d. Using the coloring $\chi \equiv 0$ proves the same lower bound for $|\boldsymbol{Y}|$. □

5.5.1.1 Choice of the walk length

We end this section by specifying what the parameters σ_1, σ_2 and τ are going to be in Lemma 5.3.8. We now make the assumption $|\boldsymbol{X}| \leq |\boldsymbol{Y}|$. Given Remark 5.1.3, this is without loss of generality; this fact would be true either in f or in g, and running steps 2, 3, 5 on f is equivalent to running steps 2, 3, 4 on g. The violation subgraphs for f and g are isomorphic. Then,

$$\sigma_1 = \sigma_{\boldsymbol{X}} := \frac{|\boldsymbol{X}|}{n^d} \quad \text{and} \quad \sigma_2 = \sigma_{\boldsymbol{Y}} := \frac{|\boldsymbol{Y}|}{n^d}$$

and set τ to be the unique power of two such that

$$\frac{1}{2}\lceil \sigma_{\boldsymbol{Y}} \cdot d^{1/2-7c\delta} \rceil < \tau - 1 \leq \lceil \sigma_{\boldsymbol{Y}} \cdot d^{1/2-7c\delta} \rceil.$$

We conclude the subsection by establishing the following upper bound on the number of vertices which are not up-persistent.

Claim 5.5.7. *We may assume that*

- *the number of vertices $\mathbf{x} \in \boldsymbol{X}$ where $f(\mathbf{x}) = 1$ that are not $(\tau - 1, \log^{-5} d)$-up-persistent is at most $d^{-6c\delta} \cdot |\boldsymbol{X}|$, and*
- *the number of vertices $\mathbf{y} \in \boldsymbol{Y}$ where $f(\mathbf{y}) = 0$ that are not $(\tau - 1, \log^{-5} d)$-up-persistent is at most $d^{-6c\delta} \cdot |\boldsymbol{Y}|$.*

Proof. The statement for points where $f(\mathbf{x}) = 1$ is implied by item (4) of Lemma 5.5.2, for otherwise the tester succeeds with the desired probability when it runs the upward path tester with walk length $\tau - 1$ (step (2) of Alg. 1).

Now we prove the statement for points where $f(\mathbf{y}) = 0$. By Claim 5.2.17, the total number of $(\tau - 1, \log^{-5} d)$-non-persistent vertices is at most $C_{per}\tau \cdot \log^5 d \cdot \frac{1}{\sqrt{d}} \cdot n^d \leq \sigma_{\boldsymbol{Y}} \cdot d^{-6c\delta} \cdot n^d$, where we have simply used $\log^5 d \ll d^{c\delta}$ and our definition of τ. $\square$

5.5.2 Using 'Persist-or-Blow-up' Lemma to obtain Down-Persistence

Lemma 5.5.2 provides a seed violation subgraph which has a large Talagrand objective and has regularity properties. Claim 5.5.7 shows that we may assume these vertices are up-persistent with respect to walk length of $\tau - 1$. However, we may not have down persistence. In particular, it could be $|\boldsymbol{X}| \ll |\boldsymbol{Y}|$ and if we try to apply Claim 5.2.17 and remove all nodes from $\boldsymbol{X}$ which are not $(\tau - 1, 0.6)$-down-persistent, we may end up removing everything. To obtain a subgraph with down-persistence properties, we need to apply a translation procedure which is encapsulated in the lemma below. The proof of the lemma is deferred to Section 5.7.

Lemma 5.5.8 (Persist-or-Blow-up Lemma). *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$ such that all vertices in G are c-typical where $c \leq 99$ and $(\ell, \log^{-5} d)$-up persistent where $1 \leq \ell \leq \sqrt{d}/\log^5(d/\varepsilon)$. Then, there exists a violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ where all vertices are $(c + \frac{\ell}{\sqrt{d}})$-typical and satisfying one of the following conditions.*

1. *Down-persistent case:*
   - (a) *All vertices in $\boldsymbol{X}'$ are $(\ell, 0.6)$-down persistent.*
   - (b) *$m(G') \geq m(G)/\log^5 d$.*
   - (c) *$D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, and $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$*
   - (d) *$D(\boldsymbol{Y}') \leq D(\boldsymbol{Y})$, and $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq \Gamma_i(\boldsymbol{Y})$.*
2. *Blow-up case:*
   - (a) *$m(G') \geq 2(1 - 3\log^{-3} d) \cdot m(G)$.*
   - (b) *$D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, and $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$*

(c) $D(\boldsymbol{Y}') \leq 2D(\boldsymbol{Y})$, *and* $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq 2\Gamma_i(\boldsymbol{Y})$.

That is, the application of the above lemma either gives the violation subgraph we need, or it gives us a violation subgraph with around double the edges. In the remainder of this section we use Lemma 5.5.8 and the graph $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ derived in the previous section to prove the following lemma.

Lemma 5.5.9 (Down-Persistent Violation Subgraph)**.** *Let $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ be the subgraph asserted in Lemma 5.5.2. There exists a natural number $s \leq \log^3 d$ and a violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ with the following properties.*

1. $m(H) \geq 2^s \frac{m(G)}{\log^7 d}$.
2. $\Gamma(\boldsymbol{A}) \leq \Gamma(\boldsymbol{X})$ *and* $\Gamma(\boldsymbol{B}) \leq 2^s \Gamma(\boldsymbol{Y})$.
3. $D(\boldsymbol{A}) \leq D(\boldsymbol{X})$ *and* $D(\boldsymbol{B}) \leq 2^s D(\boldsymbol{Y})$.
4. *All vertices in $\boldsymbol{A} \cup \boldsymbol{B}$ are $(\tau - 1, \log^{-5} d)$-up-persistent and 99-typical.*
5. *All vertices in $\boldsymbol{A}$ are $(\tau - 1, 0.6)$-down-persistent.*

Proof. We use Lemma 5.5.8 to define the following process generating a sequence of violation subgraphs. The initial graph is $G_0 = (\boldsymbol{X}_0, \boldsymbol{Y}_0, E_0)$ which is the seed regular violation subgraph obtained from Lemma 5.5.2.

For each $i \geq 1$:

1. Obtain G'_{i-1} by removing all vertices from $\boldsymbol{X}_{i-1} \cup \boldsymbol{Y}_{i-1}$ that are not $(\tau - 1, \log^{-5} d)$-up-persistent.
2. Invoke Lemma 5.5.8 with walk length $\tau - 1$ on G'_{i-1} to obtain $G_i = (\boldsymbol{X}_i, \boldsymbol{Y}_i, E_i)$.
3. If G_i satisfies the down persistence condition of Lemma 5.5.8 then halt and return G_i.

4. If G_i satisfies the blowup condition of Lemma 5.5.8, then continue.

By Lemma 5.5.8, if the process does not halt on step i, then we have the following recurrences.

- $m(G_i) \geq 2(1 - 3\log^{-3} d) \cdot m(G'_{i-1})$.
- $D(\boldsymbol{X}_i) \leq D(\boldsymbol{X}_{i-1})$, $\Gamma(\boldsymbol{X}_i) \leq \Gamma(\boldsymbol{X}_{i-1})$, $D(\boldsymbol{Y}_i) \leq 2D(\boldsymbol{Y}_{i-1})$, $\Gamma(\boldsymbol{Y}_i) \leq 2\Gamma(\boldsymbol{Y}_{i-1})$.

Furthermore, we have the following claim that bounds the number of edges lost in step (1).

Claim 5.5.10. *For every $i \geq 1$, we have $m(G'_{i-1}) \geq m(G_{i-1}) - d^{-2c\delta} \cdot 2^{i-1} \cdot m(G)$.*

Proof. By Claim 5.5.7, the number of vertices we remove from $\boldsymbol{X}_{i-1}$ in step (1) is at most $d^{-6c\delta} \cdot |\boldsymbol{X}|$ and the number of vertices we remove from $\boldsymbol{Y}_{i-1}$ in step (1) is at most $d^{-6c\delta} \cdot |\boldsymbol{Y}|$. The number of edges we remove by deleting these vertices from $\boldsymbol{Y}_{i-1}$ is at most

$$d^{-6c\delta}|\boldsymbol{Y}|D(\boldsymbol{Y}_{i-1}) \leq d^{-6c\delta}2^{i-1}|\boldsymbol{Y}|D(\boldsymbol{Y}) \leq d^{-3c\delta}2^{i-1}m(G) \tag{5.5}$$

where in the second inequality we used $D(\boldsymbol{Y}) \leq \Phi(\boldsymbol{Y})\Gamma(\boldsymbol{Y})$ and the regularity property on G (item (2) of Lemma 5.5.2).

An analogous argument bounds the number of removed edges when we delete non-persistent vertices from $\boldsymbol{X}_{i-1}$. Thus the total number of edges removed is at most $d^{-2c\delta}2^{i-1}m(G)$. $\square$

Claim 5.5.11. *If $i \leq \log^3 d$ and the process has not halted by step i, then $m(G_i) \geq \Omega(2^i m(G))$.*

Proof. For brevity, let $\alpha = 2(1 - 3\log^{-3} d)$ and $\beta = d^{-2c\delta} m(G)$. Using the above bounds, we get the recurrence

$$m(G_i) \geq \alpha \cdot m(G'_{i-1}) \geq \alpha(m(G_{i-1}) - \beta 2^{i-1}).$$

Expanding this recurrence yields $m(G_i) \geq \alpha^i m(G) - \beta \sum_{j=1}^{i} \alpha^j \cdot 2^{i-j}$. Observe that the subtracted term can be bounded as

$$\beta \sum_{j=1}^{i} \alpha^j \cdot 2^{i-j} = d^{-2c\delta} 2^i m(G) \sum_{j=1}^{i} (1 - 3\log^{-3} d)^j \leq d^{-c\delta} 2^i m(G)$$

simply using the fact that $i \leq \log^3 d \ll d^{c\delta}$. The first term is

$$\alpha^i m(G) = 2^i (1 - 3\log^{-3} d)^i m(G) \geq C \cdot 2^i m(G)$$

for some constant C. Combining the above two bounds completes the proof. $\square$

Claim 5.5.12. *The above process halts in $s \leq \log^3 d$ iterations.*

Proof. Suppose that the above process has not halted by step $i = \log^3 d$. By the previous claim, the number of edges in G_i is at least $C \cdot 2^i m(G) = C \cdot d^{\log^2 d} m(G)$ for some constant C. By Observation 5.2.6, note that $m(G) \geq T_{\Phi_\chi}(G)$ and thus is $\geq \varepsilon \cdot d^{-c\delta} \cdot n^d$ by item (1) of Lemma 5.5.2. Thus, the number of edges in G_i is at least $C \cdot \varepsilon \cdot d^{\log^2 d - c\delta} n^d$. Note that the total number of edges in the fully augmented hypergrid is at most $nd \cdot n^d$. Moreover, recall that we are assuming $nd \leq d^c$ and $\varepsilon \geq d^{-1/2}$. Therefore, $m(G_i) \gg nd \cdot n^d$ and this is a contradiction. $\square$

By Claim 5.5.12 and Lemma 5.5.8, the process halts in some $s \leq \log^3 d$ number of steps producing $G_s(\boldsymbol{X}_s, \boldsymbol{Y}_s, E_s)$ with the following properties.

- $m(G_s) \geq 2^s \cdot \frac{m(G)}{\log^6 d}$.
- All vertices in $\boldsymbol{X}_s$ are $(\tau - 1, 0.6)$-down-persistent.

- $\Gamma(\boldsymbol{X}_s) \leq \Gamma(\boldsymbol{X})$ and $\Gamma(\boldsymbol{Y}_s) \leq 2^s \Gamma(\boldsymbol{Y})$.
- $D(\boldsymbol{X}_s) \leq D(\boldsymbol{X})$ and $D(\boldsymbol{Y}_s) \leq 2^s D(\boldsymbol{Y})$.

Note that by Lemma 5.5.8 and (c) of Lemma 5.5.2, all vertices in $G_1, \ldots, G_s$ are $(98 + \frac{s\tau}{\sqrt{d}})$-typical. Moreover, by our choice of τ, we have $s\tau \ll \sqrt{d}$ and so all vertices in $G_1, \ldots, G_s$ are 99-typical.

One last time, we remove all vertices in $\boldsymbol{X}_s \cup \boldsymbol{Y}_s$ that are not $(\tau - 1, \log^{-5} d)$-up-persistent and obtain our final graph $H(\boldsymbol{A}, \boldsymbol{B}, E)$. Using a similar argument made above in eq. (5.5), the number of edges that are removed by deleting the non-persistent vertices from $\boldsymbol{Y}_s$ is at most

$$d^{-6c\delta}|\boldsymbol{Y}|D(\boldsymbol{Y}_s) \leq 2^s d^{-6c\delta}|\boldsymbol{Y}|D(\boldsymbol{Y}) \leq 2^s d^{-3c\delta} m(G) \leq d^{-3c\delta} m(G_s) \log^6 d \leq d^{-2c\delta} m(G_s)$$

and an analogous argument bounds the number of edges lost when we remove the non-persistent vertices from $\boldsymbol{X}_s$. Thus we have $m(H) \geq m(G_s)(1 - d^{-c\delta}) \geq 2^s \frac{m(G)}{\log^7 d}$ and this completes the proof of Lemma 5.5.9. $\square$

5.5.3 Using Red/Blue Lemma to Obtain the Final Red/Blue Nice Subgraph

In this section, we prove Lemma 5.3.8 using the violation subgraph $H(\boldsymbol{A}, \boldsymbol{B}, E)$ obtained in the previous section (Lemma 5.5.9) and the red/blue lemma, Lemma 5.3.4. We split into two cases depending on how many edges in H are red.

5.5.3.1 Case 1: At least half the edges of H are red

In this case, we consider the graph H_1 by simply removing all the non-red edges. We claim that H_1 makes progress towards a (σ_1, τ)-nice red subgraph (Definition 5.3.5). Condition (a) holds by definition. Condition (b) is satisfied due to Lemma 5.5.9, condition 5). Condition (e) is satisfied because $\tau - 1 \geq 0.5\sigma_{\boldsymbol{Y}} d^{0.5-7c\delta}$ and $\sigma_{\boldsymbol{Y}} \geq \sigma_{\boldsymbol{X}} = \sigma_1$. We need to establish condition (c) and (d). That is, we need to establish

(c) $\sigma_{\boldsymbol{X}} \cdot \Phi_H(\mathbf{x}) \leq \sqrt{d}$ for all $\mathbf{x} \in \boldsymbol{A}$

(d) $\sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{A}} \Phi_H(\mathbf{x}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$

Let $\boldsymbol{A'} \subseteq \boldsymbol{A}$ be the vertices $\mathbf{x} \in \boldsymbol{A}$ which have $\Phi_H(\mathbf{x}) > \frac{\sqrt{d}}{\sigma_{\boldsymbol{X}}}$. If $|\boldsymbol{A'}| \geq d^{-5c\delta}|\boldsymbol{X}|$, then simply consider $H_1(\boldsymbol{A'}, \boldsymbol{B}, E')$ by deleting all vertices not in $\boldsymbol{A'}$ from $\boldsymbol{A}$. Conditions (a), (b), (e) still hold, and (c) holds by design. Furthermore,

$$\sum_{\mathbf{x} \in \boldsymbol{A'}} \Phi_{H_1}(\mathbf{x}) \geq d^{-5c\delta}|\boldsymbol{X}| \cdot \frac{\sqrt{d}}{\sigma_{\boldsymbol{X}}} \quad \Rightarrow \quad \sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{A'}} \Phi_{H_1}(\mathbf{x}) \geq d^{-5c\delta} \cdot \frac{\varepsilon}{d^{1/2+c\delta}} \cdot n^d \cdot \sqrt{d} = \varepsilon \cdot n^d \cdot d^{-6c\delta}$$

where we used Lemma 5.5.2, part (d) for the lower bound on $|\boldsymbol{X}|$. Note that this implies something slightly stronger than condition (d) above (the exponent of ε is 1).

Therefore, we may assume $|\boldsymbol{A'}| \leq d^{-5c\delta}|\boldsymbol{X}|$. In this case, let $H_1 = (\boldsymbol{A} \setminus \boldsymbol{A'}, \boldsymbol{B}, E')$ where we simply remove the $\boldsymbol{A'}$ vertices. The number of edges this destroys is at most

$$d^{-5c\delta} D(\boldsymbol{A})|\boldsymbol{X}| \leq d^{-5c\delta} D(\boldsymbol{X})|\boldsymbol{X}| \leq d^{-2c\delta} m(G) \leq d^{-c\delta} m(H)$$

where in the second inequality we used $D(\boldsymbol{X}) \leq \Phi(\boldsymbol{X})\Gamma(\boldsymbol{X})$ and the regularity property (Lemma 5.5.2, property (b)) of G. Thus, the number of edges we've discarded is negligible, and condition (c) holds. In particular, the number of edges in H_1 is at least $m(H)/2$. We now prove condition (d) also holds.

Claim 5.5.13. $\sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A'}} \Phi_{H_1}(\mathbf{x}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$.

Proof. For any $\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A'}$, we have $\Phi_{H_1}(\mathbf{x}) \geq \frac{D(\mathbf{x})}{\Gamma(\mathbf{x})}$ and thus $\sum_{\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A'}} \Phi_H(\mathbf{x}) \geq \frac{m(H)/2}{\Gamma(\boldsymbol{A})}$.

Since $\Gamma(\boldsymbol{A}) \leq \Gamma(\boldsymbol{X})$ we have

$$\begin{aligned}\sum_{\mathbf{x} \in \boldsymbol{A} \setminus \boldsymbol{A}'} \Phi_H(\mathbf{x}) &\geq \frac{m(H)}{2\Gamma(\boldsymbol{A})} \geq \frac{2^s \cdot m(G)}{2\Gamma(\boldsymbol{X}) \log^7 d} \geq \frac{d^{-3c\delta} |\boldsymbol{X}| \Phi(\boldsymbol{X}) \Gamma(\boldsymbol{X})}{2\Gamma(\boldsymbol{X}) \log^7 d} \\ &\geq d^{-4c\delta} |\boldsymbol{X}| \Phi(\boldsymbol{X}) \geq d^{-4c\delta} \sum_{\mathbf{x} \in \boldsymbol{X}} \Phi_G(\mathbf{x}). \end{aligned} \tag{5.6}$$

where in the second inequality we used (P1) to lower bound the number of edges in H with that of G. In the third inequality we used the regularity property (property (b) of Lemma 5.5.2), in the fourth we used $d^{c\delta} \gg 2 \log^7 d$ for large enough d, and the fifth inequality uses the trivial upper bound $\Phi(\boldsymbol{X}) \geq \Phi_G(\mathbf{x})$ for all $\mathbf{x} \in \boldsymbol{X}$.

Now we apply the fact (Lemma 5.5.2, condition 1) that $T_{\Phi_\chi}(G)$ is large. Using the coloring $\chi \equiv 1$ for edges in G, we get

$$\sum_{\mathbf{x} \in \boldsymbol{X}} \sqrt{\Phi_G(\mathbf{x})} \geq T_{\Phi_\chi}(G) \geq \varepsilon \cdot d^{-c\delta} \cdot n^d \quad \Rightarrow \quad \mathbb{E}_{\mathbf{x} \in \boldsymbol{X}} [\sqrt{\Phi_G(\mathbf{x})}] \geq \frac{\varepsilon \cdot d^{-c\delta}}{\sigma_{\boldsymbol{X}}}$$

Jensen's inequality gives

$$\mathbb{E}_{\mathbf{x} \in \boldsymbol{X}} [\Phi_G(\mathbf{x})] \geq \frac{\varepsilon^2 \cdot d^{-2c\delta}}{\sigma_{\boldsymbol{X}}^2} \quad \Rightarrow \quad \frac{\sigma_{\boldsymbol{X}}^2}{|\boldsymbol{X}|} \sum_{\mathbf{x} \in \boldsymbol{X}} \Phi_G(\mathbf{x}) \geq \varepsilon^2 d^{-2c\delta} \quad \Rightarrow \sigma_{\boldsymbol{X}} \sum_{\mathbf{x} \in \boldsymbol{X}} \Phi_G(\mathbf{x}) \geq \varepsilon^2 d^{-2c\delta} n^d$$

Plugging into eq. (5.6) proves the claim. $\square$

5.5.3.2 Case 2: At most half the edges of H are red

In this case we invoke the Red/Blue lemma, Lemma 5.3.4 to obtain a violation subgraph $H_2 = (\boldsymbol{L}, \boldsymbol{R}, E')$ with the following key properties.

(P1) All edges are blue and $m(H) \geq 2^s \frac{m(G)}{7 \log^7 d}$.

(P2) $\Gamma(\boldsymbol{R}) \leq \Gamma(\boldsymbol{B}) \leq 2^s \cdot \Gamma(\boldsymbol{Y})$.

(P3) $D(\boldsymbol{R}) \leq D(\boldsymbol{B}) \leq 2^s \cdot D(\boldsymbol{Y})$.

We claim that H_s makes progress towards a (σ_2, τ)-nice blue subgraph (Definition 5.3.6). Condition (a) holds by definition. Condition (d) is satisfied because $\tau \geq 0.5\sigma_{\boldsymbol{Y}} d^{0.5-7c\delta}$ and $\sigma_{\boldsymbol{Y}} = \sigma_2$. We need to establish condition (b) and (c). That is, we need to establish

(b) $\sigma_{\boldsymbol{Y}} \cdot \Phi_H(\mathbf{y}) \leq \sqrt{d}$ for all $\mathbf{x} \in \boldsymbol{R}$

(c) $\sigma_{\boldsymbol{Y}} \sum_{\mathbf{y} \in \boldsymbol{R}} \Phi_H(\mathbf{y}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$

As in Case 1, we begin by removing low degree vertices. Let $\boldsymbol{R}' \subseteq \boldsymbol{R}$ be the vertices $\mathbf{y} \in \boldsymbol{R}$ which have $\Phi_H(\mathbf{y}) > \frac{\sqrt{d}}{\sigma_{\boldsymbol{Y}}}$. If $|\boldsymbol{R}'| \geq d^{-5c\delta}|\boldsymbol{Y}|$, then we would just focus on $H_2(\boldsymbol{R}', \boldsymbol{L}, E')$ and this would satisfy (b) and (c) for a very similar reason as in Case 1. And so, we may assume $|\boldsymbol{R}'|$ is smaller than $d^{-5c\delta}|\boldsymbol{Y}|$ and we define $H_2(\boldsymbol{L}, \boldsymbol{R} \setminus \boldsymbol{R}', E')$, and this leads to a negligible decrease in the number of edges. Condition (b) holds by design, and the proof that condition (c) holds is similar. We provide it for completeness.

Claim 5.5.14. $\sigma_{\boldsymbol{Y}} \sum_{\mathbf{y} \in \boldsymbol{R} \setminus \boldsymbol{R}'} \Phi_{H_2}(\mathbf{y}) \geq \varepsilon^2 \cdot n^d \cdot d^{-6c\delta}$.

Proof. For any $\mathbf{y} \in \boldsymbol{R} \setminus \boldsymbol{R}'$, we have $\Phi_{H_2}(\mathbf{y}) \geq \frac{D(\mathbf{y})}{\Gamma(\mathbf{y})}$ and thus $\sum_{\mathbf{y} \in \boldsymbol{R} \setminus \boldsymbol{R}'} \Phi_{H_2}(\mathbf{y}) \geq \frac{m(H)/2}{\Gamma(\boldsymbol{R})}$. Since $\Gamma(\boldsymbol{R}) \leq 2^s \cdot \Gamma(\boldsymbol{Y})$ we have

$$\begin{aligned}
\sum_{\mathbf{y} \in \boldsymbol{R} \setminus \boldsymbol{R}'} \Phi_H(\mathbf{y}) &\geq \frac{m(H)}{2\Gamma(\boldsymbol{R})} \geq \frac{2^s \cdot m(G)}{2^s \cdot 14\Gamma(\boldsymbol{Y}) \log^7 d} \geq \frac{d^{-3c\delta}|\boldsymbol{Y}|\Phi(\boldsymbol{Y})\Gamma(\boldsymbol{Y})}{14\Gamma(\boldsymbol{Y}) \log^7 d} \\
&\geq d^{-4c\delta}|\boldsymbol{Y}|\Phi(\boldsymbol{Y}) \geq d^{-4c\delta} \sum_{\mathbf{y} \in \boldsymbol{Y}} \Phi_G(\mathbf{y}). \qquad (5.7)
\end{aligned}$$

where in the second inequality we used Lemma 5.5.9, part 1, to lower bound the number of edges in H with that of G, the original seed graph from Lemma 5.5.2. In the third inequality we used the regularity property (property 2 of Lemma 5.5.2), in the fourth we used $d^{c\delta} \gg 14 \log^7 d$ for large enough d, and the fifth inequality uses the trivial upper bound $\Phi(\boldsymbol{Y}) \geq \Phi_G(\mathbf{y})$ for all $\mathbf{y} \in \boldsymbol{Y}$.

The rest of the proof is the same as Case 1 except we apply the coloring $\chi \equiv 0$ for edges in G. We omit this very similar calculation. $\square$

These two cases conclude the proof of Lemma 5.3.8. All that remains is to prove the Red/Blue lemma, Lemma 5.3.4 and the Persist-or-Blow-up lemma, Lemma 5.5.8. We prove these in the subsequent two sections, and both of these use the translation of violation subgraphs idea.

5.6 Proof of the Red/Blue Lemma, Lemma 5.3.4

Let us recall the red/blue lemma.

Lemma 5.6.1 (Red/Blue Lemma). *Let $G(\boldsymbol{X}, \boldsymbol{Y}, E)$ be a violation subgraph and $1 \leq \ell \leq \sqrt{d}/\log^5(d/\varepsilon)$ be a walk length such that the following hold.*

1. *At most half the edges are red for walk length ℓ.*
2. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are $(\ell, \log^{-5} d)$-up-persistent.*
3. *All vertices in $\boldsymbol{X} \cup \boldsymbol{Y}$ are 99-typical.*

Then there exists another violation subgraph $H(\boldsymbol{L}, \boldsymbol{R}, E')$ such that

1. *All edges are blue for walk length ℓ and $m(H) \geq m(G)/7$.*
2. *$\Gamma(\boldsymbol{L}) \leq \Gamma(\boldsymbol{X})$ and $\Gamma(\boldsymbol{R}) \leq \Gamma(\boldsymbol{Y})$.*
3. *$D(\boldsymbol{L}) \leq D(\boldsymbol{X})$ and $D(\boldsymbol{R}) \leq D(\boldsymbol{Y})$.*

Proof. We first recall the definition of $p_{\mathbf{x},\ell}(\mathbf{x}')$ in Definition 5.2.23. For a fixed $\mathbf{x}$, consider the process of sampling a hypercube $\mathbf{H} \sim \mathbb{H}(\mathbf{x})$ and then sampling $\mathbf{z} \sim \mathcal{U}_{\mathbf{H},\ell}(\mathbf{x})$. Recall from Fact 5.2.11 that this is one of three equivalent ways of expressing our random walk distribution. Given $\mathbf{x}, \mathbf{x}', \ell$, we have

$$p_{\mathbf{x},\ell}(\mathbf{x}') = \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \in \mathbf{H}_{100} \text{ and } \mathbf{z} = \mathbf{x}'\right].$$

We use these values to set up a flow problem as follows.

Recall the definition of red and blue edges (Definition 5.3.2 and Definition 5.3.3). Let B

denote the set of all edges in the fully augmented hypergrid that are blue for walk length ℓ. For every non-red edge $(\mathbf{x}, \mathbf{y})$ of G and every shift $\mathbf{s} \in \operatorname{supp}(\mathcal{US}_\ell(\mathbf{x}))$, if the edge $e = (\mathbf{x}+\mathbf{s}, \mathbf{y}+\mathbf{s})$ is blue, then we put $p_{\mathbf{x},\ell}(\mathbf{x}+\mathbf{s})$ units of flow on e.

Claim 5.6.2. *Every non-red edge of G inserts at least 0.95 units of flow in B.*

Proof. Fix a non-red edge $(\mathbf{x}, \mathbf{y})$, and let i denote its dimension. Generate $\mathbf{H} \sim \mathbb{H}(\mathbf{x})$ and $\mathbf{s} \sim \mathcal{US}_{\mathbf{H},\ell}(\mathbf{x})$. Note that it is equivalent to directly sample $\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{x})$. We then consider the random edge $e = (\mathbf{x}+\mathbf{s}, \mathbf{y}+\mathbf{s})$. We set $\mathbf{x}' = \mathbf{x}+\mathbf{s}$ and $\mathbf{y}' = \mathbf{y}+\mathbf{s}$. Let us define the following series of events. (i) $\mathcal{E}_1$: $\mathbf{s}_i = 0$. (ii) $\mathcal{E}_2$: $f(\mathbf{x}') = 1$. (iii) $\mathcal{E}_3$: $f(\mathbf{y}') = 0$. (iv) $\mathcal{E}_4$: at least half of $I(\mathbf{x}', \mathbf{y}')$ is *not* ℓ-mostly-zero-below, (v) $\mathcal{E}_5$: $\mathbf{x}, \mathbf{x}' \in \mathbf{H}_{100}$. We will show that whenever $\mathcal{E}_2$, $\mathcal{E}_3$, and $\mathcal{E}_4$ occur, the edge $(\mathbf{x}', \mathbf{y}')$ is blue by definition. Therefore, recalling the definition of $p_{\mathbf{x},\ell}(\mathbf{x}')$, the edge $(\mathbf{x}, \mathbf{y})$ inserts at least $\mathbb{P}[\wedge_{j=1}^5 \mathcal{E}_j]$ units of flow in B. Subsequently, we will show that the probability of this event is at least 0.95 and this will prove the claim.

Since $\|\mathbf{s}\|_0 \leq \ell \leq \sqrt{d}$, we have $\mathbb{P}[\mathcal{E}_1] \geq 1 - 1/\sqrt{d}$. Since $\mathbf{x}$ is $(\ell, \log^{-5} d)$-up-persistent, $\mathbb{P}[\mathcal{E}_2] \geq 1 - \log^{-5} d$. Note that conditioned in $\mathcal{E}_1$, the distribution on $\mathbf{y}+\mathbf{s}$ is identical to $\mathcal{U}_\ell(\mathbf{y})$. Thus, since $\mathbf{y}$ is $(\ell, \log^{-5} d)$-up-persistent$\mathbb{P}[\mathcal{E}_3 \mid \mathcal{E}_1] \geq 1 - \log^{-5} d$. By a union bound

$$\mathbb{P}[\mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq 1 - 3\log^{-5} d. \tag{5.8}$$

To deal with $\mathcal{E}_4$, we bring in the non-redness of our edge $(\mathbf{x}, \mathbf{y})$. By definition,

$$\mathbb{P}_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} \mathbb{P}_{\mathbf{z}' \sim \mathcal{U}_\ell(\mathbf{z})}[\mathbf{z}' \text{ is not } \ell\text{-}\mathsf{mzb}] \geq 0.99$$

In terms of shifts, we can express this bound as

$$\mathbb{P}_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} \mathbb{P}_{\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{z})}[\mathbf{z}+\mathbf{s} \text{ is not } \ell\text{-}\mathsf{mzb}] \geq 0.99$$

Since the probability of $\mathcal{E}_1$ is at least $1 - o(1)$, we have

$$\mathbb{P}_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})} \mathbb{P}_{\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{z})}[\mathbf{z} + \mathbf{s} \text{ is not } \ell\text{-}\mathsf{mzb} \mid \mathcal{E}_1] \geq 0.98$$

Note that conditioned in $\mathcal{E}_1$, the distributions $\mathcal{US}_\ell(\mathbf{z})$ and $\mathcal{US}_\ell(\mathbf{x})$ are identical. Hence,

$$\mathbb{P}_{\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{x})} \mathbb{P}_{\mathbf{z} \in I(\mathbf{x},\mathbf{y})}[\mathbf{z} + \mathbf{s} \text{ is not } \ell\text{-}\mathsf{mzb} \mid \mathcal{E}_1] \geq 0.98$$

Let $X_\mathbf{s}$ be the fraction of points in $I(\mathbf{x} + \mathbf{s}, \mathbf{y} + \mathbf{s})$ that are not ℓ-mzb. By linearity of expectation, $\mathbb{E}_\mathbf{s}[X_\mathbf{s} \mid \mathcal{E}_1] \geq 0.98$. Hence $\mathbb{E}_\mathbf{s}[1 - X_\mathbf{s} \mid \mathcal{E}_1] \leq 0.02$ and by Markov's inequality, $\mathbb{P}_\mathbf{s}[1 - X_\mathbf{s} > 0.5 \mid \mathcal{E}_1] \leq 1/50$. Hence, $\mathbb{P}_\mathbf{s}[X_\mathbf{s} \geq 0.5 \mid \mathcal{E}_1] \geq 49/50 = .98$. Since $\mathbb{P}[\mathcal{E}_1] = 1 - o(1)$, we have $\mathbb{P}[\mathcal{E}_4] = \mathbb{P}_\mathbf{s}[X_\mathbf{s} \geq 0.5] \geq 0.97$.

Combining with eq. (5.8), we have $\mathbb{P}[\wedge_{j=1}^4 \mathcal{E}_j] \geq 0.96$. When $\wedge_{j=1}^4 \mathcal{E}_j$ occurs, the edge $(\mathbf{x}', \mathbf{y}')$ is a violated edge and at least half of $I(\mathbf{x}', \mathbf{y}')$ is not ℓ-mzb. For $\mathbf{z}' \in I(\mathbf{x}', \mathbf{y}')$ that is not ℓ-mzb, by definition $\mathbb{P}_{\mathbf{w} \sim \mathcal{D}_\ell(\mathbf{z}')}[f(\mathbf{w}) = 1] \geq 0.1$. Hence,

$$\mathbb{P}_{\mathbf{z}' \in_R I(\mathbf{x}',\mathbf{y}')} \mathbb{P}_{\mathbf{w} \sim \mathcal{D}_\ell(\mathbf{z}')}[f(\mathbf{w}) = 1] \geq 0.5 \times 0.1 \geq 0.01$$

We conclude that $(\mathbf{x}', \mathbf{y}')$ is blue, whenever $\wedge_{j=1}^4 \mathcal{E}_j$ occurs.

Stepping back, with probability at least 0.96 over the shift $\mathbf{s} \sim \mathcal{US}_\ell(\mathbf{x})$, the edge $(\mathbf{x} + \mathbf{s}, \mathbf{y} + \mathbf{s})$ is blue. Finally, since all points in $\boldsymbol{X}$ are 99-typical, we have $\mathbb{P}[\mathbf{x} \in \mathbf{H}_{99}] \geq 1 - (\varepsilon/d)^5$, and conditioned on this event we have $\mathbf{x}' \in \mathbf{H}_{100}$ since $\ell \ll \sqrt{d}$. Together, we get $\mathbb{P}[\mathcal{E}_5] \geq 1 - 2(\varepsilon/d)^5 \geq 0.99$. Thus, by a union bound $\mathbb{P}[\wedge_{j=1}^5 \mathcal{E}_j] \geq 0.95$ and so the amount of flow that $(\mathbf{x}, \mathbf{y})$ inserts is at least 0.95. $\square$

Let $E' \subseteq B$ denote the set of blue edges which receive non-zero flow. Let $H(\boldsymbol{L}, \boldsymbol{R}, E')$ denote the bipartite graph on these edges. Since $\ell \leq \sqrt{d}/\log^5(d/\varepsilon)$, by the reversibility Lemma 5.2.24, $p_{\mathbf{x},\ell}(\mathbf{x}') \leq 2p_{\mathbf{x}',\ell}(\mathbf{x})$ for any $\mathbf{x} \in \boldsymbol{X}$, $\mathbf{x}' \in \boldsymbol{L}$ and $p_{\mathbf{y},\ell}(\mathbf{y}') \leq 2p_{\mathbf{y}',\ell}(\mathbf{y})$ for any $\mathbf{y} \in \boldsymbol{Y}$, $\mathbf{y}' \in \boldsymbol{R}$. Using this bound we're able to establish the desired capacity constraints on the flow as follows.

Claim 5.6.3 (Edge Congestion)**.** *The total flow on an edge* $(\mathbf{x}', \mathbf{y}') \in B$ *is at most* 2.

Proof. By construction, the total flow on an edge $(\mathbf{x}', \mathbf{y}')$ is at most

$$\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq 2 \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 2$$

since $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 1$. □

Claim 5.6.4 (Vertex Congestion)**.** *The following hold.*

1. *The total amount of flow through a vertex* $\mathbf{x}' \in \boldsymbol{L}$ *is at most* $2D(\boldsymbol{X})$.
2. *The total amount of flow through a vertex* $\mathbf{y}' \in \boldsymbol{R}$ *is at most* $2D(\boldsymbol{Y})$.
3. *For all* $i \in [d]$, *the total amount of* i*-flow through a vertex* $\mathbf{x}' \in \boldsymbol{L}$ *is at most* $2\Gamma_i(\boldsymbol{X})$.
4. *For all* $i \in [d]$, *the total amount of* i*-flow through a vertex* $\mathbf{y}' \in \boldsymbol{R}$ *is at most* $2\Gamma_i(\boldsymbol{Y})$.

Proof. The total flow through a vertex $\mathbf{x}' \in \boldsymbol{L}$ is at most

$$\begin{aligned}
\sum_{(\mathbf{x},\mathbf{y}) \in E} p_{\mathbf{x},\ell}(\mathbf{x}') &\leq D(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \quad (\text{max degree of } \mathbf{x} \in \boldsymbol{X} \text{ is } D(\boldsymbol{X})) \\
&\leq 2D(\boldsymbol{X}) \sum_{\mathbf{x} \in [n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \quad (\text{since } p_{\mathbf{x},\ell}(\mathbf{x}') \leq 2p_{\mathbf{x}',\ell}(\mathbf{x})) \\
&\leq 2D(\boldsymbol{X}) \quad \Big(\text{since} \sum_{\mathbf{x} \in [n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 1\Big)
\end{aligned}$$

and an analogous argument proves (2). For a coordinate $i \in [d]$, let $E_i \subseteq E$ denote the set of i-edges in G. The total i-flow through a vertex $\mathbf{x}' \in \boldsymbol{L}$ is at most

$$\sum_{(\mathbf{x},\mathbf{y})\in E_i} p_{\mathbf{x}}(\mathbf{x}') \leq \Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x}\in\boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \quad (\text{max } i\text{-degree of } \mathbf{x} \in \boldsymbol{X} \text{ is } \Gamma_i(\boldsymbol{X}))$$
$$\leq 2\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x}\in[n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \quad (\text{since } p_{\mathbf{x},\ell}(\mathbf{x}') \leq 2p_{\mathbf{x}',\ell}(\mathbf{x}))$$
$$\leq 2\Gamma_i(\boldsymbol{X}) \quad (\text{since } \sum_{\mathbf{x}\in[n]^d} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 1)$$

and an analogous argument proves (4). □

By Claim 5.6.2 and the fact that at least half the edges in G are not red, the total amount of flow is at least $m(G)/3$ and this flow satisfies the constraints listed in Claim 5.6.3 and Claim 5.6.4. Thus, dividing by 2 yields a flow of value $m(G)/6$ satisfying the following.

1. The flow on every edge is at most 1.
2. The total flow through any vertex in $\boldsymbol{L}$ is at most $D(\boldsymbol{X})$. The total i-flow through any vertex in $\boldsymbol{L}$ is at most $\Gamma_i(\boldsymbol{X})$.
3. The total flow through any vertex in $\boldsymbol{R}$ is at most $D(\boldsymbol{Y})$. The total i-flow through any vertex in $\boldsymbol{R}$ is at most $\Gamma_i(\boldsymbol{Y})$.

By integrality of flow, there exists an integral flow of at least $\lfloor m(G)/6 \rfloor \geq m(G)/7$ units satisfying the same capacity constraints. By item (1) above, the integral flow is a subgraph containing at least $m/7$ edges and satisfying the desired constraints listed in the lemma statement. □

5.7 Proof of the 'Persist-or-Blow-Up' Lemma, Lemma 5.5.8

Let us recall the 'Persist-or-Blow-Up' lemma.

Lemma 5.7.1 (Persist-or-Blow-up Lemma). *Consider a violation subgraph $G = (\boldsymbol{X}, \boldsymbol{Y}, E)$*

such that all vertices in G are c-typical where $c \leq 99$ and $(\ell, \log^{-5} d)$-up persistent where $1 \leq \ell \leq \sqrt{d}/\log^5(d/\varepsilon)$. Then, there exists a violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E')$ where all vertices are $(c + \frac{\ell}{\sqrt{d}})$-typical and satisfying one of the following conditions.

1. *Down-persistent case:*
   (a) *All vertices in $\boldsymbol{X}'$ are $(\ell, 0.6)$-down persistent.*
   (b) *$m(G') \geq m(G)/\log^5 d$.*
   (c) *$D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, and $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$*
   (d) *$D(\boldsymbol{Y}') \leq D(\boldsymbol{Y})$, and $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq \Gamma_i(\boldsymbol{Y})$.*
2. *Blow-up case:*
   (a) *$m(G') \geq 2(1 - 3\log^{-3} d) \cdot m(G)$.*
   (b) *$D(\boldsymbol{X}') \leq D(\boldsymbol{X})$, and $\forall i \in [d], \Gamma_i(\boldsymbol{X}') \leq \Gamma_i(\boldsymbol{X})$*
   (c) *$D(\boldsymbol{Y}') \leq 2D(\boldsymbol{Y})$, and $\forall i \in [d]$, $\Gamma_i(\boldsymbol{Y}') \leq 2\Gamma_i(\boldsymbol{Y})$.*

We first recall the definition of $p_{\mathbf{x},\ell}(\mathbf{x}')$ in Definition 5.2.23. For a fixed $\mathbf{x}$, consider the process of sampling a hypercube $\mathbf{H} \sim \mathbb{H}(\mathbf{x})$ and then sampling $\mathbf{z} \sim \mathcal{U}_{\mathbf{H},\ell}(\mathbf{x})$. Recall from Fact 5.2.11 that this is one of three equivalent ways of expressing our random walk distribution. Given $\mathbf{x}, \mathbf{x}', \ell$, we have

$$p_{\mathbf{x},\ell}(\mathbf{x}') = \mathbb{P}\left[\mathbf{x}, \mathbf{x}' \in \mathbf{H}_{100} \text{ and } \mathbf{z} = \mathbf{x}'\right].$$

We use these values to set up a flow problem as follows. For every edge $(\mathbf{x}, \mathbf{y})$ of G and $\mathbf{s} \in \operatorname{supp}(\mathcal{US}_\ell(\mathbf{x}))$, if $e = (\mathbf{x} + \mathbf{s}, \mathbf{y} + \mathbf{s})$ is a violation, then we put $p_{\mathbf{x}}(\mathbf{x} + \mathbf{s})$ units of flow on e. The flow, denoted $\mathcal{F}$, is supported on a violation subgraph $G' = (\boldsymbol{X}', \boldsymbol{Y}', E)$. Note that by Claim 5.2.22, all vertices in G' are $(c + \frac{\ell}{\sqrt{d}})$-typical.

Claim 5.7.2. *Every edge of G inserts at least $1 - \log^{-4} d$ units of flow.*

Proof. The proof of this claim is similar to that of Claim 5.6.2. Fix an edge $(\mathbf{x}, \mathbf{y}) \in G$ and let this be an i-edge. Generate $\mathbf{H} \sim \mathbb{H}(\mathbf{x})$ and a shift $\mathbf{s} \sim \mathcal{US}_{\mathbf{H},\ell}(\mathbf{x})$, and let $\mathbf{x}' = \mathbf{x} + \mathbf{s}$

and $\mathbf{y}' = \mathbf{y} + \mathbf{x}$. Consider the events: (i) $\mathcal{E}_1$: $\mathbf{s}_i = 0$, (ii) $\mathcal{E}_2$: $f(\mathbf{x}') = 1$, (iii) $\mathcal{E}_3$: $f(\mathbf{y}') = 0$, (iv)$\mathcal{E}_4$: $\mathbf{x}, \mathbf{x}' \in \mathbf{H}_{100}$. Note that the total flow inserted by $(\mathbf{x}, \mathbf{y})$ is at least $\mathbb{P}[\wedge_{i=1}^4 \mathcal{E}_i]$. $\mathbb{P}[\mathcal{E}_1] \geq 1 - 1/\sqrt{d}$, since $\|\mathbf{s}\|_0 \leq \ell \leq \sqrt{d}$. Since $\mathbf{x}, \mathbf{y}$ are both $(\ell, \log^{-5} d)$-up-persistent and $f(\mathbf{x}) = 1$, $f(\mathbf{y}) = 0$, we get $\mathbb{P}[\mathcal{E}_2], \mathbb{P}[\mathcal{E}_3] \geq 1 - \frac{1}{\log^5 d}$. Finally, since $\mathbf{x}$ is 99-typical, with probability $1 - (\varepsilon/d)^5$ we have $\mathbf{x} \in \mathbf{H}_{99}$ which implies $\mathbf{x}' \in \mathbf{H}_{100}$ since $\ell \ll \sqrt{d}$. Thus by a union bound, $\mathbb{P}[\wedge_{i=1}^5 \mathcal{E}_i] \geq 1 - 2\log^{-5} d - 1/\sqrt{d} - (\varepsilon/d)^5 \geq 1 - \log^{-4} d$. □

Claim 5.7.3 (Edge Congestion)**.** *The flow on any edge* $(\mathbf{x}', \mathbf{y}')$ *is at most* $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq (1 + \log^{-3} d)$.

Proof. Consider an edge $(\mathbf{x}', \mathbf{y}')$, which receives flow from some $(\mathbf{x}, \mathbf{y})$ in G. Flow is inserted by translations of edges, so $\mathbf{y} - \mathbf{x} = \mathbf{y}' - \mathbf{x}'$. Hence, for a given $\mathbf{x}$, there exists a unique $\mathbf{y}$ such that $(\mathbf{x}, \mathbf{y})$ inserts flow on $(\mathbf{x}', \mathbf{y}')$. By construction, the flow inserted is $p_{\mathbf{x},\ell}(\mathbf{x}')$. Thus, the total flow that $(\mathbf{x}', \mathbf{y}')$ receives is at most $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\tau}(\mathbf{x}')$. The RHS bound holds by Lemma 5.2.24 and observing that $\sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \leq 1$. □

Claim 5.7.4 (Vertex Congestion)**.** *The following hold.*

1. *For any* $\mathbf{x}' \in \boldsymbol{X}'$*, the total flow on edges incident to* $\mathbf{x}'$ *is at most*

$$D(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq D(\boldsymbol{X})(1 + \log^{-3} d).$$

2. *For any* $\mathbf{x}' \in \boldsymbol{X}'$*, the total* i*-flow on edges incident to* $\mathbf{x}'$ *is at most*

$$\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \leq \Gamma_i(\boldsymbol{X})(1 + \log^{-3} d).$$

3. *For any* $\mathbf{y}' \in \boldsymbol{Y}'$*, the total flow on edges incident to* $\mathbf{y}'$ *is at most*

$$D(\boldsymbol{Y}) \sum_{\mathbf{y} \in \boldsymbol{Y}} p_{\mathbf{y},\ell}(\mathbf{y}') \leq D(\boldsymbol{Y})(1 + \log^{-3} d).$$

4. For any $\mathbf{y}' \in \boldsymbol{Y}'$, the total i-flow on edges incident to $\mathbf{y}'$ is at most

$$\Gamma_i(\boldsymbol{Y}) \sum_{\mathbf{y} \in \boldsymbol{Y}} p_{\mathbf{y},\ell}(\mathbf{y}') \leq \Gamma_i(\boldsymbol{Y})(1 + \log^{-3} d).$$

Proof. Consider $\mathbf{x}' \in \boldsymbol{X}'$. All the i-flow inserted on edges incident to $\mathbf{x}'$ comes from i-edges $(\mathbf{x}, \mathbf{y})$ in G. Every i-edge in G inserts flow on at most a single edge incident to $\mathbf{x}'$ and there are at most $\Gamma_i(\boldsymbol{X})$ i-edges incident to any vertex $\mathbf{x} \in \boldsymbol{X}$. Hence, the total i-flow inserted by a $\mathbf{x} \in \boldsymbol{X}$ through $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X}) \cdot p_{\mathbf{x},\tau}(\mathbf{x}')$. Thus, summing over all $\mathbf{x} \in \boldsymbol{X}$ and using the reversibility Lemma 5.2.24 shows that the total i-flow on edges incident to $\mathbf{x}'$ is at most

$$\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x},\tau}(\mathbf{x}') \leq (1 + \log^{-3} d)\Gamma_i(\boldsymbol{X}) \sum_{\mathbf{x} \in \boldsymbol{X}} p_{\mathbf{x}',\tau}(\mathbf{x}) \leq (1 + \log^{-3} d)\Gamma_i(\boldsymbol{X})$$

and this proves (2). The proof of (1) is identical, with $D(\boldsymbol{X})$ replacing $\Gamma_i(\boldsymbol{X})$, and statements (3) and (4) have analogous proofs. □

We now come to a key definition in our analysis.

Definition 5.7.5 (Heavy Vertices). *A vertex $\mathbf{x}' \in \boldsymbol{X}'$ is called* heavy *if it satisfies any of the following.*

1. *There is an edge $(\mathbf{x}', \mathbf{y}')$ receiving at least $1/2$ units of flow.*
2. *The total flow on edges incident to $\mathbf{x}'$ is at least $D(\boldsymbol{X})/2$.*
3. *There exists $i \in [d]$ such that the total i-flow on edges incident to $\mathbf{x}'$ is at least $\Gamma_i(\boldsymbol{X})/2$.*

We refer to the flow passing through heavy vertices as the *heavy flow*.

Claim 5.7.6. *All heavy vertices are $(\ell, 0.6)$-down persistent.*

Proof. Consider a heavy vertex $\mathbf{x}'$. That is, $\mathbf{x}'$ satisfies one of the three conditions listed in Definition 5.7.5. Suppose it satisfies the first condition: there is some violated edge $(\mathbf{x}', \mathbf{y}')$ receiving at least $1/2$ units of flow. By Claim 5.7.3, the total flow on $(\mathbf{x}', \mathbf{y}')$ is at most

$\sum_{\mathbf{x}\in\boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}')$. Hence, $\sum_{\mathbf{x}\in\boldsymbol{X}} p_{\mathbf{x},\ell}(\mathbf{x}') \geq 1/2$. In fact, observe that we can prove the exact same inequality if $\mathbf{x}'$ satisfies the second or third condition of Definition 5.7.5, by using the upper bound given by the LHS of items (1) and (2), respectively, of Claim 5.7.4. Now, applying the reversibility Lemma 5.2.24, we have $(1+\log^{-3} d)\sum_{\mathbf{x}\in\boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \geq 1/2$. Note that $f(\mathbf{x}) = 1$ for all $\mathbf{x} \in \boldsymbol{X}$. Hence,

$$\mathbb{P}_{\mathbf{z}\sim\mathcal{D}_\ell(\mathbf{x}')}[f(\mathbf{z}) = 1] \geq \sum_{\mathbf{x}\in\boldsymbol{X}} p_{\mathbf{x}',\ell}(\mathbf{x}) \geq \frac{1}{2(1+\log^{-3} d)} \geq 0.4 \tag{5.9}$$

and so $\mathbf{x}'$ is $(\ell, 0.6)$-down-persistent. □

We are now set up to complete the proof. For convenience, we use m to denote $m(G)$. We refer to the flow on edges incident to heavy vertices as the *heavy flow*. We let $G_H(\boldsymbol{X}_H, \boldsymbol{Y}_H, E_H)$ denote the bipartite graph of all edges incident to heavy vertices, that is, $\boldsymbol{X}_H$ is the set of all heavy vertices. We refer to the flow on edges incident to non-heavy vertices as the *light flow*. We let $G_L(\boldsymbol{X}_L, \boldsymbol{Y}_L, E_L)$ denote the bipartite graph of all edges incident to non-heavy vertices, that is, $\boldsymbol{X}_L = \boldsymbol{X}' \setminus \boldsymbol{X}_H$ is the set of all non-heavy vertices. We split into two cases based on the amount of heavy flow.

5.7.1 Case 1: The total amount of heavy flow is at least $\frac{m}{\log^4 d}$

Note that by Claim 5.7.6, all vertices in $\boldsymbol{X}_H$ are $(\ell, 0.6)$-down persistent.

By Claim 5.7.3 and Claim 5.7.4, the heavy flow satisfies the following capacity constraints.

1. The flow on every edge is at most $(1 + \log^{-3} d)$.
2. For every $\mathbf{x}' \in \boldsymbol{X}_H$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})(1+\log^{-3} d)$ and the total i-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})(1 + \log^{-3} d)$.
3. For every $\mathbf{y}' \in \boldsymbol{Y}_H$, the total flow on edges incident to $\mathbf{y}'$ is at most $D(\boldsymbol{Y})(1+\log^{-3} d)$ and the total i-flow on edges incident to $\mathbf{y}'$ is at most $\Gamma_i(\boldsymbol{Y})(1 + \log^{-3} d)$.

Let us divide the flow by $(1+\log^{-3} d)$. Thus, we now have at least $\frac{m}{(1+\log^{-3} d)\log^4 d} \geq \frac{m}{\log^5 d}$ units of flow satisfying the following capacity constraints.

1. The flow on every edge is at most one.

2. For every $\mathbf{x}' \in \boldsymbol{X}_H$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})$ and the total i-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})$.

3. For every $\mathbf{y}' \in \boldsymbol{Y}_H$, the total flow on edges incident to $\mathbf{y}'$ is at most $D(\boldsymbol{Y})$ and the total i-flow on edges incident to $\mathbf{y}'$ is at most $\Gamma_i(\boldsymbol{Y})$.

By integrality of flow, there is an integral flow of at least $\frac{m}{\log^5 d}$ units satisfying the above constraints. By condition (1) above, this integral flow is a subgraph of G_H with at least $\frac{m}{\log^5 d}$ edges, and satisfying the degree bounds listed in (1c) and (1d) of the lemma statement. Thus, this subgraph satisfies case (1) of the lemma statement.

5.7.2 Case 2: The total amount of heavy flow is at most $\frac{m}{\log^4 d}$

By Claim 5.7.2, the total flow is at least $m(1-\log^{-4} d)$ units. Thus, after removing the heavy flow, the remaining light flow is at least $m(1-2\log^{-4} d)$ units. The light flow satisfies the following capacity constraints.

1. Every edge has at most $1/2$ units of flow.

2. For every $\mathbf{x}' \in \boldsymbol{X}_L$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})/2$ and the total i-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})/2$.

3. For every $\mathbf{y}' \in \boldsymbol{Y}_L$, the total flow on edges incident to $\mathbf{y}'$ is at most $(1+\log^{-3} d)D(\boldsymbol{Y})$ and the total i-flow on edges incident to $\mathbf{y}'$ is at most $(1+\log^{-3} d)\Gamma_i(\boldsymbol{Y})$.

Items (1) and (2) are simply by Definition 5.7.5 since all vertices in $\boldsymbol{X}_L$ are not heavy. Item (3) follows from RHS bound on the vertex congestion in Claim 5.7.4.

We now by rescale the flow by multiplying it by $\frac{2}{1+\log^{-3} d}$. We now have $2m\frac{(1-2\log^{-4} d)}{1+\log^{-3} d} \geq 2m(1-2\log^{-3} d)$ units of flow with the following capacity constraints:

1. Every edge has at most one unit of flow.
2. For every $\mathbf{x}' \in \boldsymbol{X}_L$, the total flow on edges incident to $\mathbf{x}'$ is at most $D(\boldsymbol{X})$ and the total i-flow on edges incident to $\mathbf{x}'$ is at most $\Gamma_i(\boldsymbol{X})$.
3. For every $\mathbf{y}' \in \boldsymbol{Y}_L$, the total flow on edges incident to $\mathbf{y}'$ is at most $2D(\boldsymbol{Y})$ and the total i-flow on edges incident to $\mathbf{y}'$ is at most $2\Gamma_i(\boldsymbol{Y})$.

By integrality of flow, we obtain an integral flow of at least $\lfloor 2m(1-3\log^{-4} d)\rfloor \geq 2m(1-3\log^{-3} d)$ units satisfying the same constraints listed above. In particular, the flow on any edge is at most one and so the integral flow is a violation subgraph with at least $2m(1-3\log^{-3} d)$ edges and satisfying the degree bounds listed in case (2) of the lemma statement.

5.8 Deferred Proofs

5.8.1 Equivalence of the Walk Distributions: Proof of Fact 5.2.11

Proof. Fix a pair (u, v) in $[n]^d$ where $u \preceq v$. We will show that the probability of sampling this pair from each distribution is the same. Let $S = \{i \in [d] \colon v_i > u_i\}$. Note that $u_j = v_j$ for all $j \neq S$. The probability of sampling the pair (u, v) from the distribution described in item (1) of Fact 5.2.11 is computed as follows.

$$\begin{aligned}
&\mathbb{P}_{\mathbf{x}\in_R [n]^d,\, \mathbf{y}\sim\mathcal{U}_\tau(\mathbf{x})}[(\mathbf{x},\mathbf{y}) = (u,v)] \\
&= \frac{1}{n^d} \sum_{R\supseteq S \,:\, |R|=\tau} \binom{d}{\tau}^{-1} \prod_{i\in S} \mathbb{P}[c_i = v_i \mid \mathbf{x} = u] \prod_{i\in R\setminus S} \mathbb{P}[c_i \leq u_i \mid \mathbf{x} = u]. \qquad (5.10)
\end{aligned}$$

Recall the distribution of q_i, I_i, c_i from Definition 5.1.1. Consider $i \in S$ and let $d_i := \min(v_i - u_i, n - (v_i - u_i))$. Note that conditioned on q_i, the total number of intervals $I_i \ni u_i$

is 2^{q_i} and the number of such intervals that contain v_i is $\max(0, 2^{q_i} - d_i)$. Thus, we have

$$\begin{aligned} i \in S \implies \mathbb{P}[c_i = v_i \mid \mathbf{x} = u] &= \mathbb{E}_{q_i}\Big[\mathbb{P}_{I_i}[v_i \in I_i]\mathbb{P}_{c_i \in I_i}[c_i = v_i \mid v_i \in I_i]\Big] \\ &= \frac{1}{\log n} \sum_{q\colon 2^{q_i} \geq d_i} \frac{2^{q_i} - d_i}{2^{q_i}} \cdot \frac{1}{2^{q_i} - 1} = \frac{1}{2} \cdot \mathbb{E}_{q_i}\left[\frac{\max(0, 2^{q_i} - d_i)}{\binom{2^{q_i}}{2}}\right]. \end{aligned} \tag{5.11}$$

For an interval $I_i \ni u_i$, let I_{i,u_i} denote the prefix of I_i preceding (not including) u_i. Note that conditioned on an interval $I_i \ni u_i$, the probability of choosing $c_i \leq u_i$ is $|I_{i,u_i}|/(2^{q_i} - 1)$. Thus, we have

$$i \in R \setminus S \implies \mathbb{P}[c_i \leq u_i \mid \mathbf{x} = u] = \mathbb{E}_{q_i}\left[\frac{1}{2^{q_i} - 1} \cdot \mathbb{E}_{I_i \ni u_i}[|I_{i,u_i}|]\right] \tag{5.12}$$

We now compute the probability of sampling (u, v) from the distribution described in item (2) of Fact 5.2.11. Recall the distribution of q_i, I_i, a_i, b_i from Definition 5.2.8. For $i \in [d]$, let $\mathcal{E}_i$ be the event that $a_i = u_i$ or $b_i = u_i$. Note that

$$\mathbb{P}[\mathcal{E}_i] = \mathbb{E}_{q_i}\left[\mathbb{P}_{I_i}[I_i \ni u_i]\mathbb{P}_{a_i < b_i \in I_i}[u_i \in \{a_i, b_i\} \mid u_i \in I_i]\right] = \mathbb{E}_{q_i}\left[\frac{2^{q_i}}{n} \cdot \frac{2}{2^{q_i}}\right] = \frac{2}{n}$$

Let $\mathcal{E}_u$ denote the event that $\mathbf{x} = u$. We have

$$\mathbb{P}[\mathcal{E}_u] = \prod_{i=1}^{d} \mathbb{P}[\mathcal{E}_i] \cdot \frac{1}{2^d} = \left(\frac{2}{n}\right)^d \frac{1}{2^d} = \frac{1}{n^d}. \tag{5.13}$$

Let $\mathcal{E}_v$ denote the event that $\mathbf{y} = v$. We have

$$\mathbb{P}\left[\mathcal{E}_v \mid \mathcal{E}_u\right] = \sum_{R \supseteq S \;:\; |R| = \tau} \binom{d}{\tau}^{-1} \prod_{i \in S} \mathbb{P}[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_u] \cdot \prod_{i \in R \setminus S} \mathbb{P}[b_i = u_i \mid \mathcal{E}_u] \tag{5.14}$$

Fix an $i \in S$ and recall $d_i := \min(v_i - u_i, n - (v_i - u_i))$. We have

$$\mathbb{P}[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_u] = \mathbb{P}[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_i] = \frac{\mathbb{P}[a_i = u_i \text{ and } b_i = v_i]}{\mathbb{P}[\mathcal{E}_i]}$$

where the numerator is

$$\mathbb{P}[a_i = u_i \text{ and } b_i = v_i] = \mathbb{E}_{q_i}\left[\mathbb{P}_{I_i}\left[I_i \supseteq [u_i, v_i]\right] \cdot \binom{2^{q_i}}{2}^{-1}\right] = \mathbb{E}_{q_i}\left[\frac{\max(0, 2^{q_i} - d_i)}{n \cdot \binom{2^{q_i}}{2}}\right]$$

and so

$$i \in S \implies \mathbb{P}[a_i = u_i \text{ and } b_i = v_i \mid \mathcal{E}_u] = \frac{1}{2} \cdot \mathbb{E}_{q_i}\left[\frac{\max(0, 2^{q_i} - d_i)}{\binom{2^{q_i}}{2}}\right] \tag{5.15}$$

which is equal to the probability computed in eq. (5.11).

Now fix an $i \in R \setminus S$. Recall the definition of I_{i,u_i}. We have

$$\mathbb{P}[b_i = u_i \mid \mathcal{E}_u] = \mathbb{P}[b_i = u_i \mid \mathcal{E}_i] = \frac{\mathbb{P}[b_i = u_i]}{\mathbb{P}[\mathcal{E}_i]}$$

where

$$\begin{aligned} \mathbb{P}[b_i = u_i] &= \mathbb{E}_{q_i}\mathbb{E}_{I_i}\left[\mathbf{1}(u_i \in I_i)\frac{|I_{u_i}|}{\binom{2^{q_i}}{2}}\right] \\ &= \mathbb{E}_{q_i}\left[\frac{1}{n}\sum_{I_i \ni u_i} |I_{i,u_i}| \binom{2^{q_i}}{2}^{-1}\right] = \frac{2}{n}\mathbb{E}_{q_i}\left[\frac{1}{2^{q_i - 1}} \cdot \mathbb{E}_{I_i \ni u_i}[|I_{i,u_i}|]\right] \end{aligned}$$

and so recalling that $\mathbb{P}[\mathcal{E}_i] = 2/n$ we have

$$i \notin R \setminus S \implies \mathbb{P}[b_i = u_i \mid \mathcal{E}_u] = \mathbb{E}_{q_i}\left[\frac{1}{2^{q_i - 1}} \cdot \mathbb{E}_{I_i \ni u_i}[|I_{i,u_i}|]\right] \tag{5.16}$$

which is equal to the probability computed in eq. (5.12). Combining eq. (5.10), eq. (5.11), eq. (5.12), eq. (5.13), eq. (5.14), eq. (5.15), eq. (5.16), we have

$$\mathbb{P}_{\mathbf{H} \sim \mathbb{H}}\mathbb{P}_{\mathbf{x} \in_R \mathbf{H},\, \mathbf{y} \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})}[(\mathbf{x}, \mathbf{y}) = (u, v)] = \mathbb{P}[\mathcal{E}_u] \cdot \mathbb{P}[\mathcal{E}_v \mid \mathcal{E}_u] = \mathbb{P}_{\mathbf{x} \in_R [n]^d,\, \mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})}[(\mathbf{x}, \mathbf{y}) = (u, v)]$$

and this proves that (1) and (2) of Fact 5.2.11 are equivalent.

To show equivalence of (1) and (3), note that we only need to show that

$$\mathbb{P}_{\mathbf{H}\sim\mathbb{H}(u),\, \mathbf{y}\sim\mathcal{U}_{\mathbf{H},\tau}(u)}[\mathbf{y} = v] = \mathbb{P}_{\mathbf{y}\sim\mathcal{U}_{\tau}(u)}[\mathbf{y} = v] \tag{5.17}$$

This is proven by an analogous calculation. The expression for $\mathbb{P}_{\mathbf{y}\sim\mathcal{U}_{\tau}(u)}[\mathbf{y} = v]$ is given by dropping the $\frac{1}{n^d}$ factor from eq. (5.10) and then plugging in the expressions obtained in eq. (5.11) and eq. (5.12). The quantity $\mathbb{P}_{\mathbf{H}\sim\mathbb{H}(u),\, \mathbf{y}\sim\mathcal{U}_{\mathbf{H},\tau}(u)}[\mathbf{y} = v]$ is precisely $\mathbb{P}[\mathcal{E}_v \mid \mathcal{E}_u]$, and an expression for this is obtained by eq. (5.14) and plugging in the expressions obtained in eq. (5.15) and eq. (5.16). Thus, (1) and (3) are equivalent and this completes the proof. □

5.8.2 Influence and Persistence Proofs

Claim 5.8.1. *If $\widetilde{I}_f > 9\sqrt{d}$, then $\widetilde{I}_f^- > \sqrt{d}$.*

Proof. Theorem 9.1 of [KMS18] asserts that for any $\mathbf{H}$, if $I_{\mathbf{H}} > 6\sqrt{d}$, then $I^-_{\mathbf{H}} > I_{\mathbf{H}}/3$. (This holds for any Boolean hypercube function.) If $\widetilde{I}_f > 9\sqrt{d}$, then by Claim 5.2.14, $\mathbb{E}_{\mathbf{H}}[I_{\mathbf{H}}] > 9\sqrt{d}$. Hence,

$$\begin{aligned} 9\sqrt{d} < \mathbb{E}_{\mathbf{H}}[I_{\mathbf{H}}] &= \mathbb{P}[I_{\mathbf{H}} \le 6\sqrt{d}]\ \mathbb{E}_{\mathbf{H}}[I_{\mathbf{H}} | I_{\mathbf{H}} \le 6\sqrt{d}]\ +\ \mathbb{P}[I_{\mathbf{H}} > 6\sqrt{d}]\ \mathbb{E}_{\mathbf{H}}[I_{\mathbf{H}} | I_{\mathbf{H}} > 6\sqrt{d}] \\ &< 6\sqrt{d} + \mathbb{P}[I_{\mathbf{H}} > 6\sqrt{d}]\mathbb{E}_{H}[3I^-_{\mathbf{H}} | I_{\mathbf{H}} > 6\sqrt{d}] \le 6\sqrt{d} + 3\mathbb{E}_{\mathbf{H}}[I^-_{\mathbf{H}}] \end{aligned}$$

Hence, $\mathbb{E}_{\mathbf{H}}[I^-_{\mathbf{H}}] > \sqrt{d}$. By Claim 5.2.14, $\widetilde{I}^-_f > \sqrt{d}$. □

Claim 5.8.2. *If $\widetilde{I}_f \le 9\sqrt{d}$, then the fraction of vertices that are not (τ, β)-persistent is at most $C_{per}\frac{\tau}{\beta\sqrt{d}}$ where C_{per} is a universal constant.*

Proof. We will analyze the random walk using the distributions described in the first and second bullet point of Fact 5.2.11 and leverage the analysis that [KMS18] use to prove their Lemma 9.3. Let α_{up} denote the fraction of vertices in the fully augmented hypergrid that are not (τ, β)-up-persistent. Using the definition of persistence and Fact 5.2.11, we have

$$\alpha_{up} \cdot \beta < \mathbb{P}_{\mathbf{x} \in_R [n]^d,\, \mathbf{y} \sim \mathcal{U}_\tau(\mathbf{x})}\left[f(\mathbf{x}) \neq f(\mathbf{z})\right] = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}}\left[\mathbb{P}_{\mathbf{x} \in_R \mathbf{H},\, \mathbf{y} \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})}\left[f(\mathbf{x}) \neq f(\mathbf{z})\right]\right]. \quad (5.18)$$

Let $\widehat{\mathcal{U}}_{\mathbf{H},\tau}(\mathbf{x})$ denote the same distribution as $\mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})$ except with the set R being a uar subset of the 0-coordinates of $\mathbf{x}$. I.e. $\widehat{\mathcal{U}}_{\mathbf{H},\tau}(\mathbf{x})$ is the *non-lazy* walk distribution on $\mathbf{H}$. 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 $\mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})$ 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 $\widehat{\mathcal{U}}_{\mathbf{H},\tau}(\mathbf{x})$. For a fixed $\mathbf{H}$ we have

$$\mathbb{P}_{\mathbf{x} \in_R \mathbf{H},\, \mathbf{y} \sim \mathcal{U}_{\mathbf{H},\tau}(\mathbf{x})}\left[f(\mathbf{x}) \neq f(\mathbf{z})\right] \leq \sum_{\ell=0}^{\tau-1} \mathbb{P}\left[f(\mathbf{x}^\ell) \neq f(\mathbf{x}^{\ell+1})\right] \leq \sum_{\ell=0}^{\tau-1} \mathbb{P}\left[f(\widehat{\mathbf{x}}^\ell) \neq f(\widehat{\mathbf{x}}^{\ell+1})\right]. \quad (5.19)$$

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 [KMS18], the edge $(\widehat{\mathbf{x}}^\ell, \widehat{\mathbf{x}}^{\ell+1})$ is distributed approximately as a uniform random edge in $\mathbf{H}$. In particular, this implies $\mathbb{P}\left[f(\widehat{\mathbf{x}}^\ell) \neq f(\widehat{\mathbf{x}}^{\ell+1})\right] \leq C \cdot 2I_{\mathbf{H}}/d$ for an absolute constant C. (Note $2I_{\mathbf{H}}/d$ is the probability of a uniform random edge in $\mathbf{H}$ being influential.) Putting eq. (5.18) and eq. (5.19) together yields $\alpha_{up} \leq \frac{4C\tau}{d}\mathbb{E}_{\mathbf{H}}[I_{\mathbf{H}}]$ and an analogous argument gives the same bound for α_{down}. Thus, by Claim 5.2.14 we have $\mathbb{E}_{\mathbf{H}}[I_{\mathbf{H}}] \leq 9\sqrt{d}$ and the fraction of (τ, β)-non-persistent vertices is at most $\frac{72C\tau}{\beta\sqrt{d}}$. Therefore, setting $C_{per} := 72C$ completes the proof. □

5.8.3 Typical Points and Reversibility Proofs

Claim 5.8.3. *For a d-dimensional hypercube $\mathbf{H}$ and $c \geq 1$, we have $|\mathbf{H}_c| \geq (1 - (\varepsilon/d)^c) \cdot 2^d$.*

Proof. Consider a uniform random point $\mathbf{x}$ in the hypercube. The Hamming weight $\|\mathbf{x}\|_1$ is

$\sum_{i=1}^{d} \mathbf{x}_i$, where each $\mathbf{x}_i$ is an iid unbiased Bernoulli. By Hoeffding's theorem, $\mathbb{P}[\Big| \|\mathbf{x}\|_1 - d/2 \Big| \geq t] \leq 2\exp(-2t^2/d)$. We set $t = \sqrt{4cd\log(d/\varepsilon)}$. The probability of not being in the c-middle layers is at most

$$2\exp(-2t^2/d) = 2\exp(-8c\log(d/\varepsilon)) = 2(\varepsilon/d)^{8c} \leq (\varepsilon/d)^c.$$

Hence, the probability of being in the c-middle layers is at least $(1 - (\varepsilon/d)^c)$. $\square$

Lemma 5.8.4 (Most Points are Typical)**.** *For any $\varepsilon \in (0,1)$ and $c \geq 6$,*

$$\mathbb{P}_{\mathbf{x} \in_R [n]^d} \left[\mathbf{x} \textit{ is } c\textit{-typical}\right] \geq 1 - (\varepsilon/d)^{c-5}$$

Proof. Given $\mathbf{x} \in [n]^d$ and a hypercube $\mathbf{H} \ni \mathbf{x}$, let $\chi(\mathbf{x}, \mathbf{H}) = \mathbf{1}(\mathbf{x} \in \mathbf{H} \setminus \mathbf{H}_c)$. By Fact 5.2.11 and Claim 5.2.19, we have

$$\mathbb{E}_{\mathbf{x} \in_R [n]^d} \mathbb{E}_{\mathbf{H} \sim \mathbb{H}(\mathbf{x})} \left[\chi(\mathbf{x}, \mathbf{H})\right] = \mathbb{E}_{\mathbf{H} \sim \mathbb{H}} \mathbb{E}_{\mathbf{x} \in_R \mathbf{H}} \left[\chi(\mathbf{x}, \mathbf{H})\right] \leq (\varepsilon/d)^c$$

Let us set $q_{\mathbf{x}} := \mathbb{E}_{\mathbf{H} \sim \mathbb{H}(\mathbf{x})}[\chi(\mathbf{x}, \mathbf{H})]$, so $\mathbb{E}_{\mathbf{x}}[q_{\mathbf{x}}] \leq (\varepsilon/d)^c$. By Markov's inequality, $\mathbb{P}_{\mathbf{x}}[q_{\mathbf{x}} \geq (\varepsilon/d)^5] \leq (\varepsilon/d)^{c-5}$. Note that when $q_{\mathbf{x}} < (\varepsilon/d)^5$, $\mathbf{x}$ is c-typical. Hence, at least a $(1 - (\varepsilon/d)^{c-5})$-fraction of points are c-typical. $\square$

Claim 5.8.5 (Translations of Typical Points)**.** *Suppose $\mathbf{x} \in [n]^d$ is c-typical. Then for a walk length $\tau \leq \sqrt{d}$, every point $\mathbf{x}' \in supp(\mathcal{U}_\tau(\mathbf{x})) \cup supp(\mathcal{D}_\tau(\mathbf{x}))$ is $(c + \frac{\tau}{\sqrt{d}})$-typical.*

Proof. We prove the claim for $\mathbf{x}' \in \mathrm{supp}(\mathcal{U}_\tau(\mathbf{x}))$. The argument for points in $\mathrm{supp}(\mathcal{D}_\tau(\mathbf{x}))$ is analogous. Let $\mathbf{H}$ be any hypercube containing $\mathbf{x}$ and $\mathbf{x}'$ and let $\|\mathbf{x}\|_{\mathbf{H}}$, $\|\mathbf{x}'\|_{\mathbf{H}}$ denote the Hamming weight of these points in $\mathbf{H}$. Observe that $\|\mathbf{x}'\|_{\mathbf{H}} \leq \|\mathbf{x}\|_{\mathbf{H}} + \tau$ and so if $\mathbf{x} \in \mathbf{H}_c$, then $\|\mathbf{x}'\|_{\mathbf{H}} \leq d/2 + \sqrt{cd\log d} + \tau$ and since $\tau \leq \sqrt{d}$, we have

$$\sqrt{cd\log d} + \tau \leq \sqrt{cd\log d + \tau\sqrt{d}\log d} = \sqrt{\left(c + \frac{\tau}{\sqrt{d}}\right) d\log d}.$$

To see that the first inequality holds, observe that by squaring both sides and rearranging terms, it is equivalent to the inequality

$$\tau^2 + 2\tau\sqrt{cd\log d} \le \tau\sqrt{d}\log d \iff \tau \le \sqrt{d}(\log d - 2\sqrt{c\log d})$$

which clearly holds by our upper bound on τ. Thus, if $\mathbf{x} \in \mathbf{H}_c$, then $\mathbf{x}' \in \mathbf{H}_{c+\frac{\tau}{\sqrt{d}}}$. Therefore, the number of hypercubes $\mathbf{H}$ for which $\mathbf{x}' \in \mathbf{H}_{c+\frac{\tau}{\sqrt{d}}}$ is at least the number of hypercubes $\mathbf{H}$ for which $\mathbf{x} \in \mathbf{H}_c$. Therefore $\mathbf{x}'$ is $(c + \frac{\tau}{\sqrt{d}})$-typical. □

Lemma 5.8.6 (Reversibility Lemma)**.** *For any points $\mathbf{x} \prec \mathbf{x}' \in [n]^d$ and walk length $\ell \le \sqrt{d}/\log^5(d/\varepsilon)$, we have*

$$p_{\mathbf{x},\ell}(\mathbf{x}') = (1 \pm \log^{-3} d)p_{\mathbf{x}',\ell}(\mathbf{x}).$$

Proof. If $t := \|\mathbf{x} - \mathbf{x}'\|_0 > \ell$, then $p_{\mathbf{x},\ell}(\mathbf{x}') = p_{\mathbf{x}',\ell}(\mathbf{x}) = 0$. So assume $t \le \ell$. Fix any $\mathbf{H}$ containing $\mathbf{x}$ and $\mathbf{x}'$ such that $\mathbf{x}, \mathbf{x}' \in \mathbf{H}_{100}$ and let x and x' denote the corresponding hypercube (bit) representations of $\mathbf{x}, \mathbf{x}'$ in $\mathbf{H}$. Let $p_{x,\ell}(x') = \mathbb{P}_{z\sim\mathcal{U}_{\mathbf{H},\ell}(x)}[z = x']$ and $p_{x',\ell}(x) = \mathbb{P}_{z\sim\mathcal{D}_{\mathbf{H},\ell}(x')}[z = x]$. It suffices to show that $p_{x,\ell}(x') = (1 \pm \log^{-3} d)p_{x',\ell}(x)$.

Let S be the set of t coordinates where x and x' differ. Let $Z(x)$ be the set of zero coordinates of the point x; analogously, define $Z(x')$. Recall that the directed upward walk making ℓ steps might not flip ℓ coordinates. The process (recall Definition 5.2.10) picks a uar set R of ℓ coordinates, and only flips the zero bits in x among R. Hence, an ℓ-length walk leads from x to x' iff $R \cap Z(x) = S$.

Let the Hamming weight of x be represented as $d/2 + e_x$, where e_x denotes the "excess". Since x is in the 100-middle layers, $|e_x| \le \sqrt{400d\log(d/\varepsilon)}$.

The sets R that lead from x to y can be constructed by picking any $\ell - t$ coordinates in $\overline{Z(x)}$ and choosing all remaining coordinates to be S. Hence,

$$p_{x,\ell}(x') = \frac{\binom{d/2+e_x}{\ell-t}}{\binom{d}{\ell}}$$

Analogously, consider the downward ℓ step walks from x'. This walk leads to x iff $R \cap \overline{Z(x')} = S$. The sets R that lead from y to x can be constructed by picking any $\ell - t$ coordinates in $Z(x')$ and choosing all remaining coordinates to be S. The size of $Z(x')$ is precisely $|Z(x)| - t = d/2 - e_x - t$. Hence,

$$p_{x',\ell}(x) = \frac{\binom{d/2-e_x-t}{\ell-t}}{\binom{d}{\ell}}$$

Taking the ratio,

$$\begin{aligned}\frac{p_{x,\ell}(x')}{p_{x',\ell}(x)} &= \frac{\binom{d/2+e_x}{\ell-t}}{\binom{d/2-e_x-t}{\ell-t}} = \frac{\prod_{i=0}^{\ell-t-1}(d/2+e_x-i)}{\prod_{i=0}^{\ell-t-1}(d/2-e_x-t-i)} = \prod_{i=0}^{\ell-t-1} \frac{d/2+e_x-i}{d/2-e_x-t-i} \\ &= \prod_{i=0}^{\ell-t-1}\left(1+\frac{2e_x+t}{d/2-e_x-t-i}\right)\end{aligned}$$

Recall that $|e_x| \leq \sqrt{400d\log(d/\varepsilon)}$, $t \leq \ell < \sqrt{d}/\log^5(d/\varepsilon)$. For convenience, let $b := \sqrt{400d\log(d/\varepsilon)}$. So $2e_x + t \leq 3b$. Also, $d/2 - e_x - t - i \geq d/3$ for all $i < \ell$. Applying these bounds,

$$\frac{p_{x,\ell}(x')}{p_{x',\ell}(x)} \leq \prod_{i=0}^{\ell-1}\left(1+\frac{3b}{d/3}\right) \leq \exp\left(\frac{9\ell b}{d}\right) = \exp\left(\frac{\sqrt{d}\cdot\sqrt{400d\log(d/\varepsilon)}}{d\log^5(d/\varepsilon)}\right) \leq 1+\log^{-3} d$$

An analogous calculation proves that $\frac{p_{x,\ell}(x')}{p_{x',\ell}(x)} \geq 1 - \log^{-3} d$. □

Part II

Directed Isoperimetry and Monotonicity Testing of Real-Valued Functions

CHAPTER 6

Directed Isoperimetric Inequalities for Real-Valued Functions

The results in this chapter were originally published in [BKR23]. The main result is a generalization of the robust directed Talagrand inequality for hypergrids (stated in Theorem 4.0.4 and Theorem 2.3.10) to real-valued functions. This generalization is achieved via the following *Boolean decomposition theorem*. Given a DAG $\mathcal{G}$ and a function $f\colon V(\mathcal{G}) \to \mathbb{R}$, we use the notation $\mathcal{S}_f^- = \{(x, y) \in E(\mathcal{G})\colon f(x) > f(y)\}$ to denote the set of all edges in $\mathcal{G}$ where f is decreasing.

Theorem 6.0.1 (Boolean Decomposition)**.** *Suppose $\mathcal{G}$ is a DAG and $f\colon V(\mathcal{G}) \to \mathbb{R}$ is a function over the vertices of $\mathcal{G}$ that is not monotone. Then, for some $k \geq 1$, there exist Boolean functions $f_1, \ldots, f_k\colon V(\mathcal{G}) \to \{0, 1\}$ and vertex-disjoint (induced) subgraphs $\mathcal{H}_1, \ldots, \mathcal{H}_k$ of $\mathcal{G}$ for which the following hold:*

1. $2\sum_{i=1}^{k} \varepsilon(f_i) \geq \varepsilon(f)$.
2. $\mathcal{S}_{f_i}^- \subseteq \mathcal{S}_f^- \cap E(\mathcal{H}_i)$ *for all* $i \in [k]$.

We prove Theorem 6.0.1 in Section 6.2. The main application of Theorem 6.0.1 is that it immediately implies the robust directed Talagrand inequality for hypergrids (Theorem 4.0.4) holds for real-valued functions. To state the generalization we will need to generalize the notion of *colorful thresholded influence* (Definition 4.0.3) to real-valued functions. As in Chapters 4 and 5, we work with the *fully augmented hypergrid* which is the DAG over vertex set $[n]^d$ and edge set

$$E = \bigcup_{i=1}^{d} E_i \quad \text{where} \quad E_i = \{(x,y)\colon x_i < y_i \text{ and } x_j = y_j \ \forall j \neq i\}. \tag{6.1}$$

Given $f\colon [n]^d \to \mathbb{R}$, an *$i$-aligned violation* is an edge $(x,y) \in E_i$ such that $f(x) > f(y)$.

Definition 6.0.2 (Colorful Thresholded Influence)**.** *Fix $f\colon [n]^d \to \mathbb{R}$ and $\chi\colon E \to \{0,1\}$. Fix a dimension $i \in [d]$ and a point $x \in [n]^d$. The colorful thresholded negative influence of x along coordinate i is denoted $\Phi_{f,\chi}(x;i)$, and has value 1 if there exists an i-aligned violation (x,y) such that $\chi(x,y) = \mathbf{1}(f(x) > f(y))$, and has value 0 otherwise. The colorful thresholded negative influence of x is $\Phi_{f,\chi}(x) = \sum_{i=1}^{d} \Phi_{f,\chi}(x;i)$.*

In words, the above definition charges violating edges colored 1 to the lower endpoint and violating edges colored 0 to the upper endpoint. Note that this is consistent with the definition given for Boolean functions (Definition 4.0.3). In Section 6.1 we prove the following inequality using Theorem 6.0.1. This inequality generalizes the robust directed Talagrand inequality proven by [KMS18] with respect to both the domain and the co-domain.

Theorem 6.0.3 (Robust Directed Talagrand Inequality for Real-Valued Functions on the Hypergrid)**.** *Let $f\colon [n]^d \to \mathbb{R}$ be ε-far from monotone, and let $\chi\colon E \to \{0,1\}$ be an arbitrary coloring of the edges of the augmented hypergrid.*

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

6.1 Directed Talagrand Inequality for Real-Valued Functions

In this section, we use our Boolean decomposition Theorem 6.0.1 and our robust directed Talagrand inequality for Boolean functions Theorem 4.0.4 to prove Theorem 6.0.3. We invoke Theorem 6.0.1 with the underlying DAG being the *fully augmented hypergrid*, which has vertex set $[n]^d$ and edge set E defined in eq. (6.1). Let $f\colon [n]^d \to \mathbb{R}$ be a non-monotone function and let $\chi\colon E \to \{0,1\}$ be an arbitrary 2-coloring of E. Given $x \in \{0,1\}^d$ and a subgraph $\mathcal{H}$ of the fully augmented hypergrid, let $\Phi_{f,\chi,\mathcal{H}}(x)$ denote the colorful thresholded influence of x (Definition 6.0.2), but restricted to the edges in $\mathcal{H}$, $E(\mathcal{H})$. Formally, $\Phi_{f,\chi,\mathcal{H}}(x;i) = 1$ iff there exists an i-aligned violation $(x,y) \in E(\mathcal{H})$ such that $\chi(x,y) = \mathbf{1}(f(x) > f(y))$, and $\Phi_{f,\chi,\mathcal{H}}(x) = \sum_{i=1}^{d} \Phi_{f,\chi,\mathcal{H}}(x;i)$.

Let $f_1,\ldots,f_k\colon [n]^d \to \{0,1\}$ be the Boolean functions and $\mathcal{H}_1,\ldots,\mathcal{H}_k$ be the vertex-disjoint subgraphs of the fully augmented hypergrid that are guaranteed by Theorem 6.0.1. Let C' denote the constant from the robust Boolean isoperimetric inequality (Theorem 4.0.4) that is hidden by Ω. We have

$$
\begin{aligned}
\mathbb{E}_{x\sim[n]^d}\left[\sqrt{\Phi_{f,\chi}(x)}\right] &\geq \mathbb{E}_x\left[\sqrt{\Phi_{f,\chi,\bigcup_{i=1}^{k}\mathcal{H}_i}(x)}\right] && (6.2)\\
&= \sum_{i=1}^{k}\mathbb{E}_x\left[\sqrt{\Phi_{f,\chi,\mathcal{H}_i}(x)}\right] && (6.3)\\
&\geq \sum_{i=1}^{k}\mathbb{E}_x\left[\sqrt{\Phi_{f_i,\chi,\mathcal{H}_i}(x)}\right] && (6.4)\\
&= \sum_{i=1}^{k}\mathbb{E}_x\left[\sqrt{\Phi_{f_i,\chi}(x)}\right] && (6.5)\\
&\geq \sum_{i=1}^{k} C'\cdot\varepsilon_{f_i} && (6.6)\\
&\geq \frac{C'\cdot\varepsilon_f}{2}. && (6.7)
\end{aligned}
$$

The inequality eq. (6.2) holds simply because $\bigcup_{i=1}^{k}\mathcal{H}_i$ is a subgraph of the fully augmented hypergrid, while the equality eq. (6.3) holds because the $\mathcal{H}_i$'s are vertex-disjoint. The inequality eq. (6.4) holds since $\mathcal{S}_{f_i}^- \subseteq \mathcal{S}_f^-$ and the equality eq. (6.5) holds since $\mathcal{S}_{f_i}^- \subseteq E(\mathcal{H}_i)$ (these are both by item 2 of Theorem 6.0.1). Finally, eq. (6.6) is due to Theorem 4.0.4 and eq. (6.7) is due to item 1 of Theorem 6.0.1.

6.2 Boolean Decomposition: Proof of Theorem 6.0.1

In this section, we prove the Boolean Decomposition Theorem 6.0.1. Our results consider any partially ordered domain, which we represent by a DAG $\mathcal{G}$. The *transitive closure* of $\mathcal{G}$, denoted $\mathrm{TC}(\mathcal{G})$, is the graph with vertex set $V(\mathcal{G})$ and edge set $\{(x, y) \colon x \prec y\}$. The *violation graph* of f is the graph $(V(\mathcal{G}), E')$, where E' is the set of edges of $\mathrm{TC}(\mathcal{G})$ violated by f.

In Section 6.2.1, we define the key notion of sweeping graphs and identify some of their important properties. In Section 6.2.2, we prove a general lemma that shows how to use a matching M in $\mathrm{TC}(\mathcal{G})$ to find vertex-disjoint sweeping graphs in $\mathcal{G}$ satisfying a "matching rearrangement" property. The techniques in Section 6.2.1 and Section 6.2.2 are inspired by the techniques of [BCS18] used to analyze Boolean functions on the hypergrid domain, $[n]^d$. In Section 6.2.3, we apply our matching decomposition lemma to a carefully chosen matching to obtain the subgraphs $\mathcal{H}_1, \ldots, \mathcal{H}_k$. Finally, in Section 6.2.4, we define the Boolean functions $f_1, \ldots, f_k$ and complete the proof of Theorem 6.0.1.

6.2.1 Sweeping Graphs and Their Properties

Given a graph $\mathcal{G}$ and two subgraphs $\mathcal{H}_1$ and $\mathcal{H}_2$, we define the union $\mathcal{H}_1 \cup \mathcal{H}_2$ to be the graph with vertex set $V(\mathcal{H}_1) \cup V(\mathcal{H}_2)$ and edge set $E(\mathcal{H}_1) \cup E(\mathcal{H}_2)$.

Definition 6.2.1 ((S, T)-Sweeping Graphs)**.** *Given a DAG $\mathcal{G}$ and $s, t \in V(\mathcal{G})$, define $\mathcal{H}(s, t)$ to be the subgraph of $\mathcal{G}$ formed by the union of all directed paths in $\mathcal{G}$ from s to t. Given two disjoint subsets $S, T \subseteq V(\mathcal{G})$, define the (S, T)-*sweeping graph*, denoted $\mathcal{H}(S, T)$, to be the union of directed paths in $\mathcal{G}$ that start from some $s \in S$ and end at some $t \in T$. That is,*

$$\mathcal{H}(S, T) = \bigcup_{(s,t) \in S \times T} \mathcal{H}(s, t).$$

Note that if $s \not\preceq t$ then $\mathcal{H}(s, t) = \emptyset$.

We now prove three properties of sweeping graphs which we use in Section 6.2.4 to analyze our functions $f_1, \ldots, f_k$. Given disjoint sets $S, T \subseteq V(\mathcal{G})$ and $z \in V(\mathcal{H}(S,T))$, define the sets

$$S(z) = \{s \in S \colon s \preceq z\} \text{ and } T(z) = \{t \in T \colon z \preceq t\}.$$

Claim 6.2.2 (Properties of Sweeping Graphs)**.** *Let $\mathcal{G}$ be a DAG and $S, T \subseteq V(\mathcal{G})$ be disjoint sets.*

1. (Property of Nodes in a Sweeping Graph): *If $z \in V(\mathcal{H}(S,T))$ then $S(z) \neq \emptyset$ and $T(z) \neq \emptyset$.*

2. (Property of Nodes Outside of a Sweeping Graph): *If $z \in V(\mathcal{G}) \setminus V(\mathcal{H}(S,T))$ then at most one of the following is true: (a) $\exists y \in V(\mathcal{H}(S,T))$ such that $z \prec y$, (b) $\exists x \in V(\mathcal{H}(S,T))$ such that $x \prec z$.*

3. (Sweeping Graphs are Induced): *If $x, y \in V(\mathcal{H}(S,T))$ and $(x,y) \in E(\mathcal{G})$ then $(x,y) \in E(\mathcal{H}(S,T))$.*

Proof. Property 1 holds by definition of the sweeping graph $\mathcal{H}(S,T)$. If $z \in V(\mathcal{H}(S,T))$, then, by definition of $\mathcal{H}(S,T)$, there exist $s \in S$ and $t \in T$ for which z belongs to some directed path from s to t. That is, $z \in V(\mathcal{H}(s,t))$. Thus $s \in S(z)$ and $t \in T(z)$, and property 1 holds.

We now prove property 2. Suppose, for the sake of contradiction, that there exist $x, y, z \in V(\mathcal{G})$ for which $x, y \in V(\mathcal{H}(S,T))$, $z \notin V(\mathcal{H}(S,T))$, and $x \prec z \prec y$. By property 1, there exist some $s \in S(x)$ and some $t \in T(y)$. Then $s \preceq x \prec z \prec y \preceq t$ and, consequently, z belongs to some directed path from s to t. Thus $z \in V(\mathcal{H}(s,t))$, and so $z \in V(\mathcal{H}(S,T))$. This is a contradiction.

We now prove property 3. Suppose $x, y \in V(\mathcal{H}(S,T))$ and $(x,y) \in E(\mathcal{G})$. By property 1, there exist $s \in S$ and $t \in T$ for which $s \preceq x$ and $y \preceq t$. Since $(x,y) \in E(\mathcal{G})$, we have $x \prec y$

and so $s \preceq x \prec y \preceq t$. Thus, the edge (x, y) belongs to a directed path from s to t. That is, $(x, y) \in E(\mathcal{H}(s, t))$ and so $(x, y) \in E(\mathcal{H}(S, T))$. $\square$

6.2.2 Matching Decomposition Lemma for DAGs

In this section, we prove the following matching decomposition lemma. Recall that $\mathrm{TC}(\mathcal{G})$ denotes the transitive closure of $\mathcal{G}$, which is the graph with vertex set $V(\mathcal{G})$ and edge set $\{(x, y) \colon x \prec y\}$. Consider a matching M in $\mathrm{TC}(\mathcal{G})$. We represent $M \colon S \to T$ as a bijection between two disjoint sets $S, T \subseteq V(\mathcal{G})$ of the same size for which $s \prec M(s)$ for all $s \in S$. For a set $S' \subseteq S$, define $M(S') = \{M(s) \colon s \in S'\}$. Note that for convenience we will sometimes abuse notation and represent M as the set of pairs, $\{(s, M(s)) \colon s \in S\}$, instead of as a bijection.

Lemma 6.2.3 (Matching Decomposition Lemma for DAGs)**.** *For every DAG $\mathcal{G}$ and every matching $M \colon S \to T$ in $\mathrm{TC}(\mathcal{G})$, there exist partitions $(S_i \colon i \in [k])$ of S and $(T_i \colon i \in [k])$ of T, where $M(S_i) = T_i$ for all $i \in [k]$, and the following hold.*

1. (Sweeping Graph Disjointness): *$V(\mathcal{H}(S_i, T_i)) \cap V(\mathcal{H}(S_j, T_j)) = \emptyset$ for all $i \neq j$, where $i, j \in [k]$.*
2. (Matching Rearrangement Property): *For all $i \in [k]$ and $(x, y) \in S_i \times T_i$, if $x \prec y$ then there exists a matching $\widehat{M} \colon S_i \to T_i$ in $\mathrm{TC}(\mathcal{G})$ for which $(x, y) \in \widehat{M}$.*

Proof. In Alg. 2, we show how to construct partitions $(S_i \colon i \in [k])$ for S and $(T_i \colon i \in [k])$ for T from a matching M in $\mathrm{TC}(\mathcal{G})$. Alg. 2 uses the following notion of conflicting pairs.

Definition 6.2.4 (Conflicting Pairs)**.** *Given a DAG $\mathcal{G}$ and four disjoint sets $X, Y, X', Y' \subset V(\mathcal{G})$, we say that the two pairs (X, Y), (X', Y') conflict if $V(\mathcal{H}(X, Y)) \cap V(\mathcal{H}(X', Y')) \neq \emptyset$.*

The following observation is apparent and by design of Alg. 2.

Algorithm 2 Algorithm for constructing conflict-free pairs from a matching M

Input: A DAG $\mathcal{G}$ and a matching $M\colon S \to T$ in $\mathrm{TC}(\mathcal{G})$.

$\mathcal{Q}_0 \leftarrow \{(\{x\},\{y\})\colon (x,y) \in M\}$ $\triangleright$ Initialize pairs using M
for $s \geq 0$ **do**
  if two pairs $(X,Y) \neq (X',Y') \in \mathcal{Q}_s$ *conflict* **then**
    $\mathcal{Q}_{s+1} \leftarrow (\mathcal{Q}_s \setminus \{(X,Y),(X',Y')\}) \cup \{(X \cup X', Y \cup Y')\}$ $\triangleright$ Merge conflicting pairs
  else
    $s^* \leftarrow s$ and **return** $\mathcal{Q}_{s^*}$ $\triangleright$ Terminate when there are no conflicts

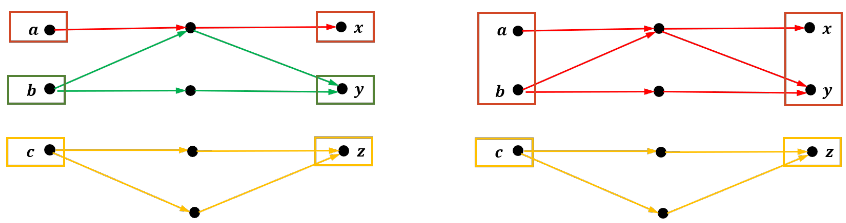


Figure 6.1: *An illustration for Alg. 2 with input matching $M = \{(a,x),(b,y),(c,z)\}$. We initialize $\mathcal{Q}_0 = \{(\{a\},\{x\}),(\{b\},\{y\}),(\{c\},\{z\})\}$. The pairs $(\{a\},\{x\})$ and $(\{b\},\{y\})$ conflict, so we merge them to obtain a new and final collection $Q_1 = \{(\{a,b\},\{x,y\}),(\{c\},\{z\})\}$.*

Observation 6.2.5 (Loop Invariants of Alg. 2)**.** *For all $s \in \{0,1,\ldots,s^*\}$, (a) $M(X) = Y$ for all $(X,Y) \in \mathcal{Q}_s$, (b) $(X\colon (X,\cdot) \in \mathcal{Q}_s)$ is a partition of S, and (c) $(Y\colon (\cdot,Y) \in \mathcal{Q}_s)$ is a partition of T.*

Given a matching $M\colon S \to T$ in $\mathrm{TC}(\mathcal{G})$, we run Alg. 2 to obtain the set $\mathcal{Q}_{s^*}$. See Fig. 6.1 for an illustration. Define $k = |\mathcal{Q}_{s^*}|$ and let $\{(S_i,T_i)\colon i \in [k]\}$ be the set of pairs in $\mathcal{Q}_{s^*}$. By Obs. 6.2.5, $(S_i\colon i \in [k])$ is a partition of S, $(T_i\colon i \in [k])$ is a partition of T, and $M(S_i) = T_i$ for all $i \in [k]$. Item 1 of Lemma 6.2.3 holds since Alg. 2 terminates at step s only when all pairs in $\mathcal{Q}_s$ are non-conflicting (recall Definition 6.2.4). Thus, to prove Lemma 6.2.3 it only remains to prove item 2. To do so, we prove the following Claim 6.2.6, that easily implies item 2. Note that while we only require Claim 6.2.6 to hold for the special case of $s = s^*$, using an inductive argument on s allows us to give a proof for all $s \in \{0,1,\ldots,s^*\}$.

Claim 6.2.6 (Rematching Claim)**.** *For all $s \in \{0,1,\ldots,s^*\}$, pairs $(X,Y) \in \mathcal{Q}_s$, and $(x,y) \in X \times Y$, there exists a matching $\widehat{M}\colon X \setminus \{x\} \to Y \setminus \{y\}$ in $\mathrm{TC}(\mathcal{G})$.*

Proof. The proof is by induction on s. For the base case, if $s = 0$, then, by inspection of Alg. 2, for $(X, Y) \in \mathcal{Q}_0$, we must have $X = \{x\}$ and $Y = \{y\}$. Thus, setting $\widehat{M} = \emptyset$ trivially proves the claim.

Now let $s > 0$. Fix some $(X, Y) \in \mathcal{Q}_s$ and $(x, y) \in X \times Y$. Let $(X_1, Y_1), (X_2, Y_2) \in \mathcal{Q}_{s-1}$ be the pairs of sets in $\mathcal{Q}_{s-1}$ for which $x \in X_1$ and $y \in Y_2$. First, if $(X_1, Y_1) = (X_2, Y_2)$, then by induction there exists a matching $\widehat{M}' \colon X_1 \setminus \{x\} \to Y_1 \setminus \{y\}$ in $\mathrm{TC}(\mathcal{G})$. Note that by definition of Alg. 2, we must have $X_1 \subseteq X$ and $Y_1 \subseteq Y$. Then the required matching is $\widehat{M} = \widehat{M}' \cup M|_{X \setminus X_1}$ where $M|_{(\cdot)}$ denotes the restriction of the original matching M to the set $(\cdot)$. Suppose $(X_1, Y_1) \neq (X_2, Y_2)$. This is the interesting case, and we give an accompanying illustration in Fig. 6.2. By definition of Alg. 2, it must be that (X_1, Y_1) and (X_2, Y_2) *conflict* (recall Definition 6.2.4) and were merged to form $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$. Thus, there exists some vertex $z \in V(\mathcal{H}(X_1, Y_1)) \cap V(\mathcal{H}(X_2, Y_2))$ and $x_1 \in X_1, y_1 \in Y_1, x_2 \in X_2, y_2 \in Y_2$ for which $x_1 \preceq z \preceq y_1$ and $x_2 \preceq z \preceq y_2$.

We now invoke the inductive hypothesis to get matchings $\widehat{M_1} \colon X_1 \setminus \{x\} \to Y_1 \setminus \{y_1\}$ and $\widehat{M_2} \colon X_2 \setminus \{x_2\} \to Y_2 \setminus \{y\}$ in $\mathrm{TC}(\mathcal{G})$. Observe that $x_2 \preceq z \preceq y_1$ and thus we can match x_2 and y_1. The required matching in $\mathrm{TC}(\mathcal{G})$ is $\widehat{M} = \widehat{M_1} \cup \widehat{M_2} \cup \{(x_2, y_1)\}$. $\square$

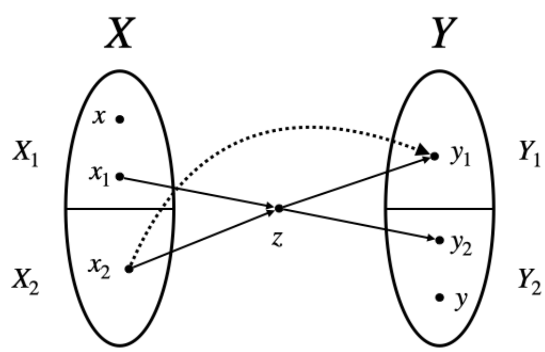


Figure 6.2: *An illustration for the case of $(X_1, Y_1) \neq (X_2, Y_2)$ in the proof of Claim 6.2.6. The solid lines represent directed paths. The dotted line represents the pair (x_2, y_1) added to obtain the final matching $\widehat{M}$. The only vertices of $X \cup Y$ not participating in $\widehat{M}$ are x and y.*

We conclude the proof of Lemma 6.2.3 by showing that Claim 6.2.6 implies item 2. We are given $(S_i, T_i) \in \mathcal{Q}_{s^*}$ for some $i \in [k]$ and $(x, y) \in S_i \times T_i$ where $x \prec y$. By Claim 6.2.6 there exists a matching $\widehat{M}'\colon S_i \setminus \{x\} \to T_i \setminus \{y\}$ in $\mathrm{TC}(\mathcal{G})$. We then set $\widehat{M} = \widehat{M}' \cup \{(x, y)\}$. Since $x \prec y$, the final matching $\widehat{M}\colon S_i \to T_i$ is a matching in $\mathrm{TC}(\mathcal{G})$ which contains the pair (x, y). $\square$

6.2.3 Specifying a Matching to Construct the Subgraphs $\mathcal{H}_1, \ldots, \mathcal{H}_k$

In this section, we apply Lemma 6.2.3 to a carefully chosen matching M in order to construct our vertex-disjoint subgraphs $\mathcal{H}_1, \ldots, \mathcal{H}_k$.

Definition 6.2.7 (Max-weight, Min-cardinality Matching)**.** *A matching M in $\mathrm{TC}(\mathcal{G})$ is a max-weight, min-cardinality matching for f if M maximizes $\sum_{(x,y)\in M}(f(x) - f(y))$ and among such matchings minimizes $|M|$.*

Henceforth, let M denote a max-weight, min-cardinality matching. Let S and T denote the set of lower and upper endpoints, respectively, of M. We use the following well-known fact on matchings in the violation graph.

Fact 6.2.8 (Corollary 2 of [FLN+02])**.** *For a DAG $\mathcal{G}$ and function $f\colon V(\mathcal{G}) \to \mathbb{R}$, the distance to monotonicity $\varepsilon(f)$ is equal to the size of the minimum vertex cover of the violation graph of f divided by $|V(\mathcal{G})|$.*

Fact 6.2.9. *M is a matching in the violation graph of f that is also maximal. That is, (a) $f(x) > f(y)$ for all $(x, y) \in M$ and (b) $|M| \geq (\varepsilon(f) \cdot |V(\mathcal{G})|)/2$.*

Proof. First, for the sake of contradiction, suppose $f(x) \leq f(y)$ for some pair $(x, y) \in M$. Then we can set $M = M \setminus \{(x, y)\}$, which can only increase $\sum_{(x,y)\in M}(f(x) - f(y))$ and will decrease $|M|$ by 1. This contradicts the definition of M. Thus, $f(x) > f(y)$ for all $(x, y) \in M$ and so M is a matching in the *violation graph of f*. Second, since M maximizes $\sum_{(x,y)\in M}(f(x) - f(y))$, it must also be a *maximal* matching in the violation graph of f. Thus, (b) follows from Fact 6.2.8 and the fact that the size of any maximal matching is at least half the size of the minimum vertex cover. $\square$

We now apply Lemma 6.2.3 to M, obtaining the partitions $(S_i : i \in [k])$ and $(T_i : i \in [k])$ for S and T, respectively, for which $M(S_i) = T_i$ for all $i \in [k]$. For each $i \in [k]$, let $\mathcal{H}_i = \mathcal{H}(S_i, T_i)$. We use the collection of sweeping graphs $\mathcal{H}_1, \ldots, \mathcal{H}_k$ to prove Theorem 6.0.1. Note that these subgraphs are all vertex-disjoint by item 1 of Lemma 6.2.3. We use item 2 of Lemma 6.2.3 to prove the following lemma regarding the (S_i, T_i) pairs. The proof crucially relies on the fact that M is a max-weight, min-cardinality matching.

Lemma 6.2.10 (Property of the Pairs (S_i, T_i))**.** *For all $i \in [k]$ and $(x, y) \in S_i \times T_i$, if $x \prec y$ then $f(x) > f(y)$.*

Proof. Suppose there exists $i \in [k]$, $x \in S_i$, and $y \in T_i$ for which $x \prec y$ and $f(x) \leq f(y)$. By item 2 of Lemma 6.2.3 there exists a matching $\widehat{M} : S \to T$ in $\mathrm{TC}(\mathcal{G})$ for which $(x, y) \in \widehat{M}$. In particular, since M and $\widehat{M}$ have identical sets of lower and upper endpoints,

$$\sum_{(s,t) \in \widehat{M}} (f(s) - f(t)) = \sum_{(s,t) \in M} (f(s) - f(t)) \text{ and } |\widehat{M}| = |M|.$$

Now set $\widehat{M}' = \widehat{M} \setminus \{(x, y)\}$ and observe that since $f(x) \leq f(y)$,

$$\sum_{(s,t) \in \widehat{M}'} (f(s) - f(t)) \geq \sum_{(s,t) \in M} (f(s) - f(t)) \text{ and } |\widehat{M}'| < |M|.$$

Therefore, M is not a max-weight, min-cardinality matching and this is a contradiction. $\square$

6.2.4 Tying it Together: Defining the Boolean Functions $f_1, \ldots, f_k$

We are now equipped to define the functions $f_1, \ldots, f_k : V(\mathcal{G}) \to \{0, 1\}$ and complete the proof of Theorem 6.0.1. First, given $i \in [k]$ and $z \in V(\mathcal{G}) \setminus V(\mathcal{H}_i)$, we say that z is *below* $\mathcal{H}_i$ if there exists $y \in V(\mathcal{H}_i)$ for which $z \prec y$, and z is *above* $\mathcal{H}_i$ if there exists $x \in V(\mathcal{H}_i)$ for which $x \prec z$. Since $\mathcal{H}_i$ is the (S_i, T_i)-sweeping graph, by item 2 of Claim 6.2.2, vertex z cannot be both below and above $\mathcal{H}_i$, simultaneously. Second, given $z \in V(\mathcal{H}_i)$, we define the set $T_i(z) = \{t \in T_i : z \preceq t\}$. Note that by item 1 of Claim 6.2.2, $T_i(z) \neq \emptyset$ for all $z \in V(\mathcal{H}_i)$, and so the quantity $\max_{t \in T_i(z)} f(t)$ is always well-defined.

Definition 6.2.11. *For each $i \in [k]$, define the function $f_i \colon V(\mathcal{G}) \to \{0,1\}$ as follows. For every $z \in V(\mathcal{G})$,*

$$f_i(z) = \begin{cases} 1, & \text{if } z \in V(\mathcal{H}_i) \text{ and } f(z) > \max_{t \in T_i(z)} f(t), \\ 0, & \text{if } z \in V(\mathcal{H}_i) \text{ and } f(z) \leq \max_{t \in T_i(z)} f(t), \\ 1, & \text{if } z \notin V(\mathcal{H}_i) \text{ and } z \text{ is above } \mathcal{H}_i, \\ 0, & \text{if } z \notin V(\mathcal{H}_i) \text{ and } z \text{ is not above } \mathcal{H}_i. \end{cases}$$

See Fig. 6.3 for an illustration of the values of f_i. We first prove item 1 of Theorem 6.0.1. Recall that $M(S_i) = T_i$ for all $i \in [k]$. Let $M_i = M|_{S_i}$ denote the matching M restricted to S_i. Consider $x \in S_i$. By Lemma 6.2.10, $f(x) > f(y)$ for all $y \in T_i$ such that $x \prec y$. Thus $f(x) > \max_{t \in T_i(x)} f(t)$ and so $f_i(x) = 1$. Now consider $y \in T_i$. Observe that $y \in T_i(y)$. Thus, clearly, $f(y) \leq \max_{t \in T_i(y)} f(t)$, and so $f_i(y) = 0$. Therefore, $f_i(x) = 1$ for all $x \in S_i$ and $f_i(y) = 0$ for all $y \in T_i$. In particular, $f_i(x) = 1 > 0 = f_i(M(x))$ for all $x \in S_i$ and so M_i is a matching in the violation graph of f_i. Thus, $\varepsilon(f_i) \geq \frac{|M_i|}{|V(\mathcal{G})|}$ for all $i \in [k]$. It follows,

$$\sum_{i=1}^{k} \varepsilon(f_i) \geq |V(\mathcal{G})|^{-1} \sum_{i=1}^{k} |M_i| = |V(\mathcal{G})|^{-1} \cdot |M| \geq |V(\mathcal{G})|^{-1} \cdot \frac{\varepsilon(f) \cdot |V(\mathcal{G})|}{2} = \frac{\varepsilon(f)}{2}$$

by the above argument and Fact 6.2.9. Thus, item 1 of Theorem 6.0.1 holds.

To prove item 2 of Theorem 6.0.1, we need to show that for all $i \in [k]$ the following hold:

$$\mathcal{S}^-_{f_i} \subseteq E(\mathcal{H}_i) \text{ and } \mathcal{S}^-_{f_i} \subseteq \mathcal{S}^-_f.$$

We first prove that $\mathcal{S}^-_{f_i} \subseteq E(\mathcal{H}_i)$. Consider an edge $(x,y) \in E(\mathcal{G}) \setminus E(\mathcal{H}_i)$. We need to show that $f_i(x) \leq f_i(y)$. First, observe that if both $x, y \in V(\mathcal{H}_i)$, then by item 3 of Claim 6.2.2, we have $(x,y) \in E(\mathcal{H}_i)$. Thus, we only need to consider the following three cases. Recall

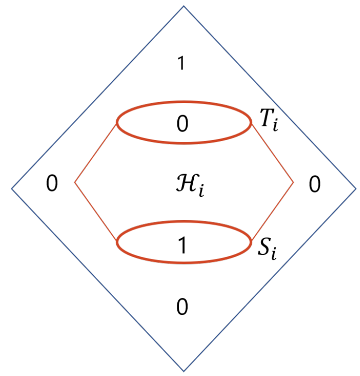


Figure 6.3: *An illustration for the Boolean function f_i of Definition 6.2.11. The diamond represents the DAG $\mathcal{G}$ whose paths are directed from bottom to top. The hexagon represents the sweeping graph $\mathcal{H}_i = \mathcal{H}(S_i, T_i)$. The value of f_i is 1 for the vertices in S_i and 0 for the vertices in T_i. For vertices outside of $\mathcal{H}_i$, its value is 1 for those vertices which are above $\mathcal{H}_i$ and 0 for vertices which are not above $\mathcal{H}_i$.*

that $f_i(x), f_i(y) \in \{0, 1\}$.

1. $x \in V(\mathcal{H}_i)$, $y \notin V(\mathcal{H}_i)$: In this case, y is above $\mathcal{H}_i$, and so $f_i(y) = 1$. Thus, $f_i(x) \leq f_i(y)$.

2. $x \notin V(\mathcal{H}_i)$, $y \in V(\mathcal{H}_i)$: In this case, x is below $\mathcal{H}_i$, and so x is *not* above $\mathcal{H}_i$ by item 2 of Claim 6.2.2. Thus, $f_i(x) = 0$, and so $f_i(x) \leq f_i(y)$.

3. $x \notin V(\mathcal{H}_i)$, $y \notin V(\mathcal{H}_i)$: If x is above $\mathcal{H}_i$, then y is above $\mathcal{H}_i$ as well, and so $f_i(x) = f_i(y) = 1$. Otherwise, x is *not* above $\mathcal{H}_i$ and so $f_i(x) = 0$. Thus, $f_i(x) \leq f_i(y)$.

Therefore, $S^-_{f_i} \subseteq E(\mathcal{H}_i)$.

We now prove that $\mathcal{S}^-_{f_i} \subseteq \mathcal{S}^-_f$. Consider an edge $(x, y) \in \mathcal{S}^-_{f_i}$. Then $f_i(x) = 1$ and $f_i(y) = 0$. Since $S^-_{f_i} \subseteq E(\mathcal{H}_i)$, we have $(x, y) \in E(\mathcal{H}_i)$ and so $x, y \in V(\mathcal{H}_i)$. By definition of the functions f_i, it holds that $f(x) > \max_{t \in T_i(x)} f(t)$ and $f(y) \leq \max_{t \in T_i(y)} f(t)$. Since $x \prec y$, then $T_i(y) \subseteq T_i(x)$, because all vertices reachable from y are also reachable from x. Therefore,

$$f(x) > \max_{t \in T_i(x)} f(t) \geq \max_{t \in T_i(y)} f(t) \geq f(y).$$

Thus $f(x) > f(y)$, and so $(x, y) \in \mathcal{S}_f^-$. As a result, $S_{f_i}^- \subseteq \mathcal{S}_f^-$ and item 2 of Theorem 6.0.1 holds. This concludes the proof of Theorem 6.0.1.

CHAPTER 7

Monotonicity Testing of Functions with Bounded Image Size

In this chapter we prove nearly matching upper and lower bounds for non-adaptive, one-sided error monotonicity testing of functions over the hypercube with image size at most r. These results were originally published in [BKR23]. Our upper bound relies on the robust directed Talagrand inequality Theorem 6.0.3 for the special case of $n = 2$, i.e. the hypercube.

Theorem 7.0.1. *There exists a non-adaptive, 1-sided error ε-tester for monotonicity of $f\colon \{0,1\}^d \to \mathbb{R}$ that makes $\widetilde{O}\Big(\min\big(\frac{r\sqrt{d}}{\varepsilon^2}, \frac{d}{\varepsilon}\big)\Big)$ queries and works for all functions f with image size r.*

Theorem 7.0.2. *There exists a constant $\varepsilon > 0$, such that for all $d, r \in \mathbb{N}$, every non-adaptive, 1-sided error ε-tester for monotonicity of functions $f\colon \{0,1\}^d \to [r]$ requires $\Omega(\min(r\sqrt{d}, d))$ queries.*

We prove Theorems 7.0.1 and 7.0.2 in Sections 7.1 and 7.2, respectively.

7.1 An $\widetilde{O}(r\sqrt{d})$ Monotonicity Tester over the Hypercube

In this section we prove Theorem 7.0.1. We show that the tester of [KMS18] for Boolean functions can be employed to test monotonicity of real-valued functions with bounded image size. The tester is simple: it queries two comparable vertices x and y and rejects if the pair exhibits a violation to monotonicity for f. The tester tries different values τ for the distance between x and y, that is, the number of coordinates on which they differ. The key step in the analysis of [KMS18] (and in our analysis) is to show that for some choice of τ, the tester will detect a violation to monotonicity with high enough probability. The extra factor of r in the query complexity of our tester arises because we are forced to choose τ which is a factor of $(r-1)$ smaller than for the Boolean case. Intuitively, the reason for this is that as the walk length τ increases, the probability that the function value stays below a certain threshold decreases. We make this precise in Section 7.1.2.

We first define the distribution from which the tester samples x and y. Following this, we present the tester as Alg. 3. Let p denote the largest integer for which $2^p \le \sqrt{d/\log d}$. In Alg. 3, we sample pairs of vertices at distance τ, where τ ranges over the powers of two up to 2^p.

Definition 7.1.1 (Pair Test Distribution). *Given parameters $b \in \{0,1\}$ and a positive integer τ, define the following distribution $\mathcal{D}_{\mathit{pair}}(b,\tau)$ over pairs $(x,y) \in (\{0,1\}^d)^2$. Sample $\mathbf{x}$ uniformly from $\{0,1\}^d$. Let $\boldsymbol{S} = \{i \in [d] \colon \mathbf{x}_i = b\}$. If $\tau > |\boldsymbol{S}|$, then set $\mathbf{y} = \mathbf{x}$. Otherwise, sample a uniformly random set $\boldsymbol{T} \subseteq \boldsymbol{S}$ of size $|\boldsymbol{T}| = \tau$. Obtain $\mathbf{y}$ by setting $\mathbf{y}_i = 1 - \mathbf{x}_i$ if $i \in \boldsymbol{T}$ and $\mathbf{y}_i = \mathbf{x}_i$ otherwise.*

Algorithm 3 Monotonicity Tester for $f \colon \{0,1\}^d \to \mathbb{R}$

Input: Parameters $\varepsilon \in (0,1)$, dimension d, and image size r; oracle access to function $f \colon \{0,1\}^d \to \mathbb{R}$.

1: **for all** $b \in \{0,1\}$ and $\tau \in \{1,2,4,\dots,2^p\}$ **do**
2: **repeat** $\widetilde{O}\left(\min\left(\frac{r\sqrt{d}}{\varepsilon^2}, \frac{d}{\varepsilon}\right)\right)$ times:
3: Sample $(\mathbf{x},\mathbf{y}) \sim \mathcal{D}_{\mathtt{pair}}(b,\tau)$.
4: **if** $b = 0$ and $f(\mathbf{x}) > f(\mathbf{y})$ **then reject**. ▷ if $b = 0$ then $\mathbf{x} \preceq \mathbf{y}$
5: **if** $b = 1$ and $f(\mathbf{x}) < f(\mathbf{y})$ **then reject**. ▷ if $b = 1$ then $\mathbf{x} \succeq \mathbf{y}$
6: **accept**.

Our tester only uses comparisons between function values, not the values themselves.

Thus, for the purposes of our analysis we can consider functions with the range $[r]$ w.l.o.g.

When $\tau = 1$, the algorithm is simply sampling edges from the d-dimensional hypercube. The distribution from which we sample is not the uniform distribution on edges, but following an argument from [KMS18], we can assume that for $\tau = 1$, our tester has the same guarantees as the edge tester.

The choice of the distance parameter τ for which the rejection probability of the tester is high depends on the existence of a certain "good" bipartite subgraph of violated edges. Our analysis differs from the analysis of [KMS18] both in how we obtain the "good" subgraph of violated edges and in the choice of the optimal distance parameter τ.

We extend the following definitions from [KMS18]. Let $G(A, B, E_{AB})$ denote a directed bipartite graph with vertex sets A and B and all edges in E_{AB} *directed* from A to B.

Definition 7.1.2 ((K,Δ)-Good Graphs)**.** *A directed bipartite graph $G(A, B, E_{AB})$ is (K,Δ)-good if for X, Y such that either $X = A$, $Y = B$ or $X = B$, $Y = A$, we have: (a) $|X| = K$. (b) Vertices in X have degree exactly Δ. (c) Vertices in Y have degree at most 2Δ. The graph G is (K,Δ)-left-good if $X = A$ and (K,Δ)-right-good if $X = B$.*

The *weight* of $x \in \{0,1\}^d$, denoted by $|x|$, is the number of coordinates of x with value 1.

Definition 7.1.3 (Persistence)**.** *Given a function $f\colon \{0,1\}^d \to [r]$ and an integer $\tau \in \left[1, \sqrt{\frac{d}{\log d}}\right]$, a vertex $x \in \{0,1\}^d$ of weight in the range $\frac{d}{2} \pm O(\sqrt{d \log d})$ is τ-*right-persistent *for f if*

$$\mathbb{P}_{\mathbf{y}}[f(\mathbf{y}) \le f(x)] > \frac{9}{10},$$

where $\mathbf{y}$ is obtained by choosing a uniformly random set $\boldsymbol{T} \subset \{i \in [d] : x_i = 0\}$ of size τ and setting $\mathbf{y}_i = 1$ if $i \in \boldsymbol{T}$ and $\mathbf{y}_i = x_i$ otherwise[1]. We define τ-left-persistence symmetrically.

We use the following technical claim implicitly shown in the analysis of the tester of [KMS18].

[1]Note that $\tau \ge |\{i \in [d] : x_i = 0\}|$ by our assumption on x and τ.

Claim 7.1.4 ([KMS18]). *Suppose there exists a (K,Δ)-right-good subgraph $G(A,B,E_{AB})$ of the directed d-dimensional hypercube, such that (a) $E_{AB} \subseteq \mathcal{S}_f^-$, (b) $K\sqrt{\Delta} = \Theta(\frac{\varepsilon(f)\cdot 2^d}{\log d})$, and (c) at least $\frac{99}{100}|B|$ of the vertices in B are $(\tau'-1)$-right-persistent for some τ' such that $\tau' \cdot \Delta \ll d$. Then there exists a constant $C' > 0$, such that for $(\mathbf{x},\mathbf{y}) \sim \mathcal{D}_{pair}(0,\tau')$,*

$$\mathbb{P}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x}) > f(\mathbf{y})] \geq \frac{C' \cdot \tau'}{d} \cdot \frac{K}{2^d} \cdot \Delta.$$

The analogous claim holds given a (K,Δ)-left-good subgraph with many $(\tau'-1)$-left-persistent vertices in A and $(\mathbf{x},\mathbf{y})$ drawn from $\mathcal{D}_{\texttt{pair}}(1,\tau')$.

In Section 7.1.1, we prove Lemma 7.1.6 which obtains a good subgraph for f satisfying conditions (a) and (b) of Claim 7.1.4. In Section 7.1.2, we prove Lemma 7.1.8 which gives an upper bound on the fraction of non-persistent vertices, enabling us to satisfy condition (c). Finally, in Section 7.1.3, we use Lemma 7.1.6 and Lemma 7.1.8 to show that the conditions of Claim 7.1.4 are satisfied. Finally, we use it to prove Theorem 7.0.1.

7.1.1 Existence of a Good Bipartite Subgraph

In this section, we prove Lemma 7.1.6 on the existence of good bipartite subgraphs for real-valued functions, which was proved in [KMS18] for the special case of Boolean functions. This lemma crucially relies on our isoperimetric inequality for real-valued functions (Theorem 6.0.3 for the special case of $n = 2$). We first state (without proof) a combinatorial result of [KMS18], which we need for our lemma.

Lemma 7.1.5 (Lemma 6.5 of [KMS18]). *Let $G(A,B,E_{AB})$ be a directed bipartite graph whose vertices have degree at most 2^s. Suppose in addition, that for any 2-coloring of its edges $\texttt{col} : E_{AB} \to \{\text{red}, \text{blue}\}$ we have*

$$\sum_{x\in A} \sqrt{\deg_{\text{red}}(x)} + \sum_{y\in B} \sqrt{\deg_{\text{blue}}(y)} \geq L, \tag{7.1}$$

where $\deg_{\text{red}}(x)$ *denotes the number of red edges incident on* x *and* $\deg_{\text{blue}}(y)$ *denotes the number of blue edges incident on* y*. Then* $G(A, B, E_{AB})$ *contains a subgraph that is* (K, Δ)*-good with* $K\sqrt{\Delta} \geq \frac{L}{8s}$*.*

We can now generalize Lemma 7.1 of [KMS18].

Lemma 7.1.6. *For all functions* $f\colon \{0,1\}^d \to \mathbb{R}$*, there exists a subgraph* $G(A, B, E_{AB})$ *of the directed, d-dimensional hypercube which is* (K, Δ)*-good, where* $K\sqrt{\Delta} = \Theta(\frac{\varepsilon(f)\cdot 2^d}{\log d})$ *and* $E_{AB} \subseteq \mathcal{S}_f^-$*.*

Proof. Our proof relies on Lemma 7.1.5. Condition eq. (7.1) is clearly reminiscent of the isoperimetric inequality in Theorem 6.0.3. We want to partition the vertices in $\{0,1\}^d$ into sets A and B such that all the violated edges are directed from A to B and apply Theorem 6.0.3 to the resulting graph. In addition, we want eq. (7.1) to hold for a big enough value of L. In the Boolean case, we can simply partition the vertices by function values. In contrast, for real-valued functions, a vertex $x \in \{0,1\}^d$ can be incident on both incoming and outgoing violated edges. To overcome this challenge we resort to the bipartiteness of the directed hypercube, where each edge is between a vertex with an odd weight and a vertex with an even weight. Partition $\mathcal{S}_f^-$ into two sets:

$$E_0 = \{(x,y) \in \mathcal{S}_f^- : |x| \text{ is even}\};$$
$$E_1 = \{(x,y) \in \mathcal{S}_f^- : |x| \text{ is odd}\}.$$

For $j \in \{0,1\}$, let V_j and W_j denote the set of lower and upper endpoints, respectively, of the edges in E_j. We consider the two subgraphs $G_j(V_j, W_j, E_j)$ for $j \in \{0,1\}$. Notice that the vertices in $V_0 \cup W_1$ have even weight and the vertices in $V_1 \cup W_0$ have odd weight. Obviously, V_0 and W_1 may not be disjoint, and similarly V_1 and W_0 may not be disjoint, and thus G_0 and G_1 may not be vertex-disjoint.

We quickly explain why we cannot simply use Lemma 7.1.5 with either G_0 or G_1. Fix a 2-coloring of the edges $E_0 \cup E_1$. By averaging, one of the graphs will have a high enough contribution to left-hand side of the isoperimetric inequality of Theorem 6.0.3. Assume this graph is G_0. As a result, condition eq. (7.1) will hold for G_0 with $L = \Omega(\varepsilon \cdot 2^d)$. However, one cannot guarantee that condition eq. (7.1) holds for *all* possible colorings of the edges of G_0. Our construction below describes how to combine G_0 and G_1 so that we can jointly "feed" them into Lemma 7.1.5.

We construct copies $\widehat{G}_0$ and $\widehat{G}_1$ of G_0 and G_1, so that $\widehat{G}_0$ contains a vertex labelled $(x, 0)$ for each vertex x of G_0, and $\widehat{G}_1$ contains a vertex $(x, 1)$ for each vertex x of G_1. For each edge (x, y) in G_0 we add an edge from $(x, 0)$ to $(y, 0)$ in $\widehat{G}_0$. We do the same for the edges of G_1. Note that each edge of $\mathcal{S}_f^-$ has exactly one copy, either in $\widehat{G}_0$ or $\widehat{G}_1$.

Let $\widehat{G}(\widehat{V}, \widehat{W}, \mathcal{S}_f^-)$ denote the union of the two vertex-disjoint graphs $\widehat{G}_0$ and $\widehat{G}_1$. That is,

$$\begin{aligned}
\widehat{V} &= \{(x, 0) \mid x \in V_0\} \cup \{(x, 1) \mid x \in V_1\}, \\
\widehat{W} &= \{(y, 0) \mid y \in W_0\} \cup \{(y, 1) \mid y \in W_1\}.
\end{aligned}$$

All the edges of $\widehat{G}$ are directed from $\widehat{V}$ to $\widehat{W}$. Although imprecise, we think of the edges of $\widehat{G}$ as $\mathcal{S}_f^-$, since each edge in $\mathcal{S}_f^-$ has exactly one copy in $\widehat{G}$.

Consider a 2-coloring $\texttt{col}\colon \mathcal{S}_f^- \to \{\text{red}, \text{blue}\}$. Observe that

$$\begin{aligned}
\sum_{(x,\cdot)\in\widehat{V}} \sqrt{I_{f,\text{red}}^-(x)} + \sum_{(y,\cdot)\in\widehat{W}} \sqrt{I_{f,\text{blue}}^-(x)} &= \sum_{x\in V_0\cup V_1} \sqrt{I_{f,\text{red}}^-(x)} + \sum_{y\in W_0\cup W_1} \sqrt{I_{f,\text{blue}}^-(y)} \\
&= \sum_{\substack{x\in\{0,1\}^d \\ |x| \text{ is even}}} \sqrt{I_{f,\text{red}}^-(x)} + \sqrt{I_{f,\text{blue}}^-(x)} + \sum_{\substack{x\in\{0,1\}^d \\ |x| \text{ is odd}}} \sqrt{I_{f,\text{red}}^-(x)} + \sqrt{I_{f,\text{blue}}^-(x)} \\
&= \sum_{x\in\{0,1\}^d} \sqrt{I_{f,\text{red}}^-(x)} + \sum_{y\in\{0,1\}^d} \sqrt{I_{f,\text{blue}}^-(y)} \geq C \cdot \varepsilon(f) \cdot 2^d,
\end{aligned}$$

where the inequality holds by Theorem 6.0.3.

By construction, $I_{f,\text{red}}^-(x) = \deg_{\text{red}}((x, \cdot))$ for all $(x, \cdot) \in \widehat{V}$ and $I_{f,\text{blue}}^-(y) = \deg_{\text{blue}}((y, \cdot))$ for all $(y, \cdot) \in \widehat{W}$. We have that condition eq. (7.1) of Lemma 7.1.5 holds with $L = C \cdot \varepsilon(f) \cdot 2^d$.

Thus, $\widehat{G}$ contains a subgraph $G_{\text{good}}(A, B, E_{AB})$ that is (K, Δ)-good with $K\sqrt{\Delta} \geq \frac{L}{8 \log d}$. Without loss of generality, assume $G_{\text{good}}(A, B, E_{AB})$ is (K, Δ)-right-good.

Let $G_{\text{good},0} = (A_0, B_0, E_{A_0 B_0})$ denote the subgraph of G_{good} lying in $\widehat{G}_0$ and let $G_{\text{good},1} = (A_1, B_1, E_{A_1 B_1})$ denote the subgraph of G_{good} lying in $\widehat{G}_1$. Since $B_0 \cap B_1 = \emptyset$, we know that either $|B_0| \geq K/2$ or $|B_1| \geq K/2$. Suppose $|B_0| \geq K/2$. Moreover, since $\widehat{G}_0$ and $\widehat{G}_1$ are vertex-disjoint subgraphs, the degree of a vertex of $A_0 \cup B_0$ in $G_{\text{good},0}$ is the same its degree in G_{good}. Thus, $G_{\text{good},0}$ is a $(K/2, \Delta)$-right-good subgraph of the d-dimensional directed hypercube for which $\frac{K}{2}\sqrt{\Delta} \geq \frac{L}{16 \log d}$.

By removing some vertices from B_0, and redefining K if necessary, we may assume that $K\sqrt{\Delta} = \Theta\left(\frac{\varepsilon(f) \cdot 2^d}{\log d}\right)$. This completes the proof of Lemma 7.1.6. $\square$

7.1.2 Bounding the Number of Non-Persistent Vertices

We prove Lemma 7.1.8 that bounds the number of non-persistent vertices for a function f and a given distance parameter τ. All results in this section also hold for τ-left-persistence.

For a function $f \colon \{0,1\}^d \to \mathbb{R}$, we define I_f^- as $\frac{|S_f^-|}{2^d}$.

Corollary 7.1.7 (Corollary of Theorem 6.6, Lemma 6.8 of [KMS18])**.** *Consider a function $h \colon \{0,1\}^d \to \{0,1\}$ and an integer $\tau \in \left[1, \sqrt{\frac{d}{\log d}}\right]$. If $I_h^- \leq \sqrt{d}$ then*

$$\mathbb{P}_{\mathbf{x} \sim \{0,1\}^d}\left[\mathbf{x} \textit{ is not } \tau\textit{-right-persistent for } h\right] = O\left(\frac{\tau}{\sqrt{d}}\right). \tag{7.2}$$

We generalize the above result to functions with image size $r \geq 2$.

Lemma 7.1.8. *Consider a function $f \colon \{0,1\}^d \to [r]$ and an integer $\tau \in \left[1, \sqrt{\frac{d}{\log d}}\right]$. If $I_f^- \leq \sqrt{d}$, then*

$$\mathbb{P}_{\mathbf{x} \sim \{0,1\}^d}\left[\mathbf{x} \textit{ is not } \tau\textit{-right-persistent for } f\right] = (r-1) \cdot O\left(\frac{\tau}{\sqrt{d}}\right).$$

Proof. For all $t \in [r]$, define the threshold function $h_t \colon \{0,1\}^d \to \{0,1\}$ as:

$$h_t(x) = \begin{cases} 1 & \text{if } f(x) > t, \\ 0 & \text{otherwise.} \end{cases}$$

Observe that for all $t \in [r]$, we have $\mathcal{S}^-_{h_t} \subseteq \mathcal{S}^-_f$, and thus $I^-_{h_t} \leq I^-_f \leq \sqrt{d}$. By Corollary 7.1.7, we have that eq. (7.2) holds for $h = h_t$ for all $t \in [r]$. Next, we point out that a vertex $x \in \{0,1\}^d$ is τ-right-persistent for f if and only if x is τ-right-persistent for the Boolean function $h_{f(x)}$. Too see this, consider a vertex z such that $x \prec z$. First, note that $h_{f(x)}(x) = 0$. Second, note that $h_{f(x)}(z) = 1$ if and only if $f(z) > f(x)$ by definition of $h_{f(x)}$. Therefore, $f(z) \leq f(x)$ if and only if $h_{f(x)}(z) \leq h_{f(x)}(x)$. Finally, note that all vertices are persistent for h_r since $h_r(x) = 0$ for all $x \in \{0,1\}^d$. Using these observations, we have

$$\begin{aligned}
\mathbb{P}_{\mathbf{x}\sim\{0,1\}^d}\left[\mathbf{x} \text{ is not } \tau\text{-right-persistent for } f\right] &= \mathbb{P}_{\mathbf{x}\sim\{0,1\}^d}\left[\mathbf{x} \text{ is not } \tau\text{-right-persistent for } h_{f(\mathbf{x})}\right] \\
&\leq \mathbb{P}_{\mathbf{x}\sim\{0,1\}^d}\left[\exists t \in [r-1] \colon \mathbf{x} \text{ is not } \tau\text{-right-persistent for } h_t\right] \\
&\leq \sum_{t=1}^{r-1} \mathbb{P}_{\mathbf{x}\sim\{0,1\}^d}\left[\mathbf{x} \text{ is not } \tau\text{-right-persistent for } h_t\right] \\
&= \sum_{t=1}^{r-1} O\left(\frac{\tau}{\sqrt{d}}\right) = (r-1) \cdot O\left(\frac{\tau}{\sqrt{d}}\right),
\end{aligned}$$

where the second inequality is by the union bound and the last equality is due to the fact that eq. (7.2) holds for all h_t, $t \in [r]$. $\square$

7.1.3 Proof of Theorem 7.0.1

In this section, we show how to use Lemma 7.1.6 and Lemma 7.1.8 to ensure that the conditions of Claim 7.1.4 hold. Once the conditions are met, we prove Theorem 7.0.1.

Proof of Theorem 7.0.1. Let $G(A, B, E_{AB})$ be the (K, Δ)-good subgraph for f which we obtain from Lemma 7.1.6. Then $K\sqrt{\Delta} = \Theta(\frac{\varepsilon(f)\cdot 2^d}{\log d})$ and $E_{AB} \subseteq \mathcal{S}^-_f$. Without loss of generality, suppose that $G(A, B, E_{AB})$ is a (K, Δ)-right-good subgraph. Note that $G(A, B, E_{AB})$ satisfies the conditions (a) and (b) of Claim 7.1.4. We define $\sigma = K/2^d$, so that $\sigma\sqrt{\Delta} = \Theta(\frac{\varepsilon(f)}{\log d})$. Before proceeding with the main analysis, we rule out some simple cases with the following claim.

Claim 7.1.9. *Suppose any of the following hold: (a) $I_f^- \geq \sqrt{d}$. (b) $r \geq \frac{\sqrt{d}}{\log d}$. (c) $\sigma \leq \frac{r \cdot \log d}{\sqrt{d}}$. Then, for $(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}_{pair}(0, 1)$, we have $\mathbb{P}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x}) > f(\mathbf{y})] \geq \widetilde{\Omega}(\frac{\varepsilon(f)^2}{r\sqrt{d}})$.*

Proof. As we remarked, for $\tau = 1$, Alg. 3 has the same guarantees as the edge tester. By definition, the edge tester rejects with probability at least $\frac{I_f^-}{d}$. Therefore, (a) implies the conclusion, since if $I_f^- \geq \sqrt{d}$, then the edge tester succeeds with probability $\Omega(\frac{1}{\sqrt{d}})$. In addition, the edge tester rejects with probability $\Omega(\frac{\varepsilon(f)}{d})$ for all real-valued functions. Thus, (b) implies the conclusion, since if $r \geq \frac{\sqrt{d}}{\log d}$, then $\frac{\varepsilon(f)}{d} \geq \frac{\varepsilon(f)^2}{r\sqrt{d}\log d}$.

To see that (c) implies the conclusion, suppose $\sigma \leq \frac{r \cdot \log d}{\sqrt{d}}$. Recall that $\sigma\sqrt{\Delta} = \Theta(\frac{\varepsilon(f)}{\log d})$. Thus,

$$\sigma \cdot \Delta = \frac{(\sigma\sqrt{\Delta})^2}{\sigma} = \sigma^{-1} \cdot \Theta\left(\left(\frac{\varepsilon(f)}{\log d}\right)^2\right) = \Omega\left(\frac{\varepsilon(f)^2\sqrt{d}}{r(\log d)^3}\right).$$

Next, recall that $E_{AB} \subseteq \mathcal{S}_f^-$ and since G is (K, Δ)-right-good, we have $|E_{AB}| = |B| \cdot \Delta = K \cdot \Delta$. Thus, $I_f^- \geq \frac{K\Delta}{2^d} = \sigma \cdot \Delta$. Therefore, the edge tester rejects with probability $\frac{I_f^-}{d} \geq \frac{\sigma\Delta}{d} \geq \Omega\left(\frac{\varepsilon(f)^2}{r\sqrt{d} \cdot (\log d)^3}\right)$. $\square$

In light of Claim 7.1.9, we henceforth assume that $I_f^- \leq \sqrt{d}$, $r \leq \frac{\sqrt{d}}{\log d}$, and $\sigma \geq \frac{r \cdot \log d}{\sqrt{d}}$. Note that this implies $\frac{r \cdot \log d}{\sqrt{d}} \leq 1$ and $\frac{r \cdot \log d}{\sqrt{d}} \leq \sigma \leq 1$. Since the tester iterates through all values of τ that are powers of 2 and at most $\sqrt{\frac{d}{\log d}}$, we can fix the unique value τ' satisfying

$$\tau' \leq \frac{\sigma}{r - 1}\sqrt{\frac{d}{\log d}} \leq 2\tau'.$$

Note that these bounds imply that $\tau' \geq \frac{1}{2} \cdot \sqrt{\log d}$. Moreover, since $I_f^- \leq \sqrt{d}$, we can apply

Lemma 7.1.8 to conclude that the fraction of vertices in $\{0,1\}^d$ which are not $(\tau'-1)$-right-persistent for f is at most $\frac{c\cdot\tau'\cdot(r-1)}{\sqrt{d}}$ for some constant $c>0$. Using our upper bound on τ', this value is at most $\frac{c\cdot\sigma}{\sqrt{\log d}} \leq \frac{\sigma}{100}$ for sufficiently large d. Since $|B| = \sigma \cdot 2^d$, we conclude that at least $\frac{99}{100}|B|$ vertices in B are $(\tau'-1)$-right-persistent. Finally, we show that $\Delta \cdot \tau' \ll d$.

$$\Delta \cdot \tau' \leq \Delta \cdot \frac{\sigma}{r-1}\sqrt{\frac{d}{\log d}} = \frac{1}{r-1}\cdot \sigma\sqrt{\Delta}\sqrt{\frac{d\Delta}{\log d}} \leq \frac{1}{r-1}\cdot \Theta\left(\frac{\varepsilon(f)}{\log d}\right)\frac{d}{\sqrt{\log d}} \ll d,$$

and therefore condition (c) of Claim 7.1.4 holds. We have shown that all conditions, (a), (b), and (c) of Claim 7.1.4 hold. Therefore, for $(\mathbf{x},\mathbf{y}) \sim \mathcal{D}_{\texttt{pair}}(0,\tau')$, we have

$$\mathbb{P}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x}) > f(\mathbf{y})] \geq \frac{C'\cdot\tau'}{d}\cdot\sigma\cdot\Delta \text{ for some constant } C'>0.$$

Using our lower bound on τ', it follows that

$$\mathbb{P}_{\mathbf{x},\mathbf{y}}[f(\mathbf{x}) > f(\mathbf{y})] \geq \frac{C'\cdot\tau'\cdot\sigma\cdot\Delta}{d} \geq \frac{1}{2}\cdot\frac{\sigma}{r-1}\sqrt{\frac{d}{\log d}}\cdot\frac{C'\cdot\sigma\cdot\Delta}{d} = \frac{C'\cdot\sigma^2\cdot\Delta}{2(r-1)\sqrt{d\log d}}.$$

Since $(\sigma\sqrt{\Delta})^2 = \Theta\left(\left(\frac{\varepsilon(f)}{\log d}\right)^2\right)$, then:

$$\mathbb{P}_{(\mathbf{x},\mathbf{y})\sim\mathcal{D}_{\texttt{pair}}(0,\tau')}[f(\mathbf{x}) > f(\mathbf{y})] \geq \frac{C'\varepsilon(f)^2}{2(r-1)\sqrt{d}(\log d)^{5/2}} = \widetilde{O}\left(\frac{\varepsilon(f)^2}{r\sqrt{d}}\right).$$

Therefore, $\widetilde{O}(\frac{r\sqrt{d}}{\varepsilon(f)^2})$ iterations of the tester with $(\mathbf{x},\mathbf{y}) \sim \mathcal{D}_{\texttt{pair}}(0,\tau')$ will suffice for the tester to detect a violation to monotonicity and reject with high probability. This concludes the proof of Theorem 7.0.1. □

7.2 An $\Omega(r\sqrt{d})$ Lower Bound for Non-adaptive One-Sided Testers

In this section, we prove Theorem 7.0.2 which gives a lower bound on the query complexity of testing monotonicity of real-valued functions with 1-sided error nonadaptive testers. Fischer et al. proved Theorem 7.0.2 for the special case of $r = 2$ [FLN+02, Theorem 19]. Our proof of Theorem 7.0.2 is a natural extension of their construction to the more general case of $r \in [2, \sqrt{d}]$.

Proof. Fix $r \in [2, \sqrt{d}]$. We show that every nonadaptive, 1-sided error tester for functions over $\{0,1\}^d$ with image size r must make $\Omega(r\sqrt{d})$ queries. This implies Theorem 7.0.2, since Blais et al. [BBM12, Theorem 1.6] proved an $\Omega(\min(d, r^2))$ lower bound for all testers.

For convenience, assume d is an odd perfect square and r divides $2\sqrt{d}+1$. We partition the points $z \in \{0,1\}^{d-1}$ into levels, according to their weight $|z|$. We group levels from the middle of the $(d-1)$-dimensional hypercube into r blocks of width w, where $w = \frac{2\sqrt{d}+1}{r}$. Specifically, for each $j \in [r]$, we define the set

$$Z_j = \left\{ z \in \{0,1\}^{d-1} \colon (j-1)w \le |z| - \left(\frac{d-1}{2} - \sqrt{d}\right) \le jw \right\}.$$

Observe that

$$\bigcup_{j=1}^{r} Z_j = \left\{ z \in \{0,1\}^{d-1} \colon -\sqrt{d} \le \left| |z| - \frac{d-1}{2} \right| \le \sqrt{d} \right\}$$

and Z_j is a block of w consecutive levels from the middle of the $(d-1)$-dimensional hypercube. For each $i \in [d]$, we define function $f_i \colon \{0,1\}^d \to [r]$ as follows. For $x \in \{0,1\}^d$ and $i \in [d]$, let x_{-i} be the point in $\{0,1\}^{d-1}$ obtained by removing the i'th coordinate from x. Given $x \in \{0,1\}^d$, we define

$$f_i(x) = \begin{cases} r & \text{if } |x_{-i}| > \frac{d-1}{2} + \sqrt{d}, \\ 1 & \text{if } |x_{-i}| < \frac{d-1}{2} - \sqrt{d}, \\ j + (1 - x_i) & \text{if } x_{-i} \in Z_j. \end{cases}$$

Claim 7.2.1. *For all $i \in [d]$, $\varepsilon(f_i) = \Omega(1)$.*

Proof. Consider the matching of edges $M = \left\{ (x,y) \colon x_i = 0,\ y_i = 1, \text{ and } x_{-i} = y_{-i} \in \bigcup_{j=1}^{r} Z_j \right\}$.

Observe that all pairs in M are edges violated by f_i and $|M| = \Omega(1) \cdot 2^d$. $\square$

Every 1-sided error tester must accept if the function values on the points it queried are consistent with a monotone function. We say that a set $Q \subseteq \{0,1\}^d$ of queries *contains a violation for a function* f if there exist $x, y \in Q$ such that $x \prec y$ and $f(x) > f(y)$. If Q does not contain a violation, then the function values on Q are consistent with a monotone function.

Claim 7.2.2. *For all sets $Q \subseteq \{0,1\}^d$ of queries,*

$$\left|\{i \in [d] \colon Q \text{ contains a violation for } f_i\}\right| < w \cdot |Q|.$$

Proof. We use the following claim due to [BCP+20].

Claim 7.2.3 (Lemma 3.18 of [BCP+20], rephrased)**.** *Let $c, d \in \mathbb{N}$ and $Q \subseteq \{0,1\}^d$. Given $x, y \in Q$, define $\texttt{cap}_c(x,y)$ as follows. If x and y differ on at least c coordinates, then let $\texttt{cap}_c(x,y)$ be the set of the first c coordinates on which x and y differ. Otherwise, let $\texttt{cap}_c(x,y)$ be the set of all coordinates on which x and y differ. Define $\texttt{cap}_c(Q) = \bigcup_{x,y\in Q} \texttt{cap}_c(x,y)$. Then $|\texttt{cap}_c(Q)| \le c(|Q|-1)$.*

By design of f_i, if Q contains a violation for f_i, then there exist $x, y \in Q$ that differ in at most w coordinates, one of which is i. Then $i \in \texttt{cap}_w(x,y)$ and thus $i \in \texttt{cap}_w(Q)$. Therefore, by Claim 7.2.3,

$$\left|\{i \in [d] \colon Q \text{ contains a violation for } f_i\}\right| \le |\texttt{cap}_w(Q)| \le w(|Q|-1) < w \cdot |Q|.$$

This completes the proof of Claim 7.2.2. $\square$

Now, consider a nonadaptive tester T with 1-sided error that makes $q = q(\varepsilon, d, r)$ queries. Let $\boldsymbol{Q} \subseteq \{0,1\}^n$ denote the random set of queries of size q made by T. Using linearity of expectation and Claim 7.2.2,

$$\sum_{i=1}^{d} \mathbb{P}[T \text{ finds a violation for } f_i] = \mathop{\mathbb{E}}_{\boldsymbol{Q}}\Big[\big|\{i \in [d] \colon \boldsymbol{Q} \text{ contains a violation for } f_i\}\big|\Big] < w \cdot q$$

and therefore there exists $i \in [d]$ such that

$$\mathbb{P}\left[T \text{ finds a violation for } f_i\right] < \frac{w \cdot q}{d} = \frac{(2\sqrt{d}+1) \cdot q}{rd} < \frac{3q}{r\sqrt{d}},$$

whereas, if T is a valid monotonicity tester, then we must have $\mathbb{P}[T \text{ finds a violation for } f_i] \geq 2/3$. Therefore, for T to be a valid monotonicity tester, we require that it makes $q \geq \frac{2}{9} r\sqrt{d} = \Omega(r\sqrt{d})$ queries. $\square$

CHAPTER 8

Approximating the Distance to Monotonicity of Real-Valued Functions

In this chapter, we prove the following theorem on approximating the distance to monotonicity of real-valued functions over the hypercube. Our proof follows by showing that the algorithm of Pallavoor et al. [PRW22] can be employed for real-valued functions, using our generalized directed Talagrand inequality for real-valued functions Theorem 6.0.3 (here we use the $n = 2$ case). This result was originally published in [BKR23].

Theorem 8.0.1. *There exists a nonadaptive $O(\sqrt{d \log d})$-approximation algorithm for the distance to monotonicity that, given a parameter $\alpha \in (0, 1)$ and oracle access to a function $f\colon \{0,1\}^d \to \mathbb{R}$ that is α-far from monotone, makes poly$(d, 1/\alpha)$ queries.*

To prove Theorem 8.0.1, it is sufficient to give a tolerant tester for monotonicity of functions $f : \{0,1\}^d \to \mathbb{R}$. A *tolerant* tester for monotonicity gets two parameters $\varepsilon_1, \varepsilon_2 \in (0, 1)$, where $\varepsilon_1 < \varepsilon_2$, and oracle access to a function f. It has to accept with probability at least $2/3$ if f is ε_1-close to monotone and reject with probability at least $2/3$ if f is ε_2-far from monotone. Our tester distinguishes functions that are $\widetilde{O}(\varepsilon/\sqrt{d})$-close to monotone from those that are ε-far. Suppose this tolerant tester has query complexity $q(\varepsilon, d)$. Then, by [PRW22, Theorem A.1], it can be converted to a distance approximation algorithm with the required approximation guarantee and query complexity $O(q(\alpha, d) \log\log(1/\alpha))$. The following lemma, proved by Pallavoor et al. for the special case of Boolean functions, states our result on tolerant testing of monotonicity. Together with the conversion procedure from tolerant testing to distance approximation discussed above, it implies Theorem 8.0.1.

Lemma 8.0.2. *There exists a fixed universal constant $c \in (0,1)$ and a nonadaptive algorithm,* `ApproxMono`*, that gets a parameter $\varepsilon \in (0, 1/2)$ and oracle access to a function $f\colon \{0,1\}^d \to \mathbb{R}$, makes $\operatorname{poly}(d, 1/\varepsilon)$ queries and returns* `close` *or* `far` *as follows:*

1. *If $\varepsilon(f) \le \frac{c \cdot \varepsilon}{\sqrt{d} \log d}$ it outputs* `close` *with probability at least 2/3.*
2. *If $\varepsilon(f) \ge \varepsilon$ it outputs* `far` *with probability at least 2/3.*

Proof. We show that Algorithm `ApproxMono` of Pallavoor et al. [PRW22], presented as Alg. 4, works for real-valued functions. At a high level, the algorithm uses the fact that a function that is far from monotone violates many edges or has a large matching of violated edges of a special type. The first subroutine estimates the number of edges violated by the function by sampling edges uniformly at random and checking if they violate monotonicity. The second subroutine estimates the size of the special type of matching of violated edges. If either of these estimates is large enough, the algorithm outputs `far`. Otherwise, it outputs `close`.

The class of matchings sought by the algorithm is parametrized by a subset of the coordinates $S \subseteq [d]$. The special property of these matchings is that one can verify locally whether a given point is matched by querying its neighbors and their neighbors.

To estimate the size of the matching parametrized by S, the algorithm estimates the probability of the following event $\texttt{Capture}(x, S, f)$. We denote by $x^{(i)}$ the point in $\{0,1\}^d$ whose i-th coordinate is equal to $1 - x_i$ and the remaining coordinates are the same as in x.

Definition 8.0.3 (Capture Event)**.** *For a function $f\colon \{0,1\}^d \to \mathbb{R}$, a set $S \subseteq [d]$, and a point $x \in \{0,1\}^d$, let $\texttt{Capture}(x, S, f)$ be the following event:*

1. *There exists an index $i \in S$ such that, for $y = x^{(i)}$, the edge between x and y is violated by f. (Note that the edge between x and y is (x, y) if $x_i = 0$; otherwise, it is (y, x).)*
2. *Neither the edges of the form $(y, y^{(j)})$ nor the edges of the form $(y^{(j)}, y)$, where $j \in S \setminus \{i\}$, are violated by f.*

Denote $\mathbb{P}_{\mathbf{x} \sim \{0,1\}^d} [\texttt{Capture}(\mathbf{x}, S, f)]$ by $\mu_f(S)$.

Observe that $\mu_f(S)$ can be estimated nonadaptively, by sampling vertices x uniformly and independently at random and querying f on x and all points that differ from x in at most two coordinates.

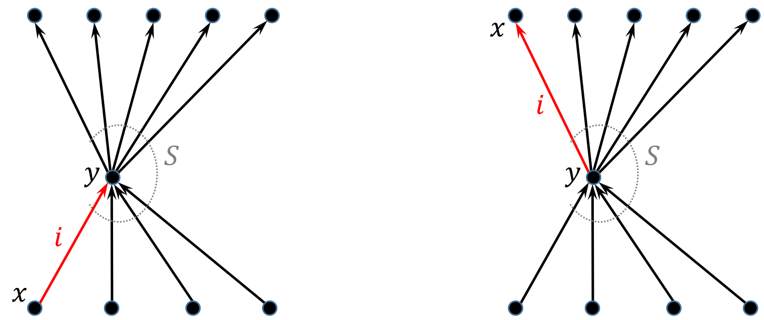


Figure 8.1: *An illustration to Definition 8.0.3. Two cases are depicted: when $x \prec y$ and when $y \prec x$.*

Algorithm 4 Algorithm `ApproxMono`

Input: Parameters $\varepsilon \in (0, 1/2)$ and dimension d; oracle access to function $f\colon \{0,1\}^d \to \mathbb{R}$.

1: Calculate $\hat{\nu}$, an estimate of the fraction of the hypercube edges that are violated by f, up to an additive error $\frac{\varepsilon}{4\sqrt{d \log d}}$.
2: **if** $\hat{\nu} \geq 3\varepsilon/(4\sqrt{d}\log d)$ **then** return `far`.
3: **for** $t \in \{1, 2, 4, \ldots, 2^{\lfloor \log_2 d \rfloor}\}$ **do**
4: Sample $S \subseteq [d]$ by including each coordinate $i \in [d]$ independently with probability $1/t$.
5: Calculate $\hat{\mu}$, an estimate of $\mu_f(S) = \mathbb{P}_{\mathbf{x} \sim \{0,1\}^d}[\mathtt{Capture}(\mathbf{x}, S, f)]$ up to an additive error $\frac{c' \cdot \varepsilon}{4\sqrt{d \log d}}$ for some constant $c' > 0$.
6: **if** $\hat{\mu} \geq \frac{3c' \cdot \varepsilon}{4\sqrt{d \log d}}$ **then** return `far`.
7: Return `close`.

The first component of the analysis is the observation that both the fraction of violated edges and $\mu_f(S)$, for every $S \subseteq [d]$, provide a good lower bound on the distance to monotonicity. We state this observation without proof because the proof for the Boolean case from [PRW22] extends to the general case verbatim. Intuitively, it tells us that, assuming that the two estimates computed by Alg. 4 are accurate, if one of the estimates is large enough then the input function is far from monotone.

Observation 8.0.4 ([PRW22])**.** *For every function $f : \{0,1\}^d \to \mathbb{R}$, the distance $\varepsilon(f)$ is at least half the fraction of the hypercube edges that are violated by f and $\varepsilon(f) \geq \mu_f(S)/2$ for all $S \subseteq [d]$.*

The second (and the main) component of the analysis for the Boolean case is [PRW22, Lemma 2.8], which relies on the robust isoperimetric inequality of [KMS18]. We generalize this lemma to real-valued functions in Lemma 8.0.5 below. Intuitively, it states that, if function f violates few edges then, for one of the $O(\log d)$ choices of the parameter tried in Step 3 of Alg. 4 for sampling set $\boldsymbol{S}$, the expectation of $\mu_f(\boldsymbol{S})$ is large in terms of $\varepsilon(f)$. That is, again assuming that the estimates computed by Alg. 4 are accurate, if none of the estimates is large enough then the input function is close to monotone.

Equipped with Observation 8.0.4 and Lemma 8.0.5, it is easy to convert the intuition above into the formal proof that the algorithm satisfies the guarantees of Lemma 8.0.2. This part of the proof uses standard techniques and is the same as for the case of Boolean functions described in [PRW22], so we omit it. This completes the proof of Lemma 8.0.2. $\square$

It remains to prove the following lemma, which crucially relies on our robust isoperimetric inequality for real-valued functions. We generalize the quantities used by Pallavoor et al. so that the proof is syntactically similar to that for the case of Boolean functions. One subtlety that arises in the case of real-valued functions is that a vertex can be incident to violated edges of both colors. In constrast, in the case of Boolean functions, each vertex can be adjacent either to violated edges going to higher-weight vertices or to violated edges going to lower-weight vertices, that is, it cannot be incident on both blue and red violated edges.

Lemma 8.0.5 (Generalized version of Lemma 2.8 of [PRW22])**.** *Let $f\colon \{0,1\}^d \to \mathbb{R}$ be ε-far from monotone, with fraction of violated edges smaller than $\frac{\varepsilon}{\sqrt{d \log d}}$. Then, for some $t \in \{1, 2, 4, \ldots, 2^{\lfloor \log_2 d \rfloor}\}$, it holds that*

$$\mathop{\mathbb{E}}_{\substack{\boldsymbol{S} \subseteq [d] \\ i \in \boldsymbol{S} \text{ w.p. } 1/t}} [\mu_f(\boldsymbol{S})] = \Omega\left(\frac{\varepsilon}{\sqrt{d \log d}}\right).$$

Proof. For $x \in \{0,1\}^d$, let $U_f^-(x)$ denote the number of violated edges incident on x (both incoming and outgoing). Consider the following 2-coloring of the edges in $\mathcal{S}_f^-$:

$$\texttt{col}((x,y)) = \begin{cases} \text{red} & \text{if } U_f^-(x) \geq U_f^-(y); \\ \text{blue} & \text{if } U_f^-(x) < U_f^-(y). \end{cases}$$

This coloring ensures that, in the isoperimetric inequality, each edge is counted towards the endpoint incident on the largest number of violated edges (and, in case of a tie, towards the lower endpoint).

The proof of [PRW22, Lemma 2.8] relies on the existence of a set $B \subseteq \{0,1\}^d$ and a color $b \in \{\text{red}, \text{blue}\}$ that satisfy the following two properties:

1. no edge violated by f has both endpoints in the set B;
2. $\frac{1}{2^d} \sum_{x \in B} \sqrt{I_{f,b}^-(x)} = \Omega(\varepsilon)$.

To obtain the set B and the color b, we partition $\{0,1\}^d$ into two sets:

$$B_{\text{even}} = \left\{x \in \{0,1\}^d : |x| \text{ is even}\right\},$$
$$B_{\text{odd}} = \left\{x \in \{0,1\}^d : |x| \text{ is odd}\right\}.$$

The sets B_{even} and B_{odd} clearly satisfy property 1. Note that, for the case of Boolean functions, Pallavoor et al. partition the domain points according to their function values instead of the parity of their weight to guarantee property 1.

By Theorem 6.0.3 (invoked for the special case of $n = 2$),

$$\sum_{x \in B_{\text{even}}} \sqrt{I_{f,\text{red}}^-(x)} + \sqrt{I_{f,\text{blue}}^-(x)} + \sum_{x \in B_{\text{odd}}} \sqrt{I_{f,\text{red}}^-(x)} + \sqrt{I_{f,\text{blue}}^-(x)} \geq C \cdot \varepsilon \cdot 2^d\,.$$

By averaging, there exist a color $b \in \{\text{red}, \text{blue}\}$ and a set $B \in \{B_{\text{even}}, B_{\text{odd}}\}$ that satisfy

$$\sum_{x \in B} \sqrt{I^-_{f,b}(x)} \geq \frac{C}{4} \cdot \varepsilon \cdot 2^d \,. \tag{8.1}$$

Therefore, property 2 also holds. Note that due to the partition into even-weight and odd-weight points we loose an extra factor of 2 as compared to Pallavoor et al. in the contribution of the set B and the color b to the isoperimetric inequality. This results in a loss by a factor of 2 (hidden in the Ω-notation) in the lower bound in Lemma 8.0.5.

The rest of the proof is the same as in [PRW22], so we only summarize the key steps. We proceed by partitioning the points $x \in B$ into buckets $B_{t,s}$ for $t, s \in \{1, 2, 4, \ldots, 2^{\lfloor \log_2 d \rfloor}\}$, where $t \geq s$, as follows:

$$B_{t,s} = \{x \in B \;:\; t \leq U^-_f(x) < 2t \text{ and } s \leq I^-_{f,b}(x) < 2s\} \,.$$

Each vertex $x \in B_{t,s}$ is incident on between t and $2t$ violated edges and between s and $2s$ edges colored b, which are counted towards x in property 2.

When the set $\boldsymbol{S}$ is chosen so that each coordinate is included with probability $1/t$, it holds for all $x \in B_{t,s}$ that the event $\texttt{Capture}(x, \boldsymbol{S}, f)$ occurs with probability $\Omega(s/t)$. Using this claim, one can lower bound the contribution of each bucket towards $\mathbb{E}_{\boldsymbol{S} \subseteq [d]}[\mu_f(\boldsymbol{S})]$. By combining the contributions of the buckets with the same value s and applying the Cauchy-Schwartz inequality, one obtains

$$\sum_{t \in \{1,2,4,\ldots,2^{\lfloor \log_2 d \rfloor}\}} \mathop{\mathbb{E}}_{\substack{\boldsymbol{S} \subseteq [d] \\ i \in \boldsymbol{S} \text{ w.p. } 1/t}} [\mu_f(\boldsymbol{S})] = \Omega\left(\frac{1}{2^d} \cdot \frac{(\sum_{t,s: t \geq s} |B_{t,s}| \sqrt{s})^2}{\sum_{t,s: t \geq s} |B_{t,s}| t}\right). \tag{8.2}$$

We lower bound the sum in the numerator using (8.1) and upper bound the sum in the denominator using the assumed upper bound on the number of violated edges. As a result, we get that the left-hand side of (8.2) is $\Omega(\varepsilon \sqrt{\log d}/\sqrt{d})$. Averaging over the $O(\log d)$ possible values of t yields Lemma 8.0.5. $\square$

Part III

Sample-Based Testing and Learning of k-Monotone Functions

CHAPTER 9

Introduction

A function $f\colon \mathcal{X} \to \mathbb{R}$ over a partial order $\mathcal{P} = (\mathcal{X}, \preceq)$ is k-*monotone* if there does not exist a chain of $k+1$ points $x_1 \prec x_2 \prec \cdots \prec x_{k+1}$ for which (a) $f(x_{i+1}) - f(x_i) < 0$ when i is odd and (b) $f(x_{i+1}) - f(x_i) > 0$ when i is even. When $k = 1$, these are the *monotone* functions, which are the non-decreasing functions with respect to $\preceq$. Monotone and k-monotone *Boolean* functions over domains $\{0,1\}^d$, $[n]^d$, and $\mathbb{R}^d$ have been the focus of a significant amount of research in property testing and computational learning theory. We give an overview of the literature in Section 9.2.

The field of property testing is concerned with the design and analysis of sub-linear time randomized algorithms for determining if a function has, or is far from having, some specific property. A key aspect in the definition of a property testing algorithm is the type of access it has to the function. Early works on property testing, e.g. [RS96, GGR98], focused on the notion of *query-based* testers, which are allowed to observe the value of the function on any point of their choosing, and since then this has become the standard model. The weaker notion of *sample-based* testers, which can only view the function on independent uniform samples, was also considered by [GGR98] and has received some attention over the years, see e.g. [KR00, BBBY12, FLV15, GR16, FH23]. Sample-based algorithms are considered more natural in many settings, for example in computational learning theory, where they are the standard model. In fact, sample-based testing and learning are closely related problems; given a learning algorithm, it is always possible to design a testing algorithm with the same sample complexity, up to an additive $\mathrm{poly}(1/\varepsilon)$ factor[1].

[1]See Lemma 11.6.1 for a precise statement. Also, note that if the learning algorithm is *proper*, then the

For many fundamental properties, there is still a large gap between how much we know in the query-based vs the sample-based models. Monotonicity (and k-monotonicity) is such a property; despite a vast body of research on query-based monotonicity testing over the hypercube $\{0,1\}^d$, the only work we know of which considers this problem in the sample-based model is [GGL+00], who gave an upper bound of $O(\sqrt{2^d/\varepsilon})$ and a matching lower bound for the case when $\varepsilon = O(d^{-3/2})$ on the number of samples needed to test monotonicity of functions $f\colon \{0,1\}^d \to \{0,1\}$. The upper bound for learning monotone Boolean functions due to [BT96, LRV22] also implies a testing upper bound of $\exp(O(\frac{1}{\varepsilon}\sqrt{d}))$. Thus, this question has been wide open for $\varepsilon \gg d^{-3/2}$.

Our work addresses this gap in the monotonicity testing literature, proving a lower bound which matches the learning upper bound for all ε at most some constant. More generally, we prove tight lower bounds for k-monotonicity testing of functions, $f\colon \{0,1\}^d \to [r]$, i.e. functions with image size at most r. To round out our results, we also give an improved learning algorithm for k-monotone functions over $\mathbb{R}^d$ under product distributions whose sample complexity matches our sample-based testing lower bound, up to poly-logarithmic factors in the exponent.

9.1 Results

Before explaining our results and the context for them, we first provide some terminology and basic notation. Given a domain $\mathcal{X}$ and a distribution μ over $\mathcal{X}$, we denote the Hamming distance between two functions $f, g\colon \mathcal{X} \to \mathbb{R}$ under μ by $d_\mu(f,g) = \mathbb{P}_{x\sim\mu}[f(x) \neq g(x)]$. We say that f is ε-far from k-monotone if $d_\mu(f,g) \geq \varepsilon$ for every k-monotone function g. The results in this paper pertain to sample-based testing and learning of k-monotone functions with respect to Hamming distance. We use the following terminology:

---

time complexity is also preserved. If the learning algorithm is *improper*, then there is a time complexity blow-up, but the sample complexity is still preserved.

- The *example oracle* for f under μ, denoted by $EX(f, \mu)$, when queried, generates an example $(x, f(x))$ where x is sampled according to μ.

- A *sample-based k-monotonicity tester* under μ is a randomized algorithm which is given access to $EX(f, \mu)$ for an arbitrary input function f and satisfies the following: (a) if f is k-monotone, then the algorithm accepts with probability at least $2/3$, and (b) if f is ε-far from k-monotone, then the algorithm rejects with probability at least $2/3$. The tester has *one-sided error* if in case (a) it accepts with probability 1.

- A *sample-based learning algorithm for k-monotone functions* under μ is a randomized algorithm which is given access to $EX(f, \mu)$ for an arbitrary k-monotone input function f and outputs a hypothesis h such that $d_\mu(f, h) \leq \varepsilon$ with probability at least $1 - \delta$. If left unspecified, $\delta = 1/3$.

In all of the above definitions if μ is unspecified, then it is the uniform distribution. Testing and learning are closely related problems; any sample-based learning algorithm can be used to construct a sample-based tester with the same sample complexity. We refer to this transformation as the testing-by-learning reduction and although this is not a new idea we provide a proof in Section 11.6 for completeness.

9.1.1 Sample-Based Testing and Learning on the Hypercube

The problem of *learning* monotone Boolean functions over the hypercube $\{0, 1\}^d$ was studied by [BT96] who proved an upper bound[2] of $\exp(O(\min\{\frac{1}{\varepsilon}\sqrt{d}, d\}))$ for improper learning and very recently by [LRV22, LV23] who obtained the same upper bound for agnostic proper learning. The improper learning upper bound was extended by [BCO+15] who showed matching upper and lower bounds of $\exp(\Theta(\min\{\frac{k}{\varepsilon}\sqrt{d}, d\}))$ for learning k-monotone Boolean

[2]We remark that any function over $\{0, 1\}^d$ can be learned exactly with $O(d2^d) = \exp(O(d))$ samples by a coupon-collector argument. Combining this with the $\exp(O(\frac{1}{\varepsilon}\sqrt{d}))$ upper bound by [BT96] yields $\exp(O(\min\{\frac{1}{\varepsilon}\sqrt{d}, d\}))$. We use this slightly clunkier notation involving the min to emphasize that our upper and lower bounds are matching in all parameter regimes.

functions for any $k \geq 1$. The testing-by-learning reduction shows that their upper bound also holds for sample-based *testing*. The only prior lower bound for sample-based testing that we're aware of is $\Omega(\sqrt{2^d/\varepsilon})$ when $\varepsilon = O(d^{-3/2})$ and $k = 1$ [GGL+00, Theorem 5]. Our main result is the following much more general lower bound for this problem, which we prove in Chapter 10.

Theorem 9.1.1 (Testing Lower Bound)**.** *There is an absolute constant $c > 0$ such that for all $\varepsilon \leq c$, every sample-based k-monotonicity tester for functions $f \colon \{0,1\}^d \to [r]$ under the uniform distribution has sample complexity*

$$\exp\left(\Omega\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

Even for the special case of sample-based monotonicity testing of Boolean functions ($k = 1$ and $r = 2$), Theorem 9.1.1 is already a new result, which matches the upper bound for learning by [BT96] and is the first lower bound to hold for $\varepsilon \gg d^{-3/2}$. Moreover, our lower bound is much more general, holding for all r, k, and is optimal in all parameters, d, r, k, ε, up to a constant factor in the exponent. We show a matching upper bound in Theorem 9.1.3.

We also note that the testing-by-learning reduction implies that the same lower bound holds for *learning* with samples. As we mentioned, this result was already known for Boolean functions (the $r = 2$ case) [BCO+15], but the general case of $r \geq 2$ was not known prior to our work.

Corollary 9.1.2 (Learning Lower Bound)**.** *There is an absolute constant $c > 0$ such that for every $\varepsilon \leq c$, every sample-based uniform-distribution learning algorithm for k-monotone functions $f \colon \{0,1\}^d \to [r]$ has sample complexity*

$$\exp\left(\Omega\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

On the upper bound side, a relatively straightforward argument extends the learning algorithm of [BCO+15] for Boolean k-monotone functions, to k-monotone functions with image size at most r. We give a short proof in Section 9.3. This shows that our lower bounds in Theorem 9.1.1 and Corollary 9.1.2 are tight.

Theorem 9.1.3 (Learning Upper Bound for Hypercubes)**.** *There is a uniform-distribution learning algorithm for k-monotone functions $f\colon \{0,1\}^d \to [r]$ which achieves error at most ε with time and sample complexity*

$$\exp\left(O\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

The testing-by-learning reduction again gives us the following corollary.

Corollary 9.1.4 (Testing Upper Bound for Hypercubes)**.** *There is a sample-based k-monotonicity tester for functions $f\colon \{0,1\}^d \to [r]$ with sample complexity*

$$\exp\left(O\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

Lastly, we consider the problem of sample-based testing with *one-sided error*. For monotonicity testing of functions $f\colon \{0,1\}^d \to \{0,1\}$ with *non-adaptive queries*, we know that one-sided and two-sided error testers achieve the same query-complexity (up to $\mathrm{polylog}(d, 1/\varepsilon)$ factors): there is a $\widetilde{O}(\sqrt{d}/\varepsilon^2)$ one-sided error upper bound due to [KMS18] and a $\widetilde{\Omega}(\sqrt{d})$ two-sided error lower bound due to [CWX17]. We show that the situation is quite different for *sample-based* monotonicity testing; while the sample complexity of two-sided error testers is $\exp(\Theta(\min\{\frac{1}{\varepsilon}\sqrt{d}, d\}))$, one-sided error testers require $\exp(\Theta(d))$ samples for all ε.

Theorem 9.1.5 (Testing with One-Sided Error)**.** *For every d, r, k, and $\varepsilon > 0$, sample-based k-monotonicity testing of functions $f\colon \{0,1\}^d \to [r]$ with one-sided error requires*

$\exp(\Theta(d))$ *samples.*

9.1.2 Sample-Based Testing and Learning in Continuous Product Spaces

Learning k-monotone Boolean-valued functions has also been studied over $\mathbb{R}^d$ with respect to product measures by [HY22] who gave an upper bound of $\exp(\widetilde{O}(\min\{\frac{k}{\varepsilon^2}\sqrt{d}, d\}))$ where $\widetilde{O}(\cdot)$ hides polylog factors of d, k, and $1/\varepsilon$. Our next result gives an upper bound which improves the dependence on ε from $1/\varepsilon^2$ to $1/\varepsilon$ in the exponent. By the same approach we used to generalize the upper bound in Theorem 9.1.3 to arbitrary $r \geq 2$, we get the same generalization for product spaces. We obtain the following upper bound which matches our lower bound for $\{0,1\}^d$ in Theorem 9.1.1 up to polylog factors of d, k, r, and $1/\varepsilon$. We say that a function $f : \mathbb{R}^d \to [r]$ is *measurable* if the set $f^{-1}(i)$ is measurable for every $i \in [r]$.

Theorem 9.1.6 (Learning Upper Bound for Product Spaces)**.** *Given an arbitrary product measure μ, there is a learning algorithm under μ for measurable k-monotone functions $f : \mathbb{R}^d \to [r]$ with time and sample complexity*

$$\exp\left(\widetilde{O}\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

The $\widetilde{O}(\cdot)$ hides polylogarithmic dependencies on d, r, k, and $1/\varepsilon$.

We prove Theorem 9.1.6 in Chapter 11. Once again the testing-by-learning reduction gives us the following corollary for sample-based testing.

Corollary 9.1.7 (Testing Upper Bound for Product Spaces)**.** *Given an arbitrary product measure μ, there is a k-monotonicity tester for measurable functions $f : \mathbb{R}^d \to [r]$ under μ with sample complexity*

$$\exp\left(\widetilde{O}\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

The $\widetilde{O}(\cdot)$ hides polylogarithmic dependencies on $d, r, k,$ and $1/\varepsilon$.

9.1.3 Proof Overviews

In this section we give an overview of our proofs for Theorem 9.1.1 and Theorem 9.1.6.

9.1.4 The Testing Lower Bound for Hypercubes

Our proof of Theorem 9.1.1 uses a family functions known as *Talagrand's random DNFs* introduced by [Tal96] which have been used by [BB21] and [CWX17] to prove lower bounds for monotonicity testing of Boolean functions $f\colon \{0,1\}^d \to \{0,1\}$ against adaptive and non-adaptive query-based testers. Very recently, they have also been used to prove lower bounds for tolerant testing by [CDL$^+$23].

To understand our construction, let us first consider the special case of monotonicity of Boolean functions, i.e. $k = 1$ and $r = 2$. We think of a DNF term as a point $t \in \{0,1\}^d$ which is said to be satisfied by $x \in \{0,1\}^d$ if $t \preceq x$, where $\preceq$ denotes the standard bit-wise partial order over $\{0,1\}^d$. Consider N randomly chosen terms $t^1, \dots, t^N$ each of width $|t^j| = w$. We will see later how to choose N and w. Let $B := \{x\colon \frac{d}{2} \le |x| \le \frac{d}{2} + \varepsilon\sqrt{d}\}$ and for each $j \in [N]$, let

$$U_j := \{x \in B\colon t^j \preceq x \text{ and } t^{j'} \npreceq x \text{ for all } j' \neq j\}$$

be the set of points in B which satisfy t^j and no other terms. Let $U := \bigcup_{j\in[N]} U_j$. Now observe that any two points lying in different U_j's are *incomparable* and therefore independently embedding an arbitrary monotone function into each U_j will result in a function which globally is monotone if one defines the function outside of U appropriately. Using this fact we can define two distributions $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$ as follows. Let A denote the set of points in $x \in \{0,1\}^d$ for which either $|x| > \frac{d}{2} + \varepsilon\sqrt{d}$ or $x \in B$ and $t^j, t^{j'} \preceq x$ for two different terms $j \neq j'$.

- $f \sim \mathcal{D}_{\mathsf{yes}}$ is drawn by setting $f(x) = 1$ if and only if $x \in A \cup \left(\bigcup_{j\in T} U_j\right)$ where $T \subseteq [N]$

contains each $j \in [N]$ with probability $1/2$, independently. Such a function is always monotone.

- $f \sim \mathcal{D}_{\texttt{no}}$ is drawn by setting $f(x) = 1$ if and only if $x \in A \cup R$ where R contains each $x \in U$ with probability $1/2$, independently. Such a function will be $\Omega(|U| \cdot 2^{-d})$-far from monotone with probability $\Omega(1)$ since its restriction with U is uniformly random.

Now, each $x \in U$ satisfies $\mathbb{P}_{f\sim\mathcal{D}_{\texttt{yes}}}[f(x) = 1] = \mathbb{P}_{f\sim\mathcal{D}_{\texttt{no}}}[f(x) = 1] = 1/2$ and for both distributions the events $f(x) = 1$ and $f(y) = 1$ are independent when x, y lie in different U_j's. Therefore, any tester will need to see at least two points from the same U_j to distinguish $\mathcal{D}_{\texttt{yes}}$ and $\mathcal{D}_{\texttt{no}}$. Roughly speaking, by birthday paradox this gives a $\Omega(\sqrt{N})$ lower bound on the number of samples. The lower bound is thus determined by the maximum number of terms N that can be used in the construction for which $|U| = \Omega(\varepsilon 2^d)$.

So how are N and w chosen? By standard concentration bounds, we have $|B| = \Omega(\varepsilon 2^d)$ and observe that a point $x \in B$ satisfies a random term with probability exactly $(|x|/d)^w$. We need U to contain a *constant fraction* of B, i.e. we need x to satisfy exactly 1 term with constant probability. The expected number of satisfied terms is $N \cdot (|x|/d)^w$ and, roughly speaking, we need this value to be $\Theta(1)$ for all $x \in B$. Applying this constraint to the case when $|x| = d/2$ forces us to pick $N \approx 2^w$. Now when $|x| = d/2 + \varepsilon\sqrt{d}$, the expected number of satisfied terms is $N \cdot 2^{-w} \cdot (1 + 2\varepsilon/\sqrt{d})^w \approx (1 + 2\varepsilon/\sqrt{d})^w$ and we are forced to choose $w \approx \sqrt{d}/\varepsilon$. The lower bound for sample-based monotonicity testing of $f\colon \{0,1\}^d \to \{0,1\}$ is then $\Omega(\sqrt{N}) \approx \exp(\Omega(\sqrt{d}/\varepsilon))$.

Let us now think about generalizing this construction to testing k-monotonicity of functions $f\colon \{0,1\}^d \to [r]$. The moral of the above argument is that the permitted number of terms is controlled by the number of distinct Hamming weights in the set B. We observe that for larger values of k and r we can partition B into $k(r-1)$ blocks as $B := B_1 \cup B_2 \cup \cdots \cup B_{k(r-1)}$ each with a window of Hamming weights of size only $\frac{\varepsilon\sqrt{d}}{k(r-1)}$. We are able to essentially repeat the above construction independently within each block wherein we can set $w \approx \frac{k(r-1)\sqrt{d}}{\varepsilon}$ and consequently $N \approx 2^{\frac{k(r-1)\sqrt{d}}{\varepsilon}}$.

For each block $i \in [k(r-1)]$, the random Talagrand DNF within block B_i is defined analogously to the above construction, except that it assigns function values from $\{i \bmod (r-1), i \bmod (r-1)+1\}$, instead of $\{0,1\}$. See Fig. 9.1 for an illustration. Since there are $k(r-1)$ blocks in total, the distribution $\mathcal{D}_{\texttt{yes}}$ only produces k-monotone functions. At the same time, a function $f \sim \mathcal{D}_{\texttt{no}}$ assigns uniform random $\{a, a+1\}$ values within each block $B_{m(r-1)+a}$. This results in a large number of long chains through $B_a \cup B_{(r-1)+a} \cup \cdots \cup B_{(k-1)(r-1)+a}$ which alternate between function value a and $a+1$. Considering the union of all such chains for $a = 0, 1, \ldots, r-2$ shows that f is $\Omega(\varepsilon)$-far from k-monotone with probability $\Omega(1)$.

9.1.5 The Learning Upper Bound for Product Spaces

As we discussed in Section 9.1, it suffices to prove Theorem 9.1.6 for the case of $r = 2$, i.e. learning functions $f : \mathbb{R}^d \to \{\pm 1\}$ under a product measure μ. We use a downsampling technique to reduce this problem to learning a discretized proxy of f over a hypergrid $[N]^d$ where $N = \Theta(kd/\varepsilon)$ with mild label noise. This technique has been used in previous works [GKW19, BCS20, HY22] and our proof borrows many technical details from [HY22].

Next, for N which is a power of 2, we observe that a k-monotone function $f : [N]^d \to \{\pm 1\}$ can be viewed as a k-monotone function over the hypercube $\{\pm 1\}^{d \log N}$ by mapping each point $x \in [N]^d$ to its bit-representation. We can then leverage a result of [BCO+15] which shows that all but a ε-fraction of the mass of the Fourier coefficients of k-monotone Boolean functions $f : \{0,1\}^d \to \{0,1\}$ is concentrated on the terms with degree at most $\frac{k\sqrt{d}}{\varepsilon}$. We can then use the Low-Degree Algorithm introduced by [LMN93] which was shown to work under random classification noise by [Kea98].

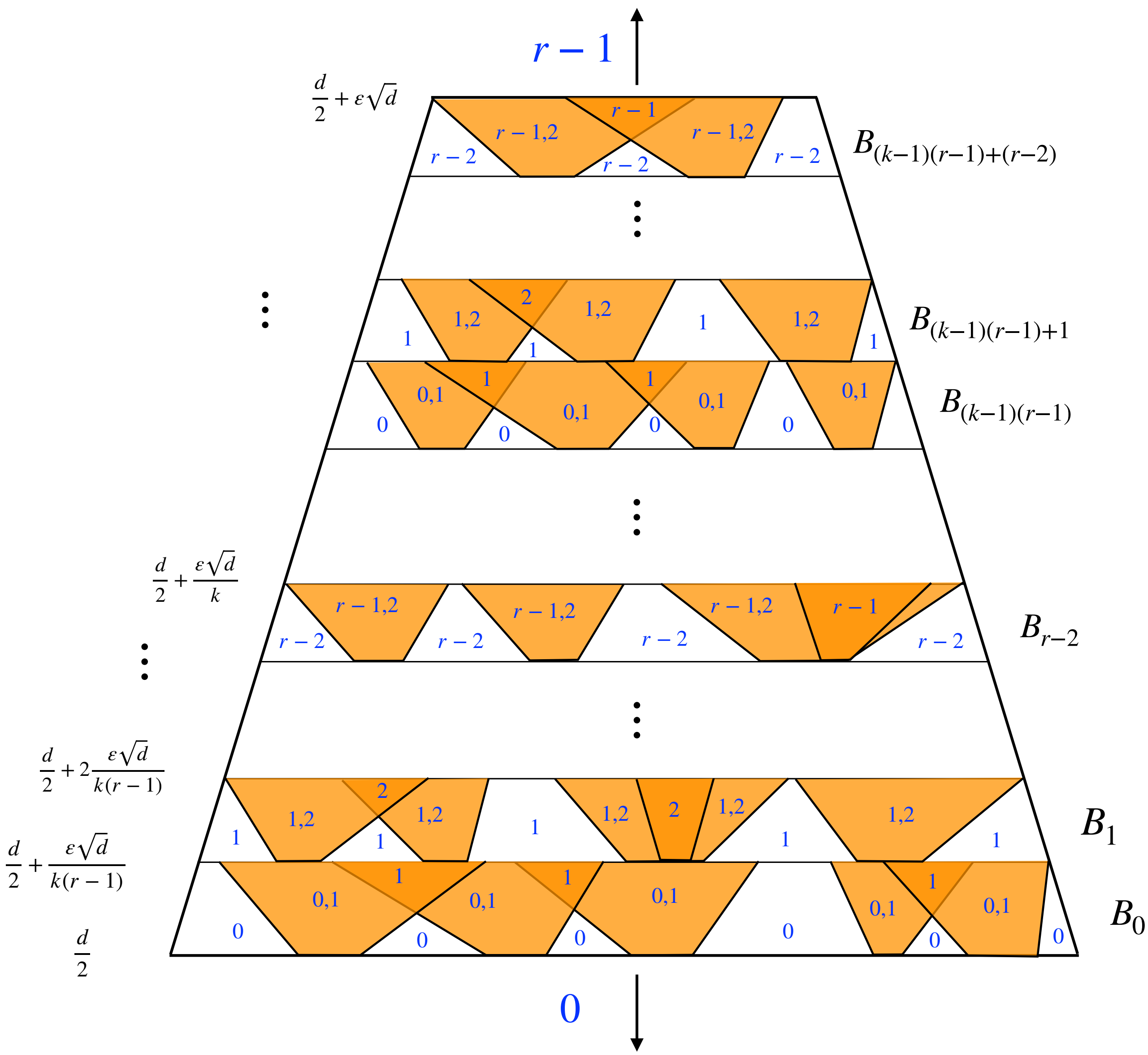


Figure 9.1: *An illustration of the construction used in our proof of Theorem 9.1.1. The image represents the set of points in the hypercube* $\{0,1\}^d$ *with Hamming weight in the interval* $[\frac{d}{2}, \frac{d}{2}+\varepsilon\sqrt{d})$*, increasing from bottom to top. The numbers on the left denote the Hamming weight of the points lying in the adjacent horizontal line. The* B_i *blocks are the sets of points contained between two adjacent horizontal lines. Each orange shaded region within* B_i *represents the set of points satisfied by a term* $t^{i,j}$*. The blue numbers represent the value that functions in the support of* $\mathcal{D}_{yes}$ *and* $\mathcal{D}_{no}$ *can take. We have used the notation "*$r-1,2$*" as shorthand for* $r-2, r-1$*.*

9.2 Related Work

Monotone functions and their generalization to k-monotone functions have been extensively studied within property testing and learning theory over the last 25 years. We highlight some of the results which are most relevant to our work. Afterwards, we discuss some selected works on sample-based property testing.

Sample-based monotonicity testing: Sample-based monotonicity testing of Boolean functions over the hypercube, $\{0,1\}^d$, was considered by [GGL+00] (see [GGL+00, Theorems 5 and 6]) who gave an upper bound of $O(\sqrt{2^d/\varepsilon})$ and a lower bound of $\Omega(\sqrt{2^d/\varepsilon})$ for $\varepsilon = O(d^{-3/2})$. Sample-based monotonicity testing over general partial orders was studied by [FLN+02] who gave a $O(\sqrt{N/\varepsilon})$ one-sided error tester for functions $f\colon D \to \mathbb{R}$ where D is any partial order on N elements. Sample-based monotonicity testing of functions on the line $f\colon [n] \to [r]$ was studied by [PRV18] who gave a one-sided error upper bound of $O(\sqrt{r/\varepsilon})$ and a matching lower bound of $\Omega(\sqrt{r})$ for all sample-based testers.

Query-based monotonicity testing: Monotonicity testing has been extensively studied in the standard query model [Ras99, EKK+00, GGL+00, DGL+99, LR01, FLN+02, HK03, AC06, HK08, ACCL07, Fis04, SS08, Bha08, BCSM12, FR10, BBM12, RRS+12, BGJ+12, CS13, CS14a, CST14, BRY14a, BRY14b, CDST15, CDJS17, KMS18, BB21, CWX17, BCS18, PRV18, BCS20, HY22, BKR23, BKKM23, BCS23b, BCS23a, CDL+23]. See Chapter 2 for an extended discussion on this body of work.

k-Monotonicity testing: The generalization to k-monotonicity testing has also been studied in the standard query model by [GKW19, CGG+19]. These works show that the query-complexity of non-adaptive one-sided error k-monotonicity testing is $\exp(\widetilde{\Theta}(\sqrt{d}))$ for all $k \geq 2$, demonstrating an interesting separation between (1-)monotonicity and 2-monotonicity.

Learning monotone functions: Monotone Boolean functions $f\colon \{0,1\}^d \to \{0,1\}$ were studied in the context of learning theory by [BT96] who showed that they can be (improperly) learned to error ε under the uniform distribution with $\exp(O(\frac{1}{\varepsilon}\sqrt{d}))$ time and samples. Very recent works [LRV22, LV23] have given *agnostic proper* learning algorithms with the same complexity.

Learning k-monotone functions: The result of [BT96] was generalized by [BCO+15] who gave upper and lower bounds of $\exp(\Theta(\frac{k}{\varepsilon}\sqrt{d}))$ for learning k-monotone Boolean functions $f\colon \{0,1\}^d \to \{0,1\}$. For Boolean functions over hypergrids $f\colon [n]^d \to \{0,1\}$, [CGG+19] gave an upper bound of $\exp(\widetilde{O}(\min(\frac{k}{\varepsilon^2}\sqrt{d}, d)))$ where $\widetilde{O}(\cdot)$ hides polylog factors of $d, k, 1/\varepsilon$. This result was generalized to functions $f\colon \mathbb{R}^d \to \{0,1\}$ under product measures by [HY22].

Sample-based property testing: The notion of sample-based property testing was first presented and briefly studied by [GGR98]. Broader studies of sample-based testing and its relationship with query-based testing have since been given by [FGL14, FLV15, GR16]. A characterization of properties which are testable with a constant number of samples was given by [BY19].

As we mentioned, sample-based algorithms are the standard model in learning theory, and learning requires at least as many samples as testing for every class of functions. Thus, it is natural to ask, when is testing *easier* than learning in terms of sample complexity? This question is referred to as *testing vs learning* and has been studied by [KR00] and more recently by [BFH21, FH23].

There has also been work studying models that interpolate between query-based and sample-based testers. For instance, [BBBY12] introduced the notion of *active testing*, where the tester may make queries, but only on points from a polynomial-sized batch of unlabeled samples drawn from the underlying distribution. This was inspired by the notion of *active learning* which considers learning problems under this access model.

Sample-based convexity testing of sets over various domains has also seen some recent attention [CFSS17a, BMR19a, BMR19b, BBH23].

9.3 Learning Functions with Bounded Image Size: Proof of Theorem 9.1.3

In this section we give a short proof showing that the learning algorithm of [BCO+15] can be extended in a relatively straightforward manner to functions $f\colon \{0,1\}^d \to [r]$ by increasing the sample-complexity by a factor of r in the exponent.

Proof of Theorem 9.1.3. [BCO+15, Theorem 1.4] proved this result for the case of $r = 2$. By standard arguments, their theorem can be strengthened slightly so that with probability at least $1-\delta$, the learner outputs a hypothesis with error at most ε, and the time and sample complexity is only larger by a multiplicative factor of $\mathrm{poly}(\varepsilon^{-1}, \delta^{-1})$. We will need this slightly stronger statement.

For each $t \in [r]$, let $f_t\colon \{0,1\}^d \to \{0,1\}$ denote the thresholded Boolean function defined as $f_t(x) := \mathbf{1}(f(x) \geq t)$. Observe that for all $x \in \{0,1\}^d$ we have $f(x) = \mathrm{argmax}_t\{f_t(x) = 1\}$. Thus, for each $t \in [r]$, run the learning algorithm of [BCO+15] with error parameters set to $\varepsilon' := \varepsilon/r$ and $\delta = 1/3r$ to obtain a hypothesis h_t. We have $\mathbb{P}[d(h_t, f_t) > \varepsilon/r] < 1/3r$. By a union bound, with probability at least $2/3$, every $t \in [r]$ satisfies $d(h_t, f_t) \leq \varepsilon/r$. Moreover, if this holds then by another union bound we have $\mathbb{P}_x[\exists t \in [r]\colon h_t(x) \neq f(x)] \leq \varepsilon$. Thus, the hypothesis $h(x) := \mathrm{argmax}_t\{h_t(x) = 1\}$ satisfies $d(h, f) \leq \varepsilon$. The number of samples used is $\mathrm{poly}(\varepsilon^{-1}, r) \cdot \exp(O(\min\{\frac{k}{\varepsilon'}\sqrt{d}, d\}))$ and this completes the proof. $\square$

9.4 Preliminaries on k-Monotonicity

We use the notation $[n] := \{0, 1, \ldots, n-1\}$.

Definition 9.4.1. *Given a poset $\mathcal{P} = (\mathcal{X}, \preceq)$ and a function $f\colon \mathcal{X} \to \mathbb{R}$, an m-alternating chain is a sequence of points $x_1 \prec x_2 \prec \cdots \prec x_m$ such that for all $i \in \{1, \ldots, m-1\}$,* *1.* $f(x_{i+1}) - f(x_i) < 0$ *when* i *is odd, and*

2. $f(x_{i+1}) - f(x_i) > 0$ *when* i *is even.*

Definition 9.4.2 (k-monotonicity)**.** *For a poset* $\mathcal{P} = (\mathcal{X}, \preceq)$*, a function* $f \colon \mathcal{X} \to \mathbb{R}$ *is called* k*-monotone if it does not have any* $(k+1)$*-alternating chains.*

Let $\mathcal{M}_{\mathcal{P},k}$ denote the set of all k-monotone functions $f \colon \mathcal{X} \to \mathbb{R}$ over the poset $\mathcal{P} = (\mathcal{X}, \preceq)$. The Hamming distance between two functions $f, g \colon \mathcal{X} \to \mathbb{R}$ is $d(f,g) = |\mathcal{X}|^{-1} \cdot |\{x \in \mathcal{X} \colon f(x) \neq g(x)\}|$. The distance to k-monotonicity of f is denoted by $\varepsilon(f, \mathcal{M}_{\mathcal{P},k}) := \min_{g \in \mathcal{M}_{\mathcal{P},k}} d(f,g)$. The following claim is our main tool for lower bounding the distance to k-monotonicity.

Claim 9.4.3. *Let* $f \colon \mathcal{X} \to \mathbb{R}$ *and* $k' \geq 3k$ *be an integer. Let* $\mathcal{C} \subset \mathcal{X}^{k'}$ *be a collection of disjoint* k'*-alternating chains for* f*. Then*

$$\varepsilon(f, \mathcal{M}_{\mathcal{P},k}) \geq \frac{1}{3|\mathcal{X}|} \cdot \left| \bigcup_{C \in \mathcal{C}} C \right| .$$

Proof. Observe that every k-monotone function $g \in \mathcal{M}_{\mathcal{P},k}$ has the following property: for every $C = (x_1, x_2, \ldots, x_{k'}) \in \mathcal{C}$, the sequence

$$\big(1, g(x_2) - g(x_1), g(x_3) - g(x_2), \ldots, g(x_{k'}) - g(x_{k'-1})\big)$$

changes sign at most $k-1$ times, whereas the sequence

$$\big(1, f(x_2) - f(x_1), f(x_3) - f(x_2), \ldots, f(x_{k'}) - f(x_{k'-1})\big)$$

changes sign exactly $k'-1$ times. We have prepended a 1 so that the first sign change occurs as soon as the function value decreases. Now, changing $f(x_i)$ can only reduce the number of times the sequence changes sign by at most 2 and so $|\{i \colon f(x_i) \neq g(x_i)\}| \geq \frac{k'-k}{2}$. Summing over all chains in $\mathcal{C}$ and normalizing yields

$$d(f,g) \geq \frac{k'-k}{2} \cdot \frac{|\mathcal{C}|}{|\mathcal{X}|} \geq \frac{k'}{3} \cdot \frac{|\mathcal{C}|}{|\mathcal{X}|} \geq \frac{1}{3|\mathcal{X}|} \cdot \left| \bigcup_{C \in \mathcal{C}} C \right|$$

where the second inequality follows from $k \leq k'/3$ and the third inequality is due to the fact that the chains in $\mathcal{C}$ are all disjoint and each of size k'. This completes the proof since this inequality holds for all $g \in \mathcal{M}_{\mathcal{P},k}$. $\square$

We use the notation $\mathcal{M}_{r,k}$ to denote the set of all k-monotone functions $f\colon \{0,1\}^d \to [r]$ over the hypercube whose image has at most r distinct values.

CHAPTER 10

Lower Bounds for Sample-Based k-Monotonicity Testing

In this chapter we prove the following lower bounds on the sample complexity of testing k-monotonicity of functions $f\colon \{0,1\}^d \to [r]$. The main result in this chapter is Theorem 10.0.1 which we prove in Section 10.1. The proof of Theorem 10.0.2 is given in Section 10.2 and is relatively straightforward.

Theorem 10.0.1 (Testing with Two-Sided Error). *There is an absolute constant $c > 0$ such that for all $\varepsilon \leq c$, every sample-based k-monotonicity tester for functions $f\colon \{0,1\}^d \to [r]$ under the uniform distribution has sample complexity*

$$\exp\left(\Omega\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

Theorem 10.0.2 (Testing with One-Sided Error). *For every d, r, k, and $\varepsilon > 0$, sample-based k-monotonicity testing of functions $f\colon \{0,1\}^d \to [r]$ with one-sided error requires $\exp(\Theta(d))$ samples.*

10.1 Two-Sided Error Lower Bound

Our proof of Theorem 10.0.1 follows the standard approach of defining a pair of distributions $\mathcal{D}_{\texttt{yes}}, \mathcal{D}_{\texttt{no}}$ over functions $f\colon \{0,1\}^d \to [r]$ which satisfy the following:

- $\mathcal{D}_{\texttt{yes}}$ is supported over k-monotone functions.

- Functions drawn from $\mathcal{D}_{\texttt{no}}$ are typically $\Omega(\varepsilon)$-far from k-monotone: $\mathbb{P}_{f\sim\mathcal{D}_{\texttt{no}}}[\varepsilon(f,\mathcal{M}_{r,k}) = \Omega(\varepsilon)] = \Omega(1)$.

- The distributions over labeled examples from $\mathcal{D}_{\texttt{yes}}$ and $\mathcal{D}_{\texttt{no}}$ are close in TV-distance.

Our construction uses a generalized version of a family functions known as random Talagrand DNFs, which were used by [BB21] and [CWX17] to prove lower bounds for testing monotonicity of Boolean functions with adaptive and non-adaptive queries.

Let r, k satisfy $rk \le \frac{\varepsilon\sqrt{d}}{4800}$. For convenience, we will assume that $\frac{k(r-1)}{\varepsilon}$ and $\sqrt{d}$ are integers and that $\frac{k(r-1)}{\varepsilon}$ divides $\sqrt{d}$. Let $L_\ell := \left\{x \in \{0,1\}^d \colon |x| = \ell\right\}$ denote the ℓ'th Hamming level of the hypercube. We partition $\bigcup_{\ell\in[0,\varepsilon\sqrt{d})} L_{d/2+\ell}$ into $k(r-1)$ blocks as follows. For each $i \in [k(r-1)]$, define

$$B_i = \bigcup_{\ell = i\cdot\frac{\varepsilon\sqrt{d}}{k(r-1)}}^{(i+1)\cdot\frac{\varepsilon\sqrt{d}}{k(r-1)}-1} L_{\frac{d}{2}+\ell}.$$

The idea of our proof is to define a random DNF within each B_i. The *width* of each DNF will be set to $w := \frac{(r-1)k\sqrt{d}}{2\varepsilon}$ and for each i, the number of terms in the DNF within B_i will be set to $N_i := 2^w \cdot e^{-i} = 2^{\frac{(r-1)k\sqrt{d}}{2\varepsilon}(1-o(1))}$. The DNF defined over B_i will assign function values from $\{i \bmod (r-1), i \bmod (r-1)+1\}$. The terms in each DNF will be chosen randomly from the following distribution. We think of terms as points $t \in \{0,1\}^d$ in the hypercube where another point x *satisfies* t if $t \preceq x$, i.e. $t_i = 1$ implies $x_i = 1$.

Definition 10.1.1 (Term distribution)**.** *A term $t \in \{0,1\}^d$ is sampled from the distribution $\mathcal{D}_{term}$ as follows. Form a (multi)-set $S \subseteq [d]$ by choosing w independent uniform samples from $[d]$. For each $a \in [d]$, let $t_a := \mathbf{1}(a \in S)$.*

10.1.1 The Distributions $\mathcal{D}_{\texttt{yes}}$ and $\mathcal{D}_{\texttt{no}}$

We now define the yes and no distributions over functions $f\colon \{0,1\}^d \to [r]$. For each $i \in [k(r-1)]$, choose terms $t^{i,1}, \ldots, t^{i,N_i}$ i.i.d. from $\mathcal{D}_{\texttt{term}}$ and let $\boldsymbol{t} = \{t^{i,j} : i \in [k(r-1)],\ j \in [N_i]\}$ denote the random set of all terms. Now, for each $i \in [k(r-1)]$ and $j \in [N_i]$, define the set

$$U_{i,j} = \left\{x \in B_i \colon x \succeq t^{i,j} \text{ and } x \not\succeq t^{i,j'} \text{ for all } j' \neq j\right\} \tag{10.1}$$

of all points in the i'th block that satisfy the j'th term *uniquely*. Let $U_i = \bigcup_{j \in [N_i]} U_{i,j}$ denote the set of points in B_i that satisfy a unique term. The following claim is key to our result and motivates our choice of w and N_i. We defer its proof to Section 10.1.2.

Claim 10.1.2. *For any $i \in [k(r-1)]$, $j \in [N_i]$, and $x \in B_i$, we have*

$$\frac{1}{20N_i} \leq \mathbb{P}_{\mathbf{t}}[x \in U_{i,j}] \leq \frac{3}{N_i}.$$

As a corollary, we have $\mathbb{P}_{\mathbf{t}}[x \in U_i] \geq 1/20$.

Functions drawn from $\mathcal{D}_{\texttt{yes}}$ are generated as follows. For each $i \in [k(r-1)]$ choose a uniform random assignment

$$\boldsymbol{\phi}_i \colon [N_i] \to \{i \bmod (r-1), i \bmod (r-1) + 1\} \text{ and let } \boldsymbol{\phi} = (\boldsymbol{\phi}_i \colon i \in [k(r-1)]).$$

For every $x \in B_i$ define

$$f_{\boldsymbol{t},\boldsymbol{\phi}}(x) = \begin{cases} i \bmod (r-1), & \text{if } \forall j \in [N_i],\ x \not\succeq t^{i,j} \\ i \bmod (r-1) + 1, & \text{if } \exists j \neq j' \in [N_i],\ x \succeq t^{i,j}, t^{i,j'} \\ \boldsymbol{\phi}_i(j), & \text{if } x \in U_{i,j}. \end{cases}$$

Functions drawn $\mathcal{D}_{\texttt{no}}$ are generated as follows. For each $i \in [k(r-1)]$ choose a uniform random function

$$\boldsymbol{r}_i \colon U_i \to \{i \bmod (r-1), i \bmod (r-1)+1\} \text{ and let } \boldsymbol{r} = (\boldsymbol{r}_i \colon i \in [k(r-1)]).$$

For each $x \in B_i$ define

$$f_{\boldsymbol{t},\boldsymbol{r}}(x) = \begin{cases} i \bmod (r-1), & \text{if } \forall j \in [N_i],\, x \not\succeq t^{i,j} \\ i \bmod (r-1)+1, & \text{if } \exists j \neq j' \in [N_i],\, x \succeq t^{i,j}, t^{i,j'} \\ \boldsymbol{r}_i(x), & \text{if } x \in U_i. \end{cases}$$

For x not belonging to any B_i: if $|x| < \frac{d}{2}$, then both the yes and no distributions assign value 0 and if $|x| \geq \frac{d}{2} + \varepsilon\sqrt{d}$, then both the yes and no distributions assign value $r-1$.

In summary, a function $f_{\boldsymbol{t},\boldsymbol{\phi}} \sim \mathcal{D}_{\mathsf{yes}}$ assigns the same random value $\boldsymbol{\phi}_i(j) \in \{i \bmod (r-1), i \bmod (r-1)+1\}$ to all points in $U_{i,j}$, which results in a k-monotone function, whereas a function $f_{\boldsymbol{t},\boldsymbol{r}} \sim \mathcal{D}_{\mathsf{no}}$ assigns an i.i.d. uniform random $\{i \bmod (r-1), i \bmod (r-1)+1\}$-value to each point in U_i, resulting in a function that is far from being k-monotone. By construction, to detect any difference between these cases a tester will need to sample at least two points from the same $U_{i,j}$. Theorem 9.1.1 follows immediately from the following three lemmas.

Lemma 10.1.3. *Every function in the support of $\mathcal{D}_{yes}$ is k-monotone.*

Proof. Consider any $f_{\mathbf{t},\boldsymbol{\phi}}(x) \in \mathrm{supp}(\mathcal{D}_{\mathsf{yes}})$. For each $a \in [k]$, consider the union of $r-1$ blocks formed by

$$Y_a := B_{a(r-1)} \cup B_{a(r-1)+1} \cup \cdots \cup B_{(a+1)(r-1)-1}.$$

Recall that if $|x| < d/2$, then $f_{\mathbf{t},\boldsymbol{\phi}}(x) = 0$ and if $|x| \geq d/2 + \varepsilon\sqrt{d}$, then $f_{\mathbf{t},\boldsymbol{\phi}}(x) = r-1$. If $d/2 \leq |x| < d/2 + \varepsilon\sqrt{d}$, then $x \in \bigcup_{a \in [k]} Y_a$. Therefore, it suffices to show that for any pair of comparable points $x \prec y \in Y_a$, we have $f_{\mathbf{t},\boldsymbol{\phi}}(x) \leq f_{\mathbf{t},\boldsymbol{\phi}}(y)$. Firstly, observe that by construction all points $z \in B_{a(r-1)+b}$ have function value $f_{\mathbf{t},\boldsymbol{\phi}}(z) \in \{b, b+1\}$. Since $x \prec y$, if x and y are in different blocks, then $x \in B_{a(r-1)+b}$ and $y \in B_{a(r-1)+b'}$ where $b < b'$ and so the inequality is satisfied. Therefore, we may assume $x, y \in B_{a(r-1)+b}$ are in the same block. Since $x \prec y$, if $t \prec x$ for some term $t \in \text{supp}(\mathcal{D}_{\texttt{term}})$, then $t \prec y$ as well. I.e. the set of terms in $B_{a(r-1)+b}$ satisfied by y is a superset of the set of terms in $B_{a(r-1)+b}$ satisfied by x. By construction, this implies $f_{\mathbf{t},\boldsymbol{\phi}}(x) \leq f_{\mathbf{t},\boldsymbol{\phi}}(y)$. □

Lemma 10.1.4. *For $f_{\boldsymbol{t},\boldsymbol{r}} \sim \mathcal{D}_{no}$, we have $\mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}[\varepsilon(f_{\boldsymbol{t},\boldsymbol{r}}, \mathcal{M}_{r,k}) = \Omega(\varepsilon)] = \Omega(1)$.*

We prove Lemma 10.1.4 in Section 10.1.4.

Lemma 10.1.5. *Given a collection of points $\boldsymbol{x} = (x_1, \ldots, x_s) \in (\{0,1\}^d)^s$ and a function $f\colon \{0,1\}^d \to [r]$, let $(\boldsymbol{x}, f(\boldsymbol{x})) = ((x_1, f(x_1)), \ldots, (x_s, f(x_s))))$ denote the corresponding collection of labelled examples. Let $\mathcal{E}_{yes}$ and $\mathcal{E}_{no}$ denote the distributions over $(\boldsymbol{x}, f(\boldsymbol{x}))$ when $\boldsymbol{x}$ consists of s i.i.d. uniform samples and $f \sim \mathcal{D}_{\boldsymbol{yes}}$ and $f \sim \mathcal{D}_{\boldsymbol{no}}$, respectively. If $s \leq 2^{\frac{(r-1)k\sqrt{d}}{5\varepsilon}}$, then the total variation distance between $\mathcal{E}_{\boldsymbol{yes}}$ and $\mathcal{E}_{\boldsymbol{no}}$ is $o(1)$.*

We prove Lemma 10.1.5 in Section 10.1.3.

10.1.2 Proof of Claim 10.1.2

Proof. Recall $w = \frac{(r-1)k\sqrt{d}}{2\varepsilon}$, $N_i = 2^w \cdot e^{-i}$, the definition of $\mathcal{D}_{\texttt{term}}$ from Definition 10.1.1, and the definition of $U_{i,j}$ from eq. (10.1). Since $x \in B_i$ we have $|x| = \frac{d}{2} + \ell$ where $\frac{i\varepsilon\sqrt{d}}{k(r-1)} \leq \ell < \frac{(i+1)\varepsilon\sqrt{d}}{k(r-1)}$. Note that $\mathbb{P}_{t \sim \mathcal{D}_{\texttt{term}}}[t \preceq x] = (|x|/d)^w$ since $t \preceq x$ iff the non-zero coordinates of t are a subset of the non-zero coordinates of x. Therefore, we have

$$\mathbb{P}_{\mathbf{t}}[x \in U_{i,j}] = \mathbb{P}_{t^{i,j}}[t^{i,j} \preceq x] \cdot \prod_{j' \in [N_i] \setminus \{j\}} \mathbb{P}_{t^{i,j'}}[t^{i,j'} \not\preceq x] = (|x|/d)^w \left(1 - (|x|/d)^w\right)^{N_i - 1}.$$

Note that the first term is upper bounded as

$$(|x|/d)^w \leq \left(\frac{\frac{d}{2} + \frac{(i+1)\cdot\varepsilon\sqrt{d}}{k(r-1)}}{d}\right)^w = \frac{1}{2^w}\left(1 + \frac{2\varepsilon}{k(r-1)\sqrt{d}} \cdot (i+1)\right)^w \leq \frac{e^{i+1+o(1)}}{2^w} \leq \frac{e^{1+o(1)}}{N_i}$$

and this immediately implies the upper bound on $\mathbb{P}_{\mathbf{t}}[x \in U_{i,j}]$. We can also lower bound this quantity by

$$(|x|/d)^w \geq \left(\frac{\frac{d}{2} + \frac{i\cdot\varepsilon\sqrt{d}}{k(r-1)}}{d}\right)^w = \frac{1}{2^w}\left(1 + \frac{2\varepsilon}{k(r-1)\sqrt{d}} \cdot i\right)^w \geq \frac{e^{i-o(1)}}{2^w} \geq \frac{1}{e^{o(1)}N_i}.$$

Now, combining our upper and lower bounds on $(|x|/d)^w$ yields

$$\mathbb{P}_{\mathbf{t}}[x \in U_{i,j}] \geq \frac{1}{e^{o(1)}N_i}\left(1 - \frac{e^{1+o(1)}}{N_i}\right)^{N_i} \geq \frac{1}{e^{o(1)}N_i}e^{-(1+o(1))\cdot e^{1+o(1)}} \geq \frac{1}{e^{e+1}N_i} \geq \frac{1}{20N_i}.$$

□

10.1.3 $\mathcal{D}_{\texttt{yes}}$ and $\mathcal{D}_{\texttt{no}}$ are Hard to Distinguish

Proof. Recall the definition of the set $U_{i,j}$ in eq. (10.1). For $a \neq b \in [s]$, let E_{ab} denote the event that x_a and x_b belong to the same $U_{i,j}$ for some $i \in [k(r-1)]$ and $j \in [N_i]$. Observe that conditioned on $\overline{\vee_{a,b} E_{ab}}$, the distributions $\mathcal{E}_{yes}$ and $\mathcal{E}_{no}$ are identical. Let $x, y \in \{0,1\}^d$ denote two i.i.d. uniform samples. We have

$$\begin{aligned}\mathbb{P}[E_{ab}] &= \mathbb{P}_{x,y,\mathbf{t}}\left[\bigvee_{i,j}(x \in U_{i,j} \wedge y \in U_{i,j})\right] \\ &= \sum_{i,j}\mathbb{P}_{x,y,\mathbf{t}}\left[x \in U_{i,j} \wedge y \in U_{i,j}\right] = \sum_{i,j}\mathbb{P}_{x,\mathbf{t}}[x \in U_{i,j}]^2 \qquad (10.2)\end{aligned}$$

where the first step holds since the $U_{i,j}$'s are disjoint and the second step holds by independence of x and y. Now, for a fixed $i \in [k(r-1)]$ and $j \in [N_i]$ we have the following: by Claim 10.1.2, for $x \in B_i$ we have $\mathbb{P}_{\mathbf{t}}[x \in U_{i,j}] \leq \frac{3}{N_i}$ and for $x \notin B_i$ we have $\mathbb{P}_{\mathbf{t}}[x \in U_{i,j}] = 0$.

Therefore $\mathbb{P}_{x,\mathbf{t}}[x \in U_{i,j}] \leq \frac{3}{N_i}$. Therefore, the RHS of eq. (10.2) is bounded as

$$\sum_{i,j} \mathbb{P}_{x,\mathbf{t}}[x \in U_{i,j}]^2 = \sum_i N_i \cdot \mathbb{P}_{x,\mathbf{t}}[x \in U_{i,j}]^2 \leq \sum_i \frac{9}{N_i} \leq rk \cdot \frac{9}{N_{r-2}}$$

since the N_i's are decreasing with respect to i. Therefore,

$$d_{TV}(\mathcal{E}_{\mathtt{yes}}, \mathcal{E}_{\mathtt{no}}) \leq \mathbb{P}_{\boldsymbol{x},\boldsymbol{t}}\left[\bigvee_{a,b \in [s]} E_{ab}\right] \leq s^2 \cdot rk \cdot \frac{9}{N_{r-2}} = o(1)$$

since $N_{r-2} = 2^{\frac{(r-1)k\sqrt{d}}{2\varepsilon}(1-o(1))} = \omega(s^2 \cdot rk)$. $\square$

10.1.4 Functions Drawn from $\mathcal{D}_{\mathtt{no}}$ are Far from k-Monotone

Proof. We will use Claim 9.4.3, restated below for the special case of r-valued functions over the hypercube. Recall that $\mathcal{M}_{r,k}$ is the set of k-monotone functions $f\colon \{0,1\}^d \to [r]$.

Claim 10.1.6. *Let $f\colon \{0,1\}^d \to [r]$ and $k' \geq 3k$ be an integer. Let $\mathcal{C} \subset (\{0,1\}^d)^{k'}$ be a collection of disjoint k'-alternating chains for f. Then*

$$\varepsilon(f, \mathcal{M}_{r,k}) \geq \frac{1}{3 \cdot 2^d} \cdot \left|\bigcup_{C \in \mathcal{C}} C\right|.$$

From the above claim, we can lower bound the distance to k-monotonicity of f by showing that it contains a collection of disjoint k'-alternating chains where $k' \geq 3k$ whose union makes up an $\Omega(\varepsilon)$-fraction of the hypercube.

Recall $U_i = U_{i,1} \cup \cdots \cup U_{i,N_i} \subseteq B_i$ and note that $f_{\mathbf{t},\boldsymbol{r}} \sim \mathcal{D}_{\mathtt{no}}$ takes values only from $\{i \bmod (r-1), i \bmod (r-1)+1\}$ in B_i. In particular, for $a \in \{0,1,\ldots,r-2\}$, let

$$X_a = B_a \cup B_{(r-1)+a} \cup B_{2(r-1)+a} \cup \cdots \cup B_{(k-1)(r-1)+a} = \bigcup_{i \in [k]} B_{i(r-1)+a} \tag{10.3}$$

and note that all points $x \in X_a$ are assigned value $f_{\mathbf{t},\boldsymbol{r}}(x) \in \{a, a+1\}$. Moreover, this value is chosen uniformly at random when $x \in \bigcup_{i \in [k]} U_{i(r-1)+a}$, which occurs with probability $\geq 1/20$ by Claim 10.1.2. Let $k'' := \frac{\varepsilon\sqrt{d}}{r-1}$ and recall that we are assuming $rk \leq \frac{\varepsilon\sqrt{d}}{4800}$ and so $k'' \geq 4800k$. We first show there exists a large collection $\mathcal{C}_a$ of length-k'' disjoint chains in X_a for all $a \in \{0, 1, \ldots, r-2\}$.

Claim 10.1.7. *For every $a \in \{0, 1, \ldots, r-2\}$, there exists a collection of vertex disjoint chains $\mathcal{C}_a \subset (X_a)^{k''}$ in X_a of length k'' of size $|\mathcal{C}_a| \geq \Omega(\frac{2^d}{\sqrt{d}})$.*

Proof. We start by showing that there is a large matching in the transitive closure of the hypercube from $L_{\frac{d}{2}}$ to $L_{\frac{d}{2}+\varepsilon\sqrt{d}-1}$. Consider the bipartite graph (U, V, E) where $U := L_{\frac{d}{2}}$, $V := L_{\frac{d}{2}+\varepsilon\sqrt{d}-1}$, and $E := \{(x, y) \in U \times V : x \prec y\}$. Observe that vertices in U have degree exactly $\Delta := \binom{\frac{d}{2}}{\varepsilon\sqrt{d}-1}$ while vertices in V have degree exactly $\binom{\frac{d}{2}+\varepsilon\sqrt{d}-1}{\varepsilon\sqrt{d}-1} \geq \Delta$. Note also that $|V| = \binom{d}{\frac{d}{2}+\varepsilon\sqrt{d}-1} \geq \Omega(\frac{2^d}{\sqrt{d}})$ by Stirling's approximation. We now use the following claim from [BBH23].

Claim 10.1.8 ([BBH23])**.** *Let (U, V, E) be a bipartite graph and $\Delta > 0$ be such that (a) each vertex $x \in U$ has degree exactly Δ and (b) each vertex $y \in V$ has degree at least Δ. Then there exists a matching $M \subseteq E$ in (U, V, E) of size $|M| \geq \frac{1}{2}|V|$.*

By the above claim and the previous observations, there exist subsets $S \subseteq L_{\frac{d}{2}}$ and $T \subseteq L_{\frac{d}{2}+\varepsilon\sqrt{d}-1}$ of size $|S| = |T| = \Omega(\frac{2^d}{\sqrt{d}})$ and a bijection $\phi \colon S \to T$ satisfying $x \prec \phi(x)$ for all $x \in S$. We now use the following routing theorem due to Lehman and Ron to obtain a collection of disjoint chains from S to T.

Theorem 10.1.9 (Lehman-Ron, [LR01])**.** *Let $a < b$ and $S \subseteq L_a$, $T \subseteq L_b$ where $m := |S| = |T|$. Moreover, suppose there is a bijection $\phi \colon S \to T$ satisfying $x \prec \phi(x)$ for all $x \in S$. Then there exist m vertex disjoint paths from S to T in the hypercube.*

Now, invoking the above theorem on our bijection $\phi \colon S \to T$ yields a collection P of $|P| \geq \Omega(\frac{2^d}{\sqrt{d}})$ vertex disjoint paths from $L_{\frac{d}{2}}$ to $L_{\frac{d}{2}+\varepsilon\sqrt{d}-1}$. For each $a \in \{0, 1, \ldots, r-2\}$, let $\mathcal{C}_a$ denote the collection of chains formed by taking a path in P and including only the vertices from X_a (recall eq. (10.3)). Note that the resulting chains in $\mathcal{C}_a$ are of length $k'' = \frac{\varepsilon\sqrt{d}}{r-1}$. This completes the proof of Claim 10.1.7. $\square$

From Claim 10.1.7, we have $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{r-2}$ where each $\mathcal{C}_a \subset (X_a)^{k''}$ is a collection of vertex disjoint chains of length $k'' \geq 4800k$ of size $|\mathcal{C}_a| \geq \Omega(\frac{2^d}{\sqrt{d}})$. Fix a chain $C = (x_1, x_2, \ldots, x_{k''}) \in \mathcal{C}_a$. Let $A(C)$ be the random variable which denotes the max-length alternating sub-chain (recall Definition 9.4.1) of C over a random $f_{\mathbf{t},\boldsymbol{r}} \sim \mathcal{D}_{\mathtt{no}}$. Fix x_j in the chain and suppose $x_j \in B_i \subseteq X_a$. By Claim 10.1.2, $\mathbb{P}_{\mathbf{t}}[x_j \in U_i] \geq 1/20$. Moreover, conditioned on $x_j \in U_i$, $f_{\mathbf{t},\boldsymbol{r}}(x_j)$ is chosen from $\{a, a+1\}$ uniformly at random. Thus, any step of the sequence

$$(1, f_{\mathbf{t},\boldsymbol{r}}(x_2) - f_{\mathbf{t},\boldsymbol{r}}(x_1), f_{\mathbf{t},\boldsymbol{r}}(x_3) - f_{\mathbf{t},\boldsymbol{r}}(x_2), \ldots, f_{\mathbf{t},\boldsymbol{r}}(x_{k''}) - f_{\mathbf{t},\boldsymbol{r}}(x_{k''-1}))$$

is non-zero *and* differs in sign from the previous non-zero step with probability at least $1/40$ and so $\mathbb{E}[A(C)] \geq k''/40$. I.e., $0 \leq \mathbb{E}[k'' - A(C)] < k''(1 - \frac{1}{40})$. Thus, using Markov's inequality we have

$$\mathbb{P}\left[A(C) < \frac{k''}{1600}\right] = \mathbb{P}\left[k'' - A(C) > k''\left(1 - \frac{1}{40}\right)\left(1 + \frac{1}{40}\right)\right] \leq \frac{1}{(1 + \frac{1}{40})} = 1 - \frac{1}{41}. \tag{10.4}$$

Now, let $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_{r-2}$ and let $Z := |\{C \in \mathcal{C} \colon A(C) \geq \frac{k''}{1600}\}|$. By eq. (10.4) we have $\mathbb{E}[Z] \geq |\mathcal{C}|/41$ and $0 \leq \mathbb{E}[|\mathcal{C}| - Z] \leq |\mathcal{C}|(1 - \frac{1}{41})$. Again using Markov's inequality, we have

$$\mathbb{P}\left[Z < \frac{|\mathcal{C}|}{1681}\right] = \mathbb{P}\left[|\mathcal{C}| - Z > |\mathcal{C}|\left(1 - \frac{1}{41}\right)\left(1 + \frac{1}{41}\right)\right] \leq \frac{1}{(1 + \frac{1}{41})} = 1 - \frac{1}{42}. \tag{10.5}$$

Now, for $C \in \mathcal{C}$ such that $A(C) \geq k''/1600$, let C' be any $(k''/1600)$-alternating sub-chain of C. Let $\mathcal{C}' = \{C' \colon C \in \mathcal{C}$ such that $A(C) \geq k''/1600\}$ which is a collection of disjoint $(k''/1600)$-alternating chains for $f_{\mathbf{t},\boldsymbol{r}}$. Now, recall that $k'' \geq 4800k$ and so $k''/1600 \geq 3k$.

Thus, if $Z \geq |\mathcal{C}|/1681$, then $|\mathcal{C}'| \geq |\mathcal{C}|/1681$ and so by Claim 10.1.6 we have

$$\varepsilon(f_{\mathbf{t},\boldsymbol{r}}, \mathcal{M}_{r,k}) \geq \frac{1}{3 \cdot 2^d} \left| \bigcup_{C' \in \mathcal{C}'} C' \right| \geq \frac{1}{3 \cdot 2^d} \cdot |\mathcal{C}'| \cdot \frac{k''}{1600} \geq \frac{k'' \cdot |\mathcal{C}|}{10,000,000 \cdot 2^d} \tag{10.6}$$

By Claim 10.1.7 we have $|\mathcal{C}| \geq (r-1) \cdot \Omega(\frac{2^d}{\sqrt{d}})$ and recall that $k'' = \frac{\varepsilon\sqrt{d}}{r-1}$. Thus, the RHS of eq. (10.6) is $\Omega(\varepsilon)$. In conclusion,

$$\mathbb{P}_{\mathbf{t},\boldsymbol{r}}\left[\varepsilon(f_{\mathbf{t},\boldsymbol{r}}, \mathcal{M}_{r,k}) \geq \Omega(\varepsilon)\right] \geq \mathbb{P}\left[Z \geq \frac{|\mathcal{C}|}{1681}\right] \geq \frac{1}{42}$$

by eq. (10.5) and this completes the proof of Lemma 10.1.4. □

10.2 One-Sided Error Lower Bound

In this section we prove Theorem 10.0.2, our upper and lower bound on sample-based testing with one-sided error over the hypercube.

Proof of Theorem 10.0.2. By a coupon-collecting argument, there is an $O(d \cdot 2^d)$ sample upper bound for *exactly learning* any function over $\{0,1\}^d$ under the uniform distribution and therefore the upper bound is trivial.

It suffices to prove the lower bound for the case of $r = 2$ and $k = 1$, i.e. for testing monotonicity of Boolean functions. We will need the following fact.

Fact 10.2.1. *Let $A \subset \{0,1\}^d$ be any anti-chain and let $\ell \colon A \to \{0,1\}$ be any labelling of A. Then there exists a monotone function $f \colon \{0,1\}^d \to \{0,1\}$ such that $f(x) = \ell(x)$ for all $x \in A$. I.e. A* shatters *the class of monotone functions.*

Now, let T be any monotonicity tester with one-sided error and let $S \subseteq \{0,1\}^d$ denote a set of s i.i.d. uniform samples. Since T has one-sided error, if the input function is monotone, then T must accept. In other words, for T to reject it must be sure without a doubt that the input function is not monotone. By Fact 10.2.1 for T to be sure the input function is not monotone, it must be that S is *not* an anti-chain. Let $f\colon \{0,1\}^d \to \{0,1\}$ be any function which is ε-far from monotone. Since T is a valid tester, it rejects f with probability at least 2/3. By the above argument we have

$$2/3 \leq \mathbb{P}_S[T \text{ rejects } f] \leq \mathbb{P}_S[S \text{ is not an anti-chain}] \leq s^2 \cdot \mathbb{P}_{x,y\sim\{0,1\}^d}[x \preceq y] \tag{10.7}$$

where the last inequality is by a union bound over all pairs of samples. We then have

$$\mathbb{P}_{x,y\sim\{0,1\}^d}[x \preceq y] = \mathbb{P}_{x,y\sim\{0,1\}^d}[x_i \leq y_i,\, \forall i \in [d]] = \prod_{i=1}^{d} \mathbb{P}_{x_i,y_i\sim\{0,1\}}[x_i \leq y_i] = (3/4)^d. \tag{10.8}$$

Thus, combining eq. (10.7) and eq. (10.8) yields $s \geq \sqrt{\frac{2}{3}(\frac{4}{3})^d} = \exp(\Omega(d))$. $\square$

CHAPTER 11

Learning Upper Bound for k-Monotone Functions in Product Spaces

In this chapter we prove the following upper bound for learning measurable k-monotone functions in $\mathbb{R}^d$.

Theorem 11.0.1. *Given an arbitrary product measure μ, there is a learning algorithm under μ for measurable k-monotone functions $f\colon \mathbb{R}^d \to [r]$ with time and sample complexity*

$$\exp\left(\widetilde{O}\left(\min\left\{\frac{rk}{\varepsilon}\sqrt{d}, d\right\}\right)\right).$$

The $\widetilde{O}(\cdot)$ hides polylogarithmic dependencies on d, r, k, and $1/\varepsilon$.

We restate the theorem below without any hidden logarithmic factors and for the case of $r = 2$. The theorem for general $r \geq 2$ can then be obtained by replacing ε with ε/r and δ by $1/3r$ following the same approach we used to prove Theorem 9.1.3 in Section 9.3.

Theorem 11.0.2. *Given an arbitrary product measure μ, there is a learning algorithm under μ which learns any measurable k-monotone function $f\colon \mathbb{R}^d \to \{\pm 1\}$ to error ε with probability $1-\delta$ with time and sample complexity*

$$\ln\left(\frac{1}{\delta}\right) \cdot \min\left\{(d\log(dk/\varepsilon))^{O\left(\frac{k}{\varepsilon}\sqrt{d\log(dk/\varepsilon)}\right)}, \left(\frac{dk}{\varepsilon}\right)^{O(d)}\right\} \tag{11.1}$$

Our proof uses downsampling to reduce our learning problem over $\mathbb{R}^d$ to learning over a hypergrid, $[N]^d$, under the uniform distribution with mild label noise. In Section 11.1 we synthesize the results from [HY22] which we borrow for our proof. In Section 11.2 we give two learning results for hypergrids whose time complexities correspond to the two arguments inside the min expression in eq. (11.1). In Section 11.3 we describe the learning algorithm and prove its correctness.

Throughout this section, let $\mu = \prod_{i=1}^{d} \mu_i$ be any product measure over $\mathbb{R}^d$ and let N be a power of two satisfying $8kd/\varepsilon \leq N \leq 16kd/\varepsilon$.

11.1 Reduction to Hypergrids via Downsampling

The idea of downsampling is to construct a grid-partition of $\mathbb{R}^d$ into N^d blocks such that (a) the measure of each block under μ is roughly N^{-d}, and (b) the function f we're trying to learn is constant on most of the blocks. Roughly speaking, this allows us to learn f under μ by learning a proxy for f over $[N]^d$ under the uniform distribution. The value of N needed to achieve this depends on what [HY22] call the "block boundary size" of the function. Formally, the downsampling procedure constructs query access to maps $\mathsf{block}\colon \mathbb{R}^d \to [N]^d$ and $\mathsf{blockpoint}\colon [N]^d \to \mathbb{R}^d$ which have various good properties which we will spell out in the rest of this section. One should think of block as mapping each point $x \in \mathbb{R}^d$ to the block of the grid-partition that x belongs to and $\mathsf{blockpoint}$ as mapping each block to some specific point contained in the block. See [HY22, Def 2.1] for a formal definition. Given these maps and a function $f\colon \mathbb{R}^d \to \{\pm 1\}$ we define the function $f^{\mathsf{block}}\colon [N]^d \to \{\pm 1\}$ as $f^{\mathsf{block}}(z) = f(\mathsf{blockpoint}(z))$. We let $\mathsf{block}(\mu)$ denote the distribution over $[N]^d$ induced by sampling $x \sim \mu$ and then taking $\mathsf{block}(x)$.

Proposition 11.1.1 (Downsampling, [HY22])**.** *Let $f\colon \mathbb{R}^d \to \{0,1\}$ be a k-monotone function and $N, Q \in \mathbb{Z}^+$. Using*

$$m := O\left(\frac{NQ^2d^2}{\min(\delta,\varepsilon)^2} \ln\left(\frac{Nd}{\delta}\right)\right)$$

samples from $\mu = \mu_1 \times \cdots \times \mu_d$, there is a downsampling procedure that constructs query *access to maps* $\mathsf{block}\colon \mathbb{R}^d \to [N]^d$ *and* $\mathsf{blockpoint}\colon [N]^d \to \mathbb{R}^d$ *such that with probability at least* $1-\delta$ *over the random samples, the following two conditions are satisfied:*

1. $\left\|\mathsf{block}(\mu) - \mathsf{unif}([N]^d)\right\|_{TV} \le \frac{\delta}{Q}$.
2. $\mathbb{P}_{x\sim\mu}\left[f(x) \neq f^{\mathsf{block}}(\mathsf{block}(x))\right] \le \varepsilon$.

The total running time and number of samples is $O(m)$.

Proof. [HY22, Prop. 2.5] shows that there is a randomized procedure using m samples from μ and $O(m)$ time which constructs the maps block and $\mathsf{blockpoint}$ such that with probability 1, we get

$$\mathbb{P}_{x\sim\mu}\left[f(x) \neq f^{\mathsf{block}}(\mathsf{block}(x))\right] \le N^{-d}\cdot \mathsf{bbs}(f,N) + \left\|\mathsf{block}(\mu) - \mathsf{unif}([N]^d)\right\|_{\mathrm{TV}} \tag{11.2}$$

where $\mathsf{bbs}(f,N)$ is the N-block boundary size of f [HY22, Def. 2.4], which is at most kdN^{d-1} when f is k-monotone [HY22, Lemma 7.1]. Thus, the first of the two quantities in the RHS is at most kd/N which is at most $\varepsilon/8$ using our definition of N. Then, [HY22, Lemma 2.7] states that

$$\mathbb{P}\left[\left\|\mathsf{block}(\mu) - \mathsf{unif}([N]^d)\right\|_{\mathrm{TV}} > \beta\right] \le 4Nd\cdot \exp\left(-\frac{\beta^2 m}{18Nd^2}\right) \tag{11.3}$$

and so invoking this lemma with $\beta := \min(\delta/4Q, \varepsilon/8)$ and $m := \frac{18Nd^2}{\beta^2}\ln\left(\frac{16Nd}{\delta}\right)$ completes the proof. $\square$

11.2 Learning over Hypergrids

For a function $f\colon \mathcal{X} \to \{\pm 1\}$ and a measure μ over $\mathcal{X}$, recall that the *example oracle* for f under μ, denoted by $EX(f,\mu)$, when queried, generates an example, $(x, f(x))$, where x is sampled from μ. Given a *noise parameter* η, the *noisy example oracle* $EX^\eta(f,\mu)$, when queried, samples x from μ, returns the true example $(x, f(x))$ with probability $1-\eta$, and returns the corrupted example $(x, -f(x))$ with probability η. This is referred to as *random classification noise* (RCN).

We prove the following two upper bounds for learning over hypergrids under RCN. The bound in Lemma 11.2.1 is relatively straightforward to prove using coupon collector arguments plus some additional work to handle the label noise. We give a proof in Section 11.5.

Lemma 11.2.1 (Coupon Collecting Learner)**.** *Let $\varepsilon, \delta \in (0,1)$, $\eta \in (0, 1/2)$, and $N \in \mathbb{Z}^+$. There is an algorithm which, given any k-monotone function $f\colon [N]^d \to \{\pm 1\}$, uses at most*

$$\widetilde{O}\left(\frac{1}{(1-2\eta)^2}\left(\log\frac{1}{\varepsilon} + \log\frac{1}{\delta}\right)\right) \cdot N^{O(d)}$$

examples from $EX^\eta(f, \mathsf{unif}([N]^d))$ and returns $h\colon [N]^d \to \{\pm 1\}$, satisfying $\mathbb{P}_h[d(f,h) \le \varepsilon] \ge 1 - \delta$.

Lemma 11.2.2 (Hypercube Mapping Learner)**.** *Let $\varepsilon, \delta \in (0,1)$, $\eta \in (0, 1/2)$, and $N \in \mathbb{Z}^+$ be a power of two. There is an algorithm which, given any k-monotone function $f\colon [N]^d \to \{\pm 1\}$, uses at most*

$$O\left(\frac{1}{\varepsilon^2(1-2\eta)^2} + \log\frac{1}{\delta}\right)(d \log N)^{O\left(\frac{k}{\varepsilon}\sqrt{d \log N}\right)}$$

examples from $EX^\eta(f, \mathsf{unif}([N]^d))$ and returns $h\colon [N]^d \to \{\pm 1\}$, satisfying $\mathbb{P}_h[d(f,h) \le \varepsilon] \ge 1 - \delta$.

Proof. Let $b\colon [N] \to \{\pm 1\}^{\log N}$ denote the bijection which maps each element of $[N]$ to its bit representation. Let $\boldsymbol{b}\colon [N]^d \to \{\pm 1\}^{d \log N}$ be defined as $\boldsymbol{b}(x) = (b(x_1), \dots, b(x_d))$. Given $f\colon [N]^d \to \{\pm 1\}$ define the function $f^{\mathsf{cube}}\colon \{\pm 1\}^{d \log N} \to \{\pm 1\}$ as $f^{\mathsf{cube}}(z) = f(\boldsymbol{b}^{-1}(z))$.

Observation 11.2.3. *If f is k-monotone over $[N]^d$, then f^{cube} is k-monotone over $\{\pm 1\}^{d \log N}$.*

Proof. Observe that if $\boldsymbol{b}(x) \prec \boldsymbol{b}(y)$ in $\{\pm 1\}^{d \log N}$, then $x \prec y$ in $[N]^d$. Thus, if $\boldsymbol{b}(x_1) \prec \cdots \prec \boldsymbol{b}(x_m)$ is an m-alternating chain for f^{cube}, then $x_1 \prec \cdots \prec x_m$ is an m-alternating chain for f. Therefore, if f^{cube} is not k-monotone, then neither is f. $\square$

Now, given Observation 11.2.3 and the bijection $\boldsymbol{b}\colon [N]^d \to \{\pm 1\}^{d\log N}$, it suffices to provide a learning algorithm for f^{cube}. This is achieved using the Low-Degree Algorithm introduced by [LMN93] which was shown by [Kea98] to be robust to classification noise. Formally, we use the following theorem, which we prove in Section 11.4 for the sake of completeness.

Theorem 11.2.4 (Low-Degree Algorithm with Classification Noise). *Let $\varepsilon, \delta \in (0,1)$ and $\eta \in (0, 1/2)$. Suppose $\mathcal{C}$ is a concept class of Boolean functions over $\{\pm 1\}^d$ such that for some fixed positive integer τ, all $f \in \mathcal{C}$ satisfy $\sum_{S\subseteq[d]\colon |S|>\tau} \widehat{f}(S)^2 \leq \varepsilon/2$. Then there is an algorithm $\mathcal{A}$ which, on any input $f \in \mathcal{C}$, uses at most*

$$O\left(\left(\frac{1}{\varepsilon^2(1-2\eta)^2} + \log\frac{1}{\delta}\right)\cdot d^{\tau}\right)$$

examples from $EX^{\eta}(f, \mathsf{unif}(\{\pm 1\}^d))$ and returns a hypothesis $h\colon \{\pm 1\}^d \to \{\pm 1\}$ where $\mathbb{P}_h[d(f,h) \leq \varepsilon] \geq 1-\delta$.

We use the following Fourier concentration lemma due to [BCO+15] for k-monotone Boolean functions.

Lemma 11.2.5 ([BCO+15]). *If $f\colon \{\pm 1\}^d \to \{\pm 1\}$ is k-monotone, then $\sum_{S\colon |S|>\frac{k\sqrt{d}}{\varepsilon}} \widehat{f}(S)^2 \leq \varepsilon$.*

By Lemma 11.2.5, we can invoke Theorem 11.2.4 with $\tau = \frac{k\sqrt{d\log N}}{\varepsilon}$, concluding the proof of Lemma 11.2.2. □

11.3 Putting it Together: Proof of Theorem 11.0.2

Proof. We now have all the tools to define the algorithm and prove its correctness.

Recall that given maps $\mathsf{block}\colon \mathbb{R}^d \to [N]^d$, $\mathsf{blockpoint}\colon [N]^d \to \mathbb{R}^d$, and a function $f\colon \mathbb{R}^d \to \{\pm 1\}$ we define the function $f^{\mathsf{block}}\colon [N]^d \to \{\pm 1\}$ as $f^{\mathsf{block}}(z) = f(\mathsf{blockpoint}(z))$. Recall that

Algorithm 5 Learning algorithm for k-monotone functions under product measure μ

Input: $\varepsilon, \delta \in (0,1)$ and access to examples from $EX(f,\mu)$ where $f\colon \mathbb{R}^d \to \{\pm 1\}$ is k-monotone.

1. Let N be a power of 2 such that $\frac{8kd}{\varepsilon} \le N \le \frac{16kd}{\varepsilon}$. Let $\mathcal{A}$ denote the learning algorithm for k-monotone functions $g\colon [N]^d \to \{\pm 1\}$ which has the smaller sample complexity of the two algorithms from Lemma 11.2.1 and Lemma 11.2.2. Let Q be the sample complexity of $\mathcal{A}$.
2. Run the downsampling procedure of Proposition 11.1.1 to obtain the maps block, $\mathsf{blockpoint}$, and access to the corresponding function $f^{\mathsf{block}}\colon [N]^d \to \{\pm 1\}$.
3. Obtain a set of Q examples $S \in (\mathbb{R}^d \times \{\pm 1\})^Q$ from $(EX(f,\mu))^Q$.
4. Let $S^{\mathsf{block}} = \{(\mathsf{block}(x), f(x)) \colon (x, f(x)) \in S\} \in ([N]^d \times \{\pm 1\})^Q$.
5. Run $\mathcal{A}$ using the sample S^{block}, which returns a hypothesis $h^{\mathsf{block}}\colon [N]^d \to \{\pm 1\}$ for f^{block}.

Return the hypothesis $h\colon \mathbb{R}^d \to \{\pm 1\}$ for $f\colon \mathbb{R}^d \to \{\pm 1\}$ defined as $h(x) = h^{\mathsf{block}}(\mathsf{block}(x))$.

$\mathsf{block}(\mu)$ is the distribution over $\mathsf{block}(x) \in [N]^d$ when $x \sim \mu$. By Proposition 11.1.1, step (2) of Alg. 11.3 results in the following items being satisfied with probability at least $1-\delta$.

1. $\left\|\mathsf{block}(\mu) - \mathsf{unif}([N]^d)\right\|_{\mathrm{TV}} \le \frac{\delta}{Q}$.
2. $\mathbb{P}_{x\sim\mu}\left[f(x) \neq f^{\mathsf{block}}(\mathsf{block}(x))\right] \le \varepsilon$.

Firstly, by item (2), an example $(\mathsf{block}(x), f(x))$ where $x \sim \mu$, is equivalent to an example $(z,b) \sim EX^{\eta}(f^{\mathsf{block}}, \mathsf{block}(\mu))$ for some $\eta \le \varepsilon$. I.e. the set $S^{\mathsf{block}} \in ([N]^d \times \{\pm 1\})^Q$ from step (4) of Alg. 11.3 is distributed according to $(EX^{\eta}(f^{\mathsf{block}}, \mathsf{block}(\mu)))^Q$. Now, as stated, Lemma 11.2.1 and Lemma 11.2.2 only hold when $\mathcal{A}$ is given a sample from $(EX^{\eta}(f^{\mathsf{block}}, \mathsf{unif}([N]^d)))^Q$. However, the following claim shows that since $\mathsf{block}(\mu)$ and $\mathsf{unif}([N]^d))$ are sufficiently close (item (1) above), the guarantees on $\mathcal{A}$ from Lemma 11.2.1 and Lemma 11.2.2 also hold when $\mathcal{A}$ is given a sample from $(EX^{\eta}(f^{\mathsf{block}}, \mathsf{block}(\mu)))^Q$.

Claim 11.3.1. *Let $\mathcal{C}\colon \mathcal{X} \to \{\pm 1\}$ be a concept class and let $\mathcal{A}$ be an algorithm which given any $f \in \mathcal{C}$, $\varepsilon, \delta \in (0,1)$, and $\eta \in [0, 1/2)$ uses a sample from $(EX^{\eta}(f, \mathsf{unif}([N]^d)))^Q$ and produces h satisfying $\mathbb{P}_{x\sim\mathsf{unif}([N]^d)}[h(x) \neq f(x)] \le \varepsilon$ with probability at least $1-\delta$. If $\mathcal{D}$ is a distribution over $[N]^d$ with $\left\|\mathcal{D} - \mathsf{unif}([N]^d)\right\|_{TV} \le \gamma$, then given a sample from $(EX^{\eta}(f,\mathcal{D}))^Q$, $\mathcal{A}$ produces h satisfying $\mathbb{P}_{x\sim\mathcal{D}}[h(x) \neq f(x)] \le \varepsilon + \gamma$ with probability at least $1 - (\delta + \gamma Q)$.*

Using Claim 11.3.1 and item (1) above, if step (2) of Alg. 11.3 succeeds, then with probability at least $1-2\delta$, step (5) produces h^{block} such that $\mathbb{P}_{z\sim\mathsf{block}(\mu)}[h^{\mathsf{block}}(z) \neq f^{\mathsf{block}}(z)] \leq 2\varepsilon$. By the triangle inequality and using our definition of h in the return statement of Alg. 11.3, we have

$$\begin{aligned}\mathbb{P}_{x\sim\mu}[h(x) \neq f(x)] &\leq \mathbb{P}_{x\sim\mu}[f(x) \neq f^{\mathsf{block}}(\mathsf{block}(x))] + \mathbb{P}_{x\sim\mu}[f^{\mathsf{block}}(\mathsf{block}(x)) \neq h^{\mathsf{block}}(\mathsf{block}(x))] \\ &= \mathbb{P}_{x\sim\mu}[f(x) \neq f^{\mathsf{block}}(\mathsf{block}(x))] + \mathbb{P}_{z\sim\mathsf{block}(\mu)}[f^{\mathsf{block}}(z) \neq h^{\mathsf{block}}(z)]. \end{aligned} \tag{11.4}$$

The first term in the RHS is at most ε by item (2) above and the second term is at most 2ε as we argued in the previous paragraph. Finally, adding up the failure probabilities of steps (2) and (5), we conclude that Alg. 11.3 produces h satisfying $\mathbb{P}_{x\sim\mu}[h(x) \neq f(x)] \leq 3\varepsilon$ with probability at least $1-3\delta$. □

11.3.1 Proof of Claim 11.3.1

Proof. It is a well-known fact that for two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$, the TV-distance between the corresponding product distributions satisfies $\left\|\mathcal{D}_1^Q - \mathcal{D}_2^Q\right\|_{TV} \leq Q\left\|\mathcal{D}_1 - \mathcal{D}_2\right\|_{TV}$ and thus we have

$$\left\|\mathcal{D}^Q - \mathsf{unif}([N]^d)^Q\right\|_{TV} \leq \gamma Q$$

Given a set of Q examples $S \in ([N]^d \times \{\pm 1\})^Q$, let $E(S)$ denote the event that the algorithm $\mathcal{A}$ fails to produce a hypothesis with error at most ε, after sampling S. First, note the distribution over labels for the distributions are the same, and therefore

$$\mathbb{P}_{S\sim(EX^{\eta}(f,\mathcal{D}))^Q}[E(S)] - \mathbb{P}_{S\sim(EX^{\eta}(f,\mathsf{unif}([N]^d)))^Q}[E(S)] = \mathbb{P}_{S\sim\mathcal{D}^Q}[E(S)] - \mathbb{P}_{S\sim\mathsf{unif}([N]^d)^Q}[E(S)]. \tag{11.5}$$

Using the definition of TV-distance we have

$$\mathbb{P}_{S\sim\mathcal{D}^Q}[E(S)] - \mathbb{P}_{S\sim\mathsf{unif}([N]^d)^Q}[E(S)] \leq \left\|\mathcal{D}^Q - \mathsf{unif}([N]^d)^Q\right\|_{TV} \leq \gamma Q \tag{11.6}$$

and therefore

$$\mathbb{P}_{S\sim(EX^{\eta}(f,\mathcal{D}))^Q}[E(S)] \le \mathbb{P}_{S\sim(EX^{\eta}(f,\mathsf{unif}([N]^d)))^Q}[E(S)] + \gamma Q \le \delta + \gamma Q \qquad (11.7)$$

where we used $\mathbb{P}_{S\sim(EX^{\eta}(f,\mathsf{unif}([N]^d)))^Q}[E(S)] \le \delta$ by the assumption in the statement of the claim. Now, conditioned on $\neg E(S)$, we have that $\mathcal{A}$ produces h satisfying $\mathbb{P}_{x\sim\mathsf{unif}([N]^d)}[h(x) \ne f(x)] \le \varepsilon$. Again using our bound on the TV-distance, we have

$$\mathbb{P}_{x\sim\mathcal{D}}[h(x) \ne f(x)] - \mathbb{P}_{x\sim\mathsf{unif}([N]^d)}[h(x) \ne f(x)] \le \left\|\mathcal{D} - \mathsf{unif}([N]^d)\right\|_{TV} \le \gamma$$

and so $\mathbb{P}_{x\sim\mathcal{D}}[h(x) \ne f(x)] \le \varepsilon + \gamma$. □

11.4 Low-Degree Algorithm with RCN: Proof of Theorem 11.2.4

In this section we prove Theorem 11.2.4, showing that concept classes with bounded Fourier degree can be learned efficiently in the presence of random classification noise (RCN). This fact is already implicit from previous works [LMN93, Kea98], but we give a proof for the sake of completeness.

For $S \subseteq [d]$, the parity function $\chi_S\colon \{\pm 1\}^d \to \{\pm 1\}$ is defined as $\chi_S(x) = \prod_{i=1}^d x_i$. The parity functions form a Fourier basis for the space of functions $f\colon \{\pm 1\}^d \to \{\pm 1\}$ and the unique representation of f is given by

$$f(x) = \sum_{S\subseteq[d]} \widehat{f}(S)\chi_S(x) \text{ where } \widehat{f}(S) = \mathbb{E}_x[\chi_S(x)f(x)]$$

is *Fourier coefficient* for f on S. The idea of the Low-Degree Algorithm is to learn f by learning its low-degree Fourier coefficients. From the definition of $\widehat{f}(S)$, observe that an estimate of $\widehat{f}(S)$ can be viewed as a call to a statistical query oracle, which returns an estimate of $\widehat{f}(S)$ to within some specified *allowed query error*, α. In [Kea98], Kearns showed how to simulate statistical query algorithms using only examples with classification noise.

Theorem 11.4.1 (Theorem 3 of [Kea98], paraphrased). *Suppose there is an algorithm which learns a concept class $\mathcal{C}$ of Boolean functions over $\{\pm 1\}^d$ to error ε, using at most Q statistical queries with allowed query error α. Then, for any $\varepsilon, \delta \in (0,1)$, there is a learning algorithm for $\mathcal{C}$ which on any input $f \in \mathcal{C}$, uses at most*

$$O\left(\left(\frac{1}{\alpha(1-2\eta)}\right)^2 \log\left(\frac{Q}{\delta}\right) + \frac{1}{\varepsilon^2}\log\left(\frac{1}{\delta\alpha(1-2\eta)}\right)\right)$$

examples from $EX^\eta(f, \mathsf{unif}(\{\pm 1\}^d))$ and outputs a hypothesis h where $\mathbb{P}_h[d(f,h) \le \varepsilon] \ge 1-\delta$.

In light of the above, we prove Theorem 11.2.4 by first giving an efficient statistical query algorithm, and then applying Theorem 11.4.1.

Proof of Theorem 11.2.4. Since we assume $\sum_{S\colon |S|>\tau} \widehat{f}(S)^2 \le \varepsilon/2$ for all $f \in \mathcal{C}$, the idea is to use a statistical query to obtain an estimate of $\widehat{f}(S)$ for all $|S| \le \tau$. Define $A := \{S \subseteq [d] : |S| \le \tau\}$ and note that

$$|A| = \sum_{i=0}^{\tau} \binom{d}{i} \le \left(\frac{de}{\tau}\right)^\tau \le O(d^\tau). \tag{11.8}$$

We define our statistical query algorithm L to do the following:

1. For each $S \in A$, make a statistical query for an estimate of $\widehat{f}(S) = \mathbb{E}_x[\chi_S(x) f(x)]$ to allowed query error $\alpha = \sqrt{\varepsilon/(2|A|)}$. Let Z_S denote the obtained estimate for $\widehat{f}(S)$.
2. Return $h\colon \{\pm 1\}^d \to \{\pm 1\}$ where $h(x) = \mathtt{sgn}\left(\sum_{S\in A} Z_S \chi_S(x)\right)$.

We now prove that this hypothesis h satisfies $\mathbb{P}_x\left[f(x) \ne h(x)\right] \le \varepsilon$. First, observe that

$$\mathbb{P}_x\left[f(x) \ne h(x)\right] = \frac{1}{4}\mathbb{E}_x\left[(f(x) - h(x))^2\right]. \tag{11.9}$$

Now, if $f(x) \ne h(x)$, then $\left(f(x) - \sum_{S\in A} Z_S\chi_S(x)\right)^2 \ge 1 = \frac{1}{4}(f(x) - h(x))^2$. In the other case, clearly if $f(x) = h(x)$, then $\left(f(x) - \sum_{S\in A} Z_S\chi_S(x)\right)^2 \ge 0 = \frac{1}{4}(f(x) - h(x))^2$. Thus, for any $x \in \{\pm 1\}^d$, this inequality holds. Combining this observation with eq. (11.9) yields

$$\mathbb{P}_x\left[f(x) \neq h(x)\right] \leq \mathbb{E}_x\left[\left(f(x) - \sum_{S \in A} Z_S \chi_S(x)\right)^2\right]. \tag{11.10}$$

In the next calculation, for $S \notin A$, let $Z_S := 0$. Now, writing $f(x) = \sum_S \widehat{f}(S)\chi_S(x)$, expanding the squared sum, applying linearity of expectation, and using the fact that $\mathbb{E}_x[\chi_S(x)] = 0$ for any $S \neq \emptyset$, the RHS of eq. (11.10) is equal to

$$\begin{aligned}
\mathbb{E}_x\left[\left(\sum_S \chi_S(x)\left(\widehat{f}(S) - Z_S\right)\right)^2\right] &= \mathbb{E}_x\left[\sum_{S,T} \chi_{S\Delta T}(x)\left(\widehat{f}(S) - Z_S\right)\left(\widehat{f}(T) - Z_T\right)\right] \\
&= \sum_S \left(\widehat{f}(S) - Z_S\right)^2. \qquad (11.11)
\end{aligned}$$

Using eq. (11.10), eq. (11.11), and the fact that $|\widehat{f}(S) - Z_S| \leq \alpha$ for $S \in A$ and $\sum_{S \notin A} \widehat{f}(S)^2 \leq \varepsilon/2$, yields

$$\begin{aligned}
d(f,h) = \mathbb{P}_x\left[f(x) \neq h(x)\right] &\leq \sum_S \left(\widehat{f}(S) - Z_S\right)^2 \\
&= \sum_{S \in A} \left(\widehat{f}(S) - Z_S\right)^2 + \sum_{S \notin A} \widehat{f}(S)^2 \leq |A| \frac{\varepsilon}{2|A|} + \varepsilon/2 = \varepsilon.
\end{aligned}$$

Thus, L makes $|A| \leq O(d^\tau)$ statistical queries to f with query error $\alpha = \sqrt{\varepsilon/(2|A|)}$ and returns a hypothesis h satisfying $d(f,h) \leq \varepsilon$. Therefore, applying Theorem 11.4.1 completes the proof of Theorem 11.2.4. □

11.5 Coupon Collecting Learner: Proof of Lemma 11.2.1

Proof. The learner is simple. Take s samples from $EX^\eta(f, \mathsf{unif}([N]^d))$ and for each $x \in [N]^d$, let m_x denote the number of times x has been sampled. Let m_x^+, m_x^- denote the number of times x has been sampled with the label $+1, -1$ respectively. The learner outputs the hypothesis $h\colon [N]^d \to \{\pm 1\}$ defined by $h(x) = \mathsf{sgn}(m_x^+ - m_x^-)$.

Claim 11.5.1. *Suppose that* $m_x \geq m := \frac{2}{(1-2\eta)^2} \ln(2/\beta)$. *Then* $\mathbb{P}[\mathsf{sgn}(m_x^+ - m_x^-) \neq f(x)] \leq \beta$.

Proof. Each label seen for x is an independent $\{\pm 1\}$-valued random variable which is equal to $f(x)$ with probability $1-\eta$ and so $\mathbb{E}[m_x^+ - m_x^-] = m_x((1-\eta)f(x) + \eta(-f(x))) = m_x(1-2\eta)f(x)$. Thus,

$$\begin{aligned}\mathbb{P}\left[\mathsf{sgn}(m_x^+ - m_x^-) \neq f(x)\right] &\leq \mathbb{P}\left[\left|(m_x^+ - m_x^-) - \mathbb{E}[m_x^+ - m_x^-]\right| \geq m_x(1-2\eta)\right] \\ &\leq 2\exp\left(-\frac{2m_x^2(1-2\eta)^2}{4m_x}\right) \leq \beta\end{aligned}$$

by Hoeffding's inequality and our bound on m_x. □

Claim 11.5.2. *Suppose we take* $s := \ln(\frac{1}{\alpha})N^{2d}$ *samples. Then* $\mathbb{P}[\exists x \in [N]^d\colon m_x = 0] < \alpha$.

Proof. For any x, $\mathbb{P}[m_x = 0] = (1 - N^{-d})^s \leq \exp(-\frac{s}{N^d}) \ll \alpha N^{-d}$ and a union bound completes the proof. □

The following claim is an immediate corollary of the previous claim.

Claim 11.5.3. *Suppose we take* $m \cdot \ln(\frac{2m}{\delta})N^{2d}$ *samples. Then* $\mathbb{P}[\exists x \in [N]^d\colon m_x < m] < \frac{\delta}{2}$.

Proof. Partition the samples into m batches of size $\ln(\frac{2m}{\delta})N^{2d}$. Invoke Claim 11.5.2 on each batch of samples with $\alpha := \frac{\delta}{2m}$. By the claim, each batch of samples contains a least 1 copy of every point in $[N]^d$ with probability at least $1 - \frac{\delta}{2m}$. Thus, by a union bound over the m batches, our sample contains at least m copies of every point in $[N]^d$ with probability at least $1 - \frac{\delta}{2}$. □

Let $m := \frac{2}{(1-2\eta)^2} \ln\left(\frac{4}{\varepsilon\delta}\right)$ and $s := m\ln(\frac{2m}{\delta})N^{2d}$. The learner takes s samples from $EX^\eta(f, \mathsf{unif}([N]^d))$. Let $\mathcal{E}$ denote the event that $m_x \geq m$ for all $x \in [N]^d$. By Claim 11.5.3, we have $\mathbb{P}[\mathcal{E}] \geq 1 - \frac{\delta}{2}$. For each $x \in [N]^d$, let $\mathcal{B}_x = \mathbf{1}(\mathsf{sgn}(m_x^+ - m_x^-) \neq f(x))$, i.e. the indicator that x is misclassified by the learner.

By Claim 11.5.1, we have

$$\mathbb{E}\left[\sum_{x\in[N]^d} \mathcal{B}_x \;\middle|\; \mathcal{E}\right] \le \frac{\varepsilon\delta}{2} N^d \text{ and thus } \mathbb{P}\left[\sum_{x\in[N]^d} \mathcal{B}_x > \varepsilon N^d \;\middle|\; \mathcal{E}\right] \le \frac{\delta}{2}$$

by Markov's inequality. Therefore,

$$\begin{aligned}
\mathbb{P}\left[\sum_{x\in[N]^d} \mathcal{B}_x > \varepsilon N^d\right] &= \mathbb{P}\left[\sum_{x\in[N]^d} \mathcal{B}_x > \varepsilon N^d \;\middle|\; \mathcal{E}\right]\mathbb{P}[\mathcal{E}] + \mathbb{P}\left[\sum_{x\in[N]^d} \mathcal{B}_x > \varepsilon N^d \;\middle|\; \neg\mathcal{E}\right]\mathbb{P}[\neg\mathcal{E}] \\
&\le \mathbb{P}\left[\sum_{x\in[N]^d} \mathcal{B}_x > \varepsilon N^d \;\middle|\; \mathcal{E}\right] + \mathbb{P}[\neg\mathcal{E}]
\end{aligned}$$

which is at most δ. The number of examples used by the learner is

$$s = m \ln\left(\frac{2m}{\delta}\right) N^{2d} = \widetilde{O}\left(\frac{1}{(1-2\eta)^2}\left(\ln\frac{1}{\varepsilon} + \ln\frac{1}{\delta}\right)\right) \cdot N^{O(d)}$$

and this completes the proof. □

11.6 Testing by Learning

Lemma 11.6.1. *Let $\mathcal{X}$ be a domain, let μ be a measure over $\mathcal{X}$, and let $\mathcal{F}\colon \mathcal{X} \to \{\pm 1\}$ be a class of Boolean-valued functions over $\mathcal{X}$. Suppose that for every $\varepsilon \in (0,1)$ there exists a learning algorithm L for $\mathcal{F}$ under μ using $s(\varepsilon)$ samples. Then for every $\varepsilon \in (0,1)$ there is an ε-tester for $\mathcal{F}$ under μ using $s(\varepsilon/4) + O(1/\varepsilon^2)$ samples.*

Proof. We define the property testing algorithm T as follows.

1. Take $s := s(\varepsilon/4, \mathcal{X})$ samples $(x_1, f(x_1)), \ldots, (x_s, f(x_s))$ and run L to obtain a hypothesis h for f.

2. Compute a function $g \in \mathcal{F}$ for which $d(h,g) = d(h,\mathcal{F})$. (We remark that this step

incurs a blowup in time-complexity, but does not require any additional samples.)

3. Take $s' := \frac{20}{\varepsilon^2}$ new samples $(z_1, f(z_1)), \ldots, (z_{s'}, f(z_{s'}))$ and let $\alpha := \frac{1}{s'} \sum_{i=1}^{s'} \mathbf{1}(g(z_i) \neq f(z_i))$ be an empirical estimate for $d(f, g)$.

4. If $\alpha \leq \frac{3\varepsilon}{4}$, then accept. If $\alpha > \frac{3\varepsilon}{4}$, then reject.

Claim 11.6.2. *If $f \in \mathcal{F}$, then $\mathbb{P}[d(f, g) \leq \varepsilon/2] \geq 5/6$.*

Proof. By the guarantee of the learning algorithm, we have $\mathbb{P}[d(f, h) \leq \varepsilon/4] \geq 5/6$. Now, since g is a function in $\mathcal{F}$ as close as possible to h, we have $d(h, g) \leq d(h, f)$. Thus, if $d(f, h) \leq \varepsilon/4$, then $d(h, g) \leq \varepsilon/4$ as well. Thus, by the triangle inequality, with probability at least $5/6$ we have $d(f, g) \leq d(f, h) + d(h, g) \leq \varepsilon/2$ as claimed. $\square$

Now, consider the quantity α from step (4) of the algorithm, T. Let X be the Bernoulli random variable which equals 1 with probability $d(f, g)$. Note that $\alpha = \frac{1}{s'} \sum_{i=1}^{s'} X_i$ where the X_i's are independent copies of X. Using Hoeffding's inequality we have

$$\mathbb{P}\left[|\alpha - d(f, g)| \geq \frac{\varepsilon}{4}\right] = \mathbb{P}\left[\left|\sum_{i=1}^{s'} X_i - s' \cdot d(f, g)\right| \geq \frac{s'\varepsilon}{4}\right] \leq 2\exp\left(-\frac{2 \cdot (s'\varepsilon/4)^2}{s'}\right) = \frac{2}{e^{\frac{s'\varepsilon^2}{8}}}$$

which is at most $1/6$ when $s' \geq 8\ln(12)/\varepsilon^2$. We can now argue that the tester T succeeds with probability at least $2/3$. There are two cases to consider.

1. $f \in \mathcal{F}$: By Claim 11.6.2, $d(f, g) > \varepsilon/2$ with probability less than $1/6$ and by the above calculation $|\alpha - d(f, g)| \geq \varepsilon/4$ with probability at most $1/6$. By a union bound, with probability at least $2/3$ neither event occurs, and conditioned on this we have $\alpha \leq d(f, g) + \varepsilon/4 \leq 3\varepsilon/4$ and the algorithm accepts.

2. $d(f, \mathcal{F}) \geq \varepsilon$: Then $d(f, g) \geq \varepsilon$ since $g \in \mathcal{F}$. Again, $|\alpha - d(f, g)| < \varepsilon/4$ with probability at least $5/6$ and conditioned on this event occurring we have $\alpha > d(f, g) - \varepsilon/4 \geq 3\varepsilon/4$ and the algorithm rejects.

Therefore, T satisfies the conditions needed for Lemma 11.6.1. $\square$

Part IV

Testing and Learning Convex Sets in the Ternary Hypercube

CHAPTER 12

Introduction

A subset $S \subseteq [m]^n$ of the hypergrid is *discrete convex* if it is the intersection of a convex set $C \subseteq \mathbb{R}^n$ with the grid, $S = C \cap [m]^n$, or equivalently if $S = [m]^n \cap \mathsf{Conv}(S)$ where $\mathsf{Conv}(S)$ is the convex hull of S. Discrete convex sets may not even be *connected* (see Figure 12.1), which, along with some of their other unpleasant features, makes them difficult to handle algorithmically and analytically, the most famous example being the difference between linear programming and integer linear programming.

We are interested in testing and learning discrete convex sets. A learning algorithm should output an approximation of an unknown convex set S by using membership queries to S, while a testing algorithm should decide whether an unknown set S is either convex or ϵ-*far* from convex, meaning that $\mathsf{dist}(S, T) > \epsilon$ for all convex sets T, where $\mathsf{dist}(S, T)$ is the measure of the symmetric difference.

Convexity is particularly interesting for property testing because it can be defined by a local condition: a set $S \subseteq \mathbb{R}^n$ is convex if and only if for every 3 colinear points x, y, z, if $x, z \in S$ then $y \in S$. This means that, to certify the non-convexity of a (continuous) set, it suffices to provide 3 colinear points that violate this condition. Speaking informally, property testing results, especially testing with *one-sided error*, are statements about the difficulty of finding such a certificate of non-membership to the property, when the object S is ϵ-far from satisfying the property. But, the fact that convexity is *defined* by a local condition does not make it easy to find violations of the condition when a set is far from convex. This is particularly evident for *discrete* convex sets where, unlike continuous sets, there may not be *any* lines which witness non-convexity, and one must instead look for up to $n+1$ points that violate Carathéodory's theorem.

We are aware of no non-trivial algorithms for testing or learning discrete convex sets in high dimensional grids $[m]^d$ when m is small. Prior works on testing and learning convex sets include:

1. The analysis of convexity testers, such as the *line tester* and more general *convex hull testers*, which are designed to simply "spot-check" for violations of the local conditions that define convexity in $\mathbb{R}^n$ [RV04, BB20]. These works show that these spot-checkers are not very efficient, requiring $2^{\Omega(n)}$ queries to detect sets that are $\Omega(1)$-far from convex.

2. Testing or learning convex sets in two dimensions, including the continuous square $[0,1]^2$ [Sch92, BMR19a] or the discrete grid $[m]^2$ [Ras03, BMR19b, BMR22].

3. Testing convexity in high dimensions with *samples*, either in the continuous setting [CFSS17b, HY22] or discrete setting [HY22], and learning convex sets from random examples of the set [RG09] or from Gaussian samples [KOS08].

When $m \gg \mathrm{poly}(d)$, a "downsampling" or "gridding" approach can reduce to the case $m = \mathrm{poly}(d)$ [CFSS17b, HY22], but once m is small the only known algorithm for testing or learning is brute-force. So let us see what happens when we make m as small as possible. When $m = 2$, testing and learning convex sets in $[m]^n \equiv \{0,1\}^n$ is trivial, because every subset of $\{0,1\}^d$ is convex and therefore testing is as easy as possible (the tester may simply accept on every input) and learning is as hard as possible (requiring $\Omega(2^n)$ queries).

The story changes significantly when $m = 3$, so that $[m]^n$ is equivalent to the ternary hypercube $\{0,\pm1\}^n$, where the difficulties of handling high-dimensional discrete convex sets suddenly become evident. Although this is the simplest domain where where high-dimensional discrete convex sets are non-trivial, little is known about the structure of discrete convex sets on the ternary hypercube that would help in designing testing and learning

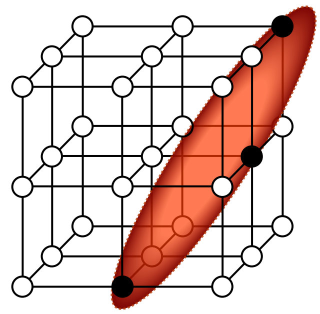


Figure 12.1: *Example of a convex set in* $\{0, \pm 1\}^3$. *The black dots are the set and the convex red ellipsoid contains them. Note that the set may not be "connected" on the hypergrid.*

algorithms. In this paper we will give the first results towards understanding testing and learning discrete sets in high dimensions by focusing on the ternary hypercube.

12.1 Results

For two sets $S, T \subseteq \{0, \pm 1\}^n$, we define $\mathsf{dist}(S, T) := \frac{|S \Delta T|}{3^n}$, where $S \Delta T$ denotes the symmetric difference. A set $S \subseteq \{0, \pm 1\}^n$ is ε-far from convex if for every (discrete) convex set $T \subseteq \{0, \pm 1\}^n$, $\mathsf{dist}(S, T) \geq \varepsilon$. Given $\varepsilon > 0$, a *convexity tester* is a randomized algorithm which is given membership oracle access to an input $S \subseteq \{0, \pm 1\}^n$ and must satisfy

1. If S is convex then the algorithm accepts with probability at least $2/3$.

2. If S is ε-far from convex then the algorithm rejects with probability at least $2/3$.

The tester is *one-sided* if it must accept convex sets S with probability 1 instead of $2/3$. A tester is *non-adaptive* if it chooses its set of queries before receiving the answers to any of the queries and it is *sample-based* if its queries are independently and uniformly random.

A learning algorithm is given membership oracle access to a convex set $S \subseteq \{0, \pm 1\}^n$ and must output (with probability at least $2/3$) a set $T \subseteq \{0, \pm 1\}^n$ with $\mathsf{dist}(S, T) < \varepsilon$; it is *proper* if its output T must be convex.

12.1.1 The Edge Boundary and Influence of Convex Sets

One of the most important things to know about a set is its *edge boundary*. The edge set of the ternary hypercube is defined as

$$E = \left\{ (x, y) \in (\{0, \pm 1\}^n)^2 \colon \sum_{i=1}^{n} |x_i - y_i| = 1 \right\}. \tag{12.1}$$

Observe that $|E| = 2n \cdot 3^{n-1}$. We will identify a set $S \subseteq \{0, \pm 1\}^n$ with its characteristic function and write $S(x) = 1$ if $x \in S$ and $S(x) = 0$ otherwise. An edge (x, y) is on the boundary of S if $S(x) \neq S(y)$. The influence of a set $S \subseteq \{0, \pm 1\}^n$ is its normalized boundary size:

$$\mathbb{I}(S) := \frac{1}{3^n} \cdot |\{(u, v) \in E \colon S(u) \neq S(v)\}| = \frac{2n}{3} \cdot \mathbb{P}_{(u,v) \sim E}[S(u) \neq S(v)]. \tag{12.2}$$

Before we state our results, consider some examples. Two important classes of convex sets in $\{0, \pm 1\}^n$ are halfspaces and balls, which often have minimal "boundary size" in various settings.

Example 12.1.1 (Halfspaces). *A halfspace is a set $H = \{x \in \{0, \pm 1\}^n : \langle v, x \rangle < \tau\}$ where $v \in \mathbb{R}^n$ and $\tau \in \mathbb{R}$. To maximize the influence, we want τ to be small, say $\tau = 0$, and we want $v \approx \vec{1}$. The probability that a random edge (x, y) is on the boundary is at most the probability that a uniformly random $x \sim \{0, \pm 1\}^n$ satisfies $|\langle \vec{1}, x \rangle| \leq 1$, and it is not difficult to show that this is at most $O\left(\frac{1}{\sqrt{n}}\right)$, giving an estimate of $O(\sqrt{n})$ for the maximum influence of a halfspace.*

Example 12.1.2 (Balls). *A ball is a set $B_r = \{x \in \{0, \pm 1\}^n : \|x\|_2^2 < r\}$ where $r \in \mathbb{R}$ is the radius. The average (squared) norm $\mathbb{E}[\|x\|_2^2]$ for $x \sim \{0, \pm 1\}^n$ is the same as the expected number of nonzero coordinates of x, which is $\frac{2}{3}n$, so to maximize the edge boundary we think of $r \approx \frac{2}{3}n$. Similar to above, the probability that $x \sim \{0, \pm 1\}^n$ is close enough to this threshold to find a boundary edge is $O\left(\frac{1}{\sqrt{n}}\right)$, again giving an estimate of $O(\sqrt{n})$ for the* *maximum influence.*

Our first result shows that there are convex sets with significantly larger influence, which can be obtained by taking S to be the intersection of roughly $3^{\Theta(\sqrt{n})}$ random halfspaces with thresholds $\tau = \Theta(n^{3/4})$; we think of these sets as interpolating between the halfspaces and the ball. Our construction is inspired by [Kan14], who showed bounds on the influence of intersections random halfspaces on the hypercube $\{0,1\}^n$, and we note that similar constructions also achieve maximal surface area under the Gaussian distribution on $\mathbb{R}^n$ [Naz03].

Theorem 12.1.3. *There exists a convex set $S \subseteq \{0, \pm 1\}^n$ with influence $\mathbb{I}(S) = \Omega(n^{3/4})$.*

Our main result on the influence of convex sets is that this construction is essentially optimal: we show a matching upper bound (up to log factors) for any convex set in $\{0, \pm 1\}^n$. Due to the discrete nature of the domain, our proof of this theorem is significantly different from the previous techniques that have been used to bound the surface area of convex sets in continuous domains.

Theorem 12.1.4. *If $S \subseteq \{0, \pm 1\}^n$ is convex, then $\mathbb{I}(S) = O(n^{3/4} \log^{1/4} n)$.*

12.1.2 Sample-Based Learning and Testing

As an application of our bounds on the influence, we show using standard Fourier analysis that any set $S \subseteq \{0, \pm 1\}^n$ can be approximated with error ε by a polynomial of degree $\mathbb{I}(S)/\varepsilon$. Using Theorem 12.1.4 and the "Low-Degree Algorithm" of Linial, Mansour, and Nisan [LMN93] then gives us the following upper bound for learning.

Theorem 12.1.5. *There is a uniform-distribution learning algorithm for convex sets in $\{0, \pm 1\}^n$ which achieves error at most ε with time and sample complexity $3^{\widetilde{O}(n^{3/4}/\varepsilon)}$. The $\widetilde{O}(\cdot)$ hides a factor of $\log^{1/4} n$.*

A corollary of Theorem 12.1.5 is that the same upper bound on the sample complexity holds for sample-based testing, due to the testing-by-learning reduction (which is slightly non-standard because the learner is not *proper*, see Section 11.6).

Corollary 12.1.6. *There is a sample-based convexity tester for sets in* $\{0, \pm 1\}^n$ *with sample complexity* $3^{\widetilde{O}(n^{3/4}/\varepsilon)}$ *where the* $\widetilde{O}(\cdot)$ *hides a factor of* $\log^{1/4} n$.

To complement our upper bounds, we prove also a lower bound for sample-based testing. Here we remark that one of our motivations for studying convex sets in $\{0, \pm 1\}^n$ is their similarity (in an informal sense) to *monotone* functions on $\{0, 1\}^n$; an analogy between monotone functions on $\{0, 1\}^n$ and convex sets in Gaussian space was proposed in [DNS22] and we are interested in this analogy for *discrete* convex sets. Our lower bound for sample-based testing discrete convex sets uses a version of Talagrand's random DNFs, which were used previously to prove lower bounds for testing monotonicity on $\{0, 1\}^n$ [BB21, CWX17].

Theorem 12.1.7. *For sufficiently small constant* $\varepsilon > 0$*, every sample-based convexity tester for sets in* $\{0, \pm 1\}^n$ *has sample complexity* $3^{\Omega(\sqrt{n})}$.

Again, the testing-by-learning reduction of Lemma 11.6.1 implies that this lower bound also holds for learning.

Corollary 12.1.8. *For sufficiently small constant* $\varepsilon > 0$*, sample-based learning convex sets in* $\{0, \pm 1\}^n$ *requires at least* $3^{\Omega(\sqrt{n})}$ *samples.*

12.1.3 Non-Adaptive One-Sided Testing

A convexity tester with one-sided error is one that finds a *witness* of non-convexity with probability at least $2/3$ when the tested set is ε-far from convex. A convexity tester is *non-adaptive* if it must choose its set of membership queries before receiving any of the query results. Bounds on non-adaptive one-sided error testing therefore have a natural combinatorial interpretation as bounds on the likelihood of blindly finding a witness of non-convexity in a random substructure of the domain.

Our first result shows that there is a non-adaptive one-sided error tester with sub-exponential query complexity $3^{o(n)}$. In contrast, a similar bound for the Gaussian setting is not yet known to exist.

Theorem 12.1.9. *For every $\varepsilon > 0$, there is a non-adaptive convexity tester with one-sided error for sets in $\{0, \pm 1\}^n$ that has query complexity $3^{\widetilde{O}\left(\sqrt{n \ln 1/\varepsilon}\right)}$ where the $\widetilde{O}(\cdot)$ notation is hiding an extra $\ln n$ term.*

Next, we show that Theorem 12.1.9 is essentially tight, in that the exponential dependence on $\sqrt{n}$ in its bound is unavoidable.

Theorem 12.1.10. *For sufficiently small constant $\varepsilon > 0$, every non-adaptive convexity tester with one-sided error for sets in $\{0, \pm 1\}^n$ has query complexity at least $3^{\Omega(\sqrt{n})}$.*

Our Theorem 12.1.7 above showed that $3^{\Omega(\sqrt{n})}$ is required for *sample-based* testing. For one-sided error testers, we can improve this lower bound to show that non-adaptive testers are significantly more powerful than sample-based testers for one-sided testing.

Theorem 12.1.11. *For sufficiently small constant $\varepsilon > 0$, sample-based convexity testing in $\{0, \pm 1\}^n$ with one-sided error requires $3^{\Theta(n)}$ samples.*

This theorem also includes a matching upper bound. The upper bound in Theorem 12.1.11 is trivial because a coupon-collector argument shows that one can learn any set $S \subseteq \{0, \pm 1\}^n$ exactly using $O(n3^n)$ samples. A slightly improved bound of $O\left(3^n \cdot \frac{1}{\varepsilon} \log(1/\varepsilon)\right)$ also holds by a general upper bound on one-sided error testing via the VC dimension [BFH21].

12.2 Techniques

The discrete nature of the ternary hypercube, in contrast to the continuity of the domains $\mathbb{R}^n$ or $[0,1]^n$, provides a new angle in the study of convexity which leads to the development of a new set of *combinatorial* techniques and tools. In this section we give a brief overview of the techniques we use to prove each of our theorems.

12.2.1 The Edge Boundary and Influence of Convex Sets

Influence Upper Bound: Our proof of Theorem 12.1.4, which gives an upper bound on the edge boundary of a convex set, is accomplished by relating the number of boundary edges to the expected number of sign-changes of one-dimensional random processes. This is done by constructing a distribution $\mathcal{D}$ over the edge-set E of the ternary hypercube, such that (a) $\mathcal{D}$ is "close" to the uniform distribution over E and (b) the probability that a random edge drawn from $\mathcal{D}$ is influential for our convex set $S \subseteq \{0, \pm 1\}^n$ is equal to the expected number of sign-changes of a certain random process. This process is defined by considering a random walk $\boldsymbol{X}^{(0)}, \ldots, \boldsymbol{X}^{(m)}$ of length $m \approx n^{1/2}$ where $\boldsymbol{X}^{(0)}$ is a random point from the middle layers of $\{0, \pm 1\}^n$ and each $\boldsymbol{X}^{(s)}$ is obtained by flipping a random 0-valued bit of $\boldsymbol{X}^{(s-1)}$ to a uniform random $\{\pm 1\}$-value; the process finally draws $\boldsymbol{s} \sim [m]$ uniformly at random and outputs the edge $(\boldsymbol{X}^{(\boldsymbol{s}-1)}, \boldsymbol{X}^{(\boldsymbol{s})})$.

The crux of the argument is to bound the expected number of times this random walk enters and leaves the set S. Since S is convex, it can be written as an intersection of halfspaces $S = H_1 \cap H_2 \cap \cdots \cap H_k$ of the form $H_i = \{x \in \{0, \pm 1\}^n \colon \langle x, v^{(i)} \rangle < \tau_i\}$ where $v^{(i)} \in \mathbb{R}^n$ and $\tau_i \in \mathbb{R}$. For each halfspace H_i, we define a corresponding one-dimensional random walk $\boldsymbol{W}_i(s) = \langle \boldsymbol{X}^{(s)}, v^{(i)} \rangle - \tau_i$ and observe that the original random walk crosses the boundary of H_i at step s if and only if $\boldsymbol{W}_i$ *changes sign* at step s. Then the number of times the walk $\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)}, \ldots$ crosses the boundary of $S = \bigcap_i H_i$ is the number of times the *maximum* of the processes $\boldsymbol{M} = \max_i \boldsymbol{W}_i$ changes sign. Therefore, our goal is to bound the expected number of sign-changes for $\boldsymbol{M}$, which we accomplish by using Sparre Andersen's fluctuation theorem [Spa54] (as stated in [BB23]) to relate this quantity to the number of sign-changes of a uniform random walk.

High-Influence Set Construction: Our proof of Theorem 12.1.3 is inspired by the proof of [Kan14, Theorem 2] which constructs a set in the Boolean hypercube $\{\pm 1\}^n$ with influence $\Omega(\sqrt{n \log k})$ by considering an intersection of k random halfspaces each of which is at distance $\approx \sqrt{n \log k}$ from the origin. In particular, when $k \approx 2^{\sqrt{n}}$ the construction has influence $\approx n^{3/4}$ and when $k \approx 2^n$ the set has influence $\approx n$. On the ternary hypercube $\{0, \pm 1\}^n$, the behaviour is different: here, halfspaces exhibit a "density increment" behaviour as their threshold moves away from the origin, which prevents the influence from increasing as k grows past $2^{\sqrt{n}}$, when $\Omega(\sqrt{n \log k})$ matches our upper bound of $\widetilde{O}(n^{3/4})$.

We can summarize this "density increment" phenomenon as follows. Most of the edges of $\{0, \pm 1\}^n$ occur in the middle layer $\{x \in \{0, \pm 1\}^n : \|x\|_1 = \frac{2}{3}n \pm O(\sqrt{n})\} = \bigcup_{\ell=-O(\sqrt{n})}^{O(\sqrt{n})} \{x : \|x\|_1 = \frac{2}{3}n + \ell\}$. A convex set is an intersection of halfspaces, but for convenience we consider its complement which is a union of halfspaces, and has the same influence. Consider the "density" or measure of the halfspace with normal vector $\vec{1}$ at distance τ from the origin on the points $\{x : \|x\|_1 = \frac{2n}{3} + \ell\}$:

$$\rho(\ell, \tau) := \mathbb{P}_{x \in \{0, \pm 1\}^n :\, \|x\|_1 = \frac{2n}{3} + \ell} \left[\sum_i x_i > \tau \right] .$$

Suppose that there is a fixed value ρ such that $\rho(\ell, \tau) \approx \rho$ up to constant factors for *all* $\ell = \pm O(\sqrt{n})$ simultaneously. Then we can take $k \approx \frac{1}{\rho}$ random halfspaces with threshold τ and combine their boundary edges, since they will be essentially disjoint on the whole middle layer, and it is not hard to show that the influence of the resulting union is roughly τ. It happens that the condition of $\rho(\ell, \tau)$ being approximately equal for all values $\ell = \pm O(\sqrt{n})$ holds for τ up to $\tau \approx n^{3/4}$ but for $\tau \gg n^{3/4}$ the intersection of the halfspace with the set $\{x : \|x\|_1 = \frac{2n}{3} + \ell\}$ grows extremely fast with ℓ making $\rho(-\sqrt{n}, \tau) \ll \rho(\sqrt{n}, \tau)$, and the intersection of halfspaces with threshold τ quickly approaches the ball with influence $O(\sqrt{n})$ (see Example 12.1.2).

12.2.2 Sample-Based Learning and Testing

Learning Upper Bound: Our proof of Theorem 12.1.5 follows by combining our upper bound on the influence from Theorem 12.1.4 with the Low-Degree Algorithm of Linial, Mansour, and Nisan [LMN93]. In particular, using Fourier analysis over $\{0, \pm 1\}^n$ in combination with Theorem 12.1.4 we can show that for convex sets, a $(1 - \varepsilon)$-fraction of the Fourier mass is on the coefficients with degree at most $\widetilde{O}(n^{3/4})/\varepsilon$. Then we may use the Low-Degree Algorithm for learning the convex sets; see Section 14.1.3. Since the ternary hypercube is a non-standard domain, we state the necessary Fourier analysis for functions over $\{0, \pm 1\}^n$ in Section 14.1.1, which follows [O'D14, Chapter 8]. One technical difference between Fourier analysis over the Boolean and ternary hypercubes is that the standard Fourier basis over $\{\pm 1\}^n$ is given by the parity functions which are bounded in $[0, 1]$, whereas any Fourier basis over $\{0, \pm 1\}^n$ will have functions taking value $2^{O(n)}$ on some elements $x \in \{0, \pm 1\}^n$. Nevertheless, with some care, we show that the Low-Degree Algorithm still works.

Sample-Based Testing Lower Bound: Our proof of Theorem 12.1.7 uses a family of functions known as *Talagrand's random DNFs* adapted to the ternary hypercube. As we mentioned, this family of functions has been used to prove lower bounds for monotonicity testing [BB21, CWX17]. Our adapted version is described as follows. Each "term" of the DNF is chosen to be a random point $t \in \{0, \pm 1\}^n$ with $\|t\|_1 = \sqrt{n}$. We then say that a point $x \in \{0, \pm 1\}^n$ "satisfies" t if $x_i = t_i$ for all $i \in [n]$ where $t_i \in \{\pm 1\}$. After choosing N random terms $t^{(1)}, \ldots, t^{(N)}$ we define the disjoint regions of $\{0, \pm 1\}^n$ given by $U_1, \ldots, U_N$ where U_i is the set of points $x \in \{0, \pm 1\}^n$ with $\|x\|_1 \in [2n/3 \pm \sqrt{n}]$ which satisfy a unique term. Choosing $N = 3^{\sqrt{n}}$ results in $\bigcup_{i=1}^{N} U_i$ covering a constant fraction of the domain. We then define two distributions $\mathcal{D}_{\texttt{yes}}$ and $\mathcal{D}_{\texttt{no}}$ as follows. Recall that B_r is the radius-r ball in the ternary cube (Example 12.1.2) and let D denote the set of points $x \in \{0, \pm 1\}^n$ with $\|x\|_1 \in [2n/3 \pm \sqrt{n}]$ that don't satisfy *any* term.

- $S \sim \mathcal{D}_{\texttt{yes}}$ is drawn by setting $S = B_{\frac{2n}{3}-\sqrt{n}} \cup D \cup \left(\bigcup_{i \in T} U_i\right)$ where T includes each $i \in [N]$ independently with probability $1/2$. Such a set is always convex.
- $S \sim \mathcal{D}_{\texttt{no}}$ is drawn by setting $S = B_{\frac{2n}{3}-\sqrt{n}} \cup D \cup C$ where C includes each $x \in \bigcup_{i=1}^N U_i$ independently with probability $1/2$. Informally, this set will be $\Omega(1)$-far from convex with constant probability since its intersection with the middle layers is random.

For both distributions, each point $x \in \bigcup_{i=1}^N U_i$ satisfies $\mathbb{P}_S[x \in S] = 1/2$ and if $x \in U_i$ and $y \in U_j$ where $i \neq j$, then the events $x \in S$ and $y \in S$ are independent. Thus, to distinguish $\mathcal{D}_{\texttt{yes}}$ and $\mathcal{D}_{\texttt{no}}$ one has to see at least two points from the same U_i and this gives our sample complexity lower bound.

12.2.3 Non-Adaptive One-Sided Testing

The proofs of Theorems 12.1.9 to 12.1.11 all rely on a partial order $\preceq$ defined on $\{0, \pm 1\}^n$, which we call the *outward-oriented poset*, that has the origin 0^n as the minimum element and the corners of the cube $\{\pm 1\}^n$ as the maximum elements. (See Section 12.4.1 for the formal definition of this poset and a discussion of its properties and history.) For any $y \in \{0, \pm 1\}^n$, we define $\mathsf{Up}(y) := \{x \in \{0, \pm 1\}^n : y \preceq x\}$ to represent the set of points above y in this poset.

Non-Adaptive One-Sided Upper Bound: An important property of the outward-oriented poset in the context of testing convexity is that any point y in the convex hull of a set of points $X \subseteq \{0, \pm 1\}^n$ is also in the convex hull of $X \cap \mathsf{Up}(y)$. Conversely, if a set $S \subseteq \{0, \pm 1\}^n$ is *not* convex, then there is a certificate of non-convexity of the form (X, y) where $y \notin S$ is in the convex hull of $X \subseteq S$, and $X \subseteq \mathsf{Up}(y)$. This property implies that a convexity tester can search for certificates of non-convexity by repeatedly choosing a random point y and querying all points in $\mathsf{Up}(y)$. A naïve implementation of this idea leads to a query complexity that is significantly larger than the bound in the theorem. However, the ternary hypercube satisfies a strong concentration of measure property: almost all of the points in the ternary hypercube have $\frac{2}{3}n \pm O(\sqrt{n})$ non-zero coordinates. As a result, we can refine the convexity tester to only query the points in $\mathsf{Up}(y)$ whose number of non-zero coordinates is at most $\frac{2}{3}n + O(\sqrt{n})$ to obtain the desired query complexity. The details of the proof of Theorem 12.1.9 are presented in Chapter 15.

Non-Adaptive One-Sided Lower Bound: The lower bound in Theorem 12.1.10 is obtained by considering the class of *anti-slabs*, which are defined by choosing a vector $v \in \{0, \pm 1\}^n$ with $n/2$ non-zero coordinates and taking the set of points $\{x \in \{0, \pm 1\}^n : |\langle v, x \rangle| > \tau\}$. It is quite easy to find certificates of non-convexity for anti-slabs—the three points $-x$, 0^n, and x obtained by choosing x uniformly at random in the ternary hypercube forms such a certificate with reasonably large probability whenever τ is small enough. However, we can eliminate these certificates of non-convexity if we "truncate" the anti-slabs by including the set of points whose number of non-zero coordinates is below $\frac{2}{3}n - O(\sqrt{n})$, and excluding the points whose number of non-zero coordinates is above $\frac{2}{3}n + O(\sqrt{n})$. We show that any certificate of non-convexity for these truncated anti-slabs must have two points x, z with a large difference between $\langle v, x \rangle$ and $\langle v, z \rangle$, but on the other hand, any small set of queries has a low probability of including such a pair when v is chosen at random.

Sample-Based One-Sided Lower Bound: Finally, the proof of the lower bound Theorem 12.1.11 again uses the outward-oriented poset and the connection between convex hulls and the upwards sets $\mathsf{Up}(y)$ to show that any set of $3^{o(n)}$ samples is unlikely to draw any point y that is contained in the convex hull of the other sampled points and thus to have any possibility of identifying a certificate of non-convexity of any set.

12.3 Discussion and Open Problems

As far as we know, we are the first to study convex sets and their associated algorithmic problems on the ternary hypercube. Thus there are many possible questions one could ask. In this section we discuss a few such questions which we find most interesting.

Learning and sample-based testing. The most obvious question which our work leaves open is that of determining the true sample complexity of learning and sample-based testing of convex sets in the ternary hypercube, where our results leave a gap of $3^{\Omega(\sqrt{n})}$ vs. $3^{\widetilde{O}(n^{3/4})}$. By Theorem 12.1.3, our upper bound of $\widetilde{O}(n^{3/4})$ on the influence of convex sets is tight up to a factor of $\log^{1/4} n$, and therefore to improve our learning upper bound would require another method.

Question 12.3.1. *Can we close the gap of $3^{\Omega(\sqrt{n})}$ vs. $3^{\widetilde{O}(n^{3/4})}$ for learning convex sets and for sample-based convexity testing in $\{0, \pm 1\}^n$?*

Testing with two-sided error. Our results for testing with queries apply only to case of one-sided error. Earlier work on testing convex sets under the Gaussian distribution on $\mathbb{R}^n$ with samples showed that, in that setting, two-sided error was more efficient than one-sided [CFSS17b].

Question 12.3.2. *Is there a two-sided error non-adaptive tester for domain $\{0, \pm 1\}^n$ with better query complexity than our one-sided error tester?*

Our lower bound technique does not suffice for two-sided error. This is because the class of anti-slabs, which we proved are hard to distinguish from convex sets using a one-sided tester, can be distinguished from convex sets with *two-sided* error using only $O(n)$ samples. To do so, one may use the standard testing-by-learning reduction of [GGR98], together with an $O(n)$ bound on the VC dimension of the anti-slabs (which are essentially the union of two halfspaces).

Testing convexity in other domains. Our results show that queries can be more effective than samples for testing discrete convex sets in some high-dimensional domains. Is this true for all discrete high-dimensional domains?

Question 12.3.3. *What are the sample and query complexities for testing discrete convexity over general hypergrids* $[m]^n$*?*

Note that our techniques do not immediately generalize to larger hypergrids, so answering the last question even for the hypergrid $\{0, \pm 1, \pm 2\}^n$ requires some new ideas.

It would also be interesting to see if the gap between sample and query complexity also holds for continuous sets.

Question 12.3.4. *Can queries improve upon the bounds of [CFSS17b, HY22] for testing convex sets with samples in* $\mathbb{R}^n$ *under the Gaussian distribution?*

It is not clear if there is a formal connection between testing convex sets on the domain $\{0, \pm 1\}^n$ and on the domain $\mathbb{R}^n$ under the standard Gaussian distribution. One might expect a connection here because the uniform distribution on $\{0, \pm 1\}^n$ acts similarly to the Gaussian in certain ways when $n \to \infty$. But we do not see how to construct direct reductions between these two settings for the problem of convexity testing. Also, there is an intriguing analogy between monotone subsets of $\{\pm 1\}^n$ and convex subsets of $\mathbb{R}^n$ in the Gaussian space [DNS22]. How do convex subsets of $\{0, \pm 1\}^n$ fit into this analogy?

12.4 Preliminaries: Convexity on the Ternary Hypercube

The main object of study in this paper is the *ternary hypercube*, an analogue of the Boolean hypercube over the ternary set $\{0, \pm 1\}^n$. This set can be viewed as a discrete subset of $\mathbb{R}^n$, as a (hyper)grid graph in which two points $x, y \in \{0, \pm 1\}^n$ are connected by an edge if and only if $\sum_{i=1}^n |x_i - y_i| = 1$, and as a poset that we will describe in more detail in the subsection below.

The study of the ternary hypercube and more general grid graphs goes back at least to Bollobás and Leader [BL91]. As a poset, its study goes back at least to Metropolis and Rota [MR78]. The ternary hypercube appears to have some particularly elegant structure that is not necessarily shared by larger hypergrids. We describe some of these fundamental properties in the following subsections.

12.4.1 The Outward-Oriented Poset

We define a partial order over $\{0, \pm 1\}^n$, which puts the origin 0^n as the minimum element and the corners $\{\pm 1\}^n$ as the maximum elements.

Definition 12.4.1 (Outward-Oriented Poset)**.** *We denote by* $(\{0, \pm 1\}^n, \preceq)$ *the* n*-wise product of the partial order defined by* $0 \prec 1$ *and* $0 \prec -1$*. Equivalently, we write* $y \preceq x$ *when* $\forall i \in [n] \colon (y_i \neq 0 \implies x_i = y_i)$.

The outward-oriented poset can easily be extended to a lattice (by adding a global maximum point), though since we do not need this extension we do not pursue it here. The outward-oriented poset appears naturally in many different contexts and, as a result, has received different names. For instance, it arises in the study of the faces of the Boolean hypercube [MR78], where it is sometimes called the "cubic lattice", and in the study of partial Boolean functions (see, e.g., [Eng97]). We use the name "outward-oriented poset" to emphasize the fact that this poset is distinct from the partial order inherited from $\mathbb{R}^n$.

Definition 12.4.2 (Upper Shadow)**.** *For any point* $y \in \{0, \pm 1\}^n$*, the* upper shadow *of* y *is the set*

$$\mathsf{Up}(y) := \{x \in \{0, \pm 1\}^n : y \preceq x\} \ .$$

12.4.2 Convexity and Witnesses of Non-Convexity

Given a set of points $X \subseteq \{0, \pm 1\}^n$, we denote the convex hull of X by

$$\mathsf{Conv}(X) := \left\{ \sum_{x \in X} \lambda_x x \colon \sum_{x \in X} \lambda_x = 1 \text{ and } \lambda_x \geq 0, \, \forall x \in X \right\}.$$

Definition 12.4.3 (Discrete Convexity)**.** *A set $S \subseteq \{0, \pm 1\}$ is* convex *if $S = \mathsf{Conv}(S) \cap \{0, \pm 1\}^n$.*

Let $\Delta(S, T)$ denote the cardinality of the symmetric difference between S and T. Given $S \subseteq \{0, \pm 1\}^n$, we define $\mathsf{dist}(S, \mathbf{convex})$ as the minimum, over all convex sets $T \subseteq \{0, \pm 1\}^n$, of $\Delta(S, T) \cdot 3^{-n}$. For brevity, we also sometimes use the notation $\varepsilon(S) := \mathsf{dist}(S, \mathbf{convex})$. If $\varepsilon(S) \geq \varepsilon$ for some $\varepsilon \in (0, 1)$, then we say that S is ε-far from convex.

Definition 12.4.4 (Violating Pairs)**.** *Consider $S \subseteq \{0, \pm 1\}^n$. If $X \subseteq S$ and $y \in \mathsf{Conv}(X) \cap \{0, \pm 1\}^n$, but $y \notin S$, then we call (X, y) a* violating pair *for S. The pair is called* minimal *if $y \notin \mathsf{Conv}(X')$ for any strict subset $X' \subset X$.*

All of our results exploit the following key property of the outward-oriented poset. This fact captures the structure of $\{0, \pm 1\}^n$ which we use throughout the paper.

Fact 12.4.5. *If a violating pair (X, y) is minimal, then $X \subseteq \mathsf{Up}(y)$.*

Proof. We have $y = \sum_{x \in X} \lambda_x x$ where $\sum_{x \in X} \lambda_x = 1$. Moreover, the minimality of (X, y) implies that $\lambda_x > 0$ for all $x \in X$. Now, let $i \in [n]$ be some coordinate where $y_i \neq 0$. We need to show that $x_i = y_i$ for all $x \in X$. Without loss of generality, suppose $y_i = 1$. Thus, we have $1 = \sum_{x \in X} \lambda_x x_i$. If $x_i < 1$ for some $x \in X$, then we would have $\sum_{x \in X} \lambda_x x_i < 1$, which is a contradiction. □

Fact 12.4.6. *Let $S \subseteq \{0, \pm 1\}^n$. The following two statements are equivalent.*

- *S is not convex.*

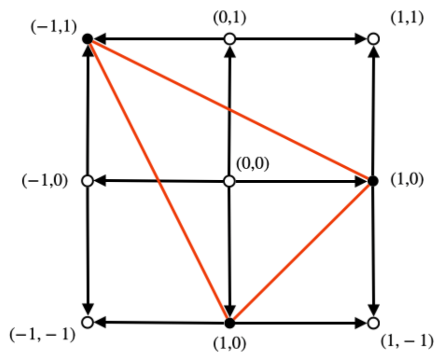


Figure 12.2: *An illustration of* $\{0, \pm 1\}^2$. *Arrows indicate the direction of the partial order. The red triangle shows the convex hull of* $X := \{(-1, 1), (1, 0), (0, 1)\}$, *which contains the origin. I.e.* $(X, (0, 0))$ *is a minimal violating pair for* X.

- *There exists a minimal violating pair* (X, y) *for* S.

Proof. Suppose there exists a minimal violating pair (X, y) for S. Since $X \subseteq S$, we have $\mathsf{Conv}(X) \subseteq \mathsf{Conv}(S)$ and so $y \in \mathsf{Conv}(S)$. Thus, $y \notin S$ implies S is not convex. Now suppose S is not convex. Then there exists $y \in (\mathsf{Conv}(S) \cap \{0, \pm 1\}^n) \setminus S$. Let $X \subseteq S$ be a minimal set of points such that $y \in \mathsf{Conv}(X)$. The pair (X, y) is a minimal violating pair for S. $\square$

Fact 12.4.7. *Consider* $S, Q \subseteq \{0, \pm 1\}^n$. *If* Q *does not contain any* $X \cup \{y\}$ *such that* (X, y) *is a violating pair for* S, *then there exists a convex set* S' *such that* $S' \cap Q = S \cap Q$.

Proof. Let $S' = \mathsf{Conv}(S \cap Q)$ and consider an arbitrary $y \in Q$. We need to show that $y \in S$ if and only if $y \in S'$. Clearly, $y \in S$ implies $y \in S'$. Now suppose $y \in S'$ and note this implies $y \in \mathsf{Conv}(S \cap Q) \subseteq \mathsf{Conv}(S)$. Thus, if $y \notin S$, then $(S \cap Q, y)$ is a violating pair for S and this contradicts our assumption about Q. $\square$

The following corollary is crucial for proving our lower bounds in Section 15.2 and Section 15.3.

Corollary 12.4.8. *Let* T *be a convexity tester for sets* $S \subseteq \{0, \pm 1\}^n$ *with* 1*-sided error. Suppose* T *rejects a set* S *after querying a set* Q. *Then* Q *contains some* $X \cup \{y\}$ *such that* (X, y) *is a minimal violating pair for* S.

12.4.3 Concentration of Mass in the Ternary Hypercube

For $x \in \{0, \pm 1\}^n$, observe that $\|x\|_1 = \|x\|_2^2$ is precisely the number of non-zero coordinates of x. Moreover, each coordinate of a uniformly random x is non-zero with probability $2/3$, and so $\mathbb{E}_{x \in \{0,\pm 1\}^n}[\|x\|_1] = \frac{2n}{3}$. Standard concentration inequalities yield the following bound on the number of points $x \in \{0, \pm 1\}^n$ where $\|x\|_1$ is far from this expectation.

Fact 12.4.9. *For every $\tau \geq 0$,*

$$\mathbb{P}_{x \in \{0,\pm 1\}^n}\left[\left|\|x\|_1 - \frac{2n}{3}\right| > \tau\right] \leq 2\exp(-\tau^2/2n).$$

Proof. We have $\|x\|_1 = \sum_{i=1}^n X_i$ where $X_i = 1$ with probability $2/3$ and $X_i = 0$ with probability $1/3$. Thus, the bound follows immediately from Hoeffding's inequality. $\square$

Given $\tau \geq 0$, we use the following notation to denote the inner, middle, and outer layers of $\{0, \pm 1\}^n$ with respect to distance τ:

$$\begin{aligned}
\mathsf{Inn}(\tau) &:= \left\{x \colon \|x\|_1 - \frac{2n}{3} < -\tau\right\}, \\
\mathsf{Mid}(\tau) &:= \left\{x \colon \left|\|x\|_1 - \frac{2n}{3}\right| \leq \tau\right\}, \\
\mathsf{Out}(\tau) &:= \left\{x \colon \|x\|_1 - \frac{2n}{3} > \tau\right\}.
\end{aligned} \tag{12.3}$$

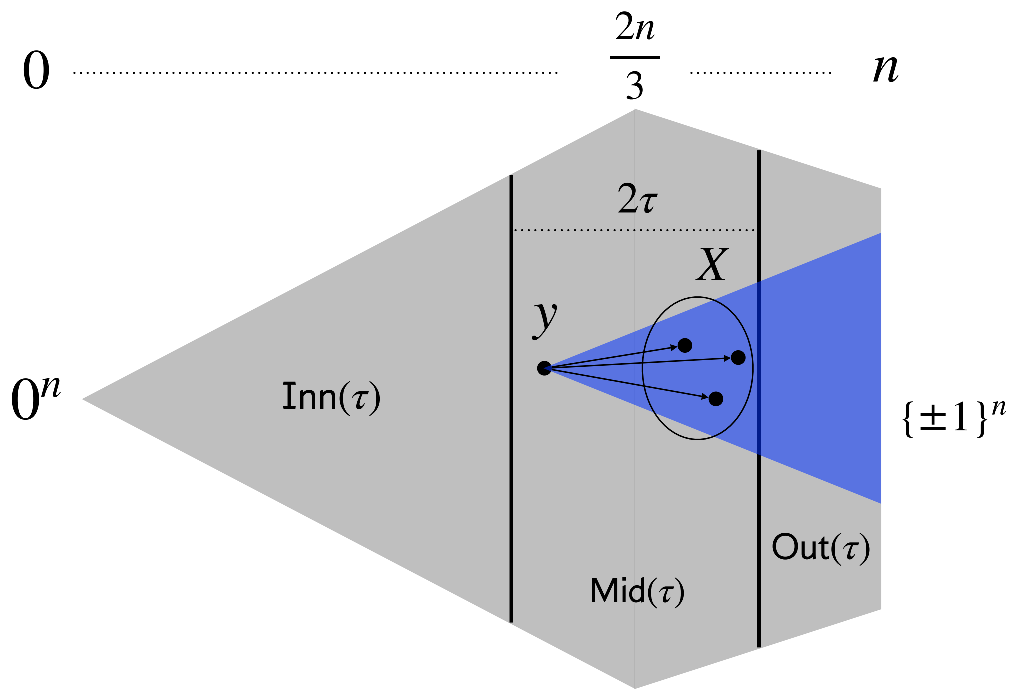


Figure 12.3: *This figure shows a pictorial representation of $\{0, \pm 1\}^n$ as a poset. Any vertical slice represents the set of all points with some fixed number of non-zero coordinates, and this number is increasing from left to right. The left-most point is the origin and the right-most points are the vertices of the hypercube $\{\pm 1\}^n$. The outward-oriented poset goes from left to right. The shaded blue region emanating from y is the set $\mathsf{Up}(y)$ of points above y in the partial order. The set X represents some minimal set of points for which $y \in \mathsf{Conv}(X)$ and thus $y \prec x$ for all $x \in X$, by Fact 12.4.5.*

CHAPTER 13

The Influence of Convex Sets

In this chapter we prove that the maximum edge boundary of convex sets in $\{0, \pm 1\}^n$ is $\widetilde{\Theta}(n^{3/4}) \cdot 3^n$, or equivalently that the influence is $\widetilde{\Theta}(n^{3/4})$.

13.1 Upper Bound

We prove that convex sets in the ternary hypercube have influence $\widetilde{O}(n^{3/4})$. The main idea in the proof is to relate the influence of a convex set S to the number of sign-changes in the maximum of a set of one-dimensional random walks. The proof will consider a random walk $\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(m)}$ starting from a random position in the middle layer of the ternary hypercube and moving randomly "outward" for $m = O\left(\sqrt{\frac{n}{\log n}}\right)$ steps, and count the number of influential edges crossed near the "middle layers" by relating them to one-dimensional random walks. We begin in Section 13.1.1 with definitions regarding the one-dimensional random walks that we require and then in Section 13.1.2 show how they relate to the number of influential edges of S; finally, in Section 13.1.3 we prove the necessary bound on the number of sign-changes of the one-dimensional random walks.

Notation. In this section it will be convenient to use bold letters like $\boldsymbol{X}$ for random variables, with the non-bold letter X being reserved for a fixed instantiation of $\boldsymbol{X}$.

13.1.1 One-Dimensional Random Walks and the Max-Walk

Let us define the types of one-dimensional random walks that will be necessary for our proof.

Definition 13.1.1 (Random Walks)**.** *Let* $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$*. Fix any permutation* $\sigma : [m] \to [m]$ *and sign vector* $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m) \in \{\pm 1\}^m$*. For any* $a \in \mathbb{R}$*, we define the function* $W_x^{+a}(t; \sigma, \varepsilon)$ *for* $t \in \{0\} \cup [m]$ *as*

$$W_x^{+a}(t; \sigma, \varepsilon) := \begin{cases} a & \text{if } t = 0 \\ a + \sum_{i=1}^{t} \varepsilon_i x_{\sigma(i)} & \text{if } t > 0\,. \end{cases}$$

The random walk $\boldsymbol{W}_x^{+a}$ *is defined by choosing a uniformly random permutation* $\boldsymbol{\sigma}$ *and vector* $\boldsymbol{\varepsilon} \sim \{\pm 1\}^m$ *and setting* $\boldsymbol{W}_x^{+a}(t) = W_x^{+a}(t; \boldsymbol{\sigma}, \boldsymbol{\varepsilon})$ *for every* t*. If* $a = 0$ *we drop the superscript.*

The main quantity of interest to us is the number of sign-changes of a random walk, defined as follows.

Definition 13.1.2 (Crossing Number)**.** *Let* $W : \{0\} \cup [m] \to \mathbb{R}$ *be any sequence. We define the* crossing number $C(W)$ *as the number of sign-changes of* W*, defined as the number of times* $t \in [m]$ *such that either* $W(t) \geq 0 > W(t-1)$ *or* $W(t) < 0 \leq W(t-1)$*.*

An important feature of our random walks will be that they have the Distinct Subset-Sum (DSS).

Definition 13.1.3 (DSS Random Walk)**.** *We say a sequence* $x \in \mathbb{R}^m$ *has the Distinct Subset-Sum (DSS) property if for every two disjoint subsets* $A, B \subseteq [m]$*, it holds that* $\sum_{a \in A} x_a \neq \sum_{b \in B} x_b$*. In particular, the random walk* $\boldsymbol{W}_x$ *satisfies*

$$\forall t \in [m]\,, \; \mathop{\mathbb{P}}_{\boldsymbol{\sigma}, \boldsymbol{\varepsilon}} \left[W_x(t; \boldsymbol{\sigma}, \boldsymbol{\varepsilon}) = 0 \right] = 0\,.$$

Note that, if x *has the DSS property, then so does any subsequence of* x*.*

We will require an upper bound on the crossing number of *max-walks*, which are random walks defined as the maximum of a set of *constituent* walks of the type defined above.

Definition 13.1.4 (Max-Walk)**.** *Let X be a set of sequences $x \in \mathbb{R}^m$, and let $a : X \to \mathbb{R}$. For a fixed permutation σ and vector $\varepsilon \in \{\pm 1\}^m$, define*

$$M_X^{+a}(t; \sigma, \varepsilon) := \max_{x \in X} W_x^{+a(x)}(t; \sigma, \varepsilon)\,,$$

and let the random walk $\boldsymbol{M}_X^{+a}$ be defined as

$$\boldsymbol{M}_X^{+a}(t) := M_X^{+a}(t; \boldsymbol{\sigma}, \boldsymbol{\varepsilon})$$

where $\boldsymbol{\sigma}, \boldsymbol{\varepsilon}$ are chosen uniformly at random.

The main fact about max-walks that we require is the following, which we prove in Section 13.1.3.

Lemma 13.1.5 (Max-Walk Crossing Number)**.** *Let X be a set of sequences $x \in \mathbb{R}^m$, each having the DSS property, and let $a : X \to \mathbb{R}$. Then*

$$\mathbb{E}\left[C(\boldsymbol{M}_X^{+a})\right] = O(\sqrt{m})\,.$$

13.1.2 Upper Bound on the Number of Influential Edges of a Convex Set

We now prove the following upper bound on the influence of any convex set in the ternary hypercube, restated below for convenience.

Theorem 12.1.4. *If $S \subseteq \{0, \pm 1\}^n$ is convex, then $\mathbb{I}(S) = O(n^{3/4} \log^{1/4} n)$.*

We require the following basic property of discrete convex sets.

Proposition 13.1.6. *Let $S \subseteq \{0, \pm 1\}^n$ be any discrete convex set. Then there is a finite set of vectors $V \subseteq \mathbb{R}^n$ and thresholds $\tau : V \to \mathbb{R}$, where each $v \in V$ defines a halfspace $H_v := \{x \in \{0, \pm 1\}^n \;:\; \langle v, x \rangle < \tau(v)\}$, such that $S = \bigcap_{v \in V} H_v$. One may also assume that V satisfies the property that, for every $v \in V$ and every two disjoint subsets $A, B \subseteq [n]$, $\sum_{i \in A} v_i \neq \sum_{j \in B} v_j$.*

Proof. Since S is the intersection of its convex hull $\mathsf{Conv}(S)$ with $\{0, \pm 1\}^n$, it may be written as the intersection of $\{0, \pm 1\}^n$ with a finite set of halfspaces with normal vectors V and thresholds $\tau : V \to \mathbb{R}$, and one may assume that none of the points in $\{0, \pm 1\}^n$ lie on the hyperplane boundary of any of the halfspaces. Then there is some $\delta > 0$ such that the minimum distance between a hyperplane and a point of $\{0, \pm 1\}^n$ is at least $\delta \cdot n$. For each $v \in V$, apply independent random perturbations to each coordinate to obtain $v'_i = v_i + r_i$ where r_i is drawn from $[-\delta, \delta]$ uniformly at random. With probability 1, the resulting set $V' = \{v' \ : \ v \in V\}$ satisfies the required conditions. $\square$

Proof of Theorem 12.1.4. Recall the definition of the edge-set E of the ternary cube from eq. (12.1) and the set $\mathsf{Mid}(\ell)$ from eq. (12.3). Given $\ell > 0$, let

$$E_\ell = \{(u, v) \in E \colon u, v \in \mathsf{Mid}(\ell)\}$$

denote the set of edges lying in the middle ℓ layers of $\{0, \pm 1\}^n$. We consider the following process which samples a random edge in $\{0, \pm 1\}^n$. Define $\ell := \sqrt{2n \log n}$ and $m := \sqrt{\frac{n}{\log n}}$. Let $\mathcal{D}$ denote the distribution over edges defined by the following procedure.

1. Sample $\boldsymbol{X}^{(0)} \sim \mathsf{Mid}(\ell)$.
2. Choose a random subset $\boldsymbol{T} \subseteq \{i \colon \boldsymbol{X}^{(0)}_i = 0\}$ with $|\boldsymbol{T}| = m$ of coordinates where $\boldsymbol{X}^{(0)}$ has a 0.
3. Let $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_m) \in \{\pm 1\}^m$ be independent Rademacher random variables and let $\boldsymbol{\sigma} \colon [m] \to \boldsymbol{T}$ be a random bijection.
4. For each $s \in [m]$, let $\boldsymbol{X}^{(s)} = \boldsymbol{X}^{(s-1)} + \boldsymbol{\varepsilon}_s e_{\boldsymbol{\sigma}(s)} = \boldsymbol{X}^{(0)} + \sum_{i=1}^{s} \boldsymbol{\varepsilon}_i e_{\boldsymbol{\sigma}(i)}$ where e_j is the unit vector with a 1 in coordinate j.
5. Choose $\boldsymbol{s} \sim [m]$ and return the edge $(\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{X}^{(s-1)}, \boldsymbol{X}^{(s)})$.

Note that the above process can be equivalently defined as obtaining $\boldsymbol{X}^{(s)}$ by selecting a uniform random coordinate $\boldsymbol{i}$ where $\boldsymbol{X}_{\boldsymbol{i}}^{(s-1)} = 0$ and flipping that bit to a random value in $\{\pm 1\}$, with equal probability. This results in a random walk $\boldsymbol{X}^{(0)}, \boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(m)}$ of length m where each $(\boldsymbol{X}^{(s-1)}, \boldsymbol{X}^{(s)})$ is a random out-going edge from $\boldsymbol{X}^{(s-1)}$. We use two main claims regarding this random walk to complete the proof of the theorem. The first is that choosing an edge $(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{D}$ is approximately the same as choosing a uniformly random edge from the middle layers.

Claim 13.1.7. *Fix any $z \in \mathsf{Mid}(\ell)$ and $s \in [m]$. Then $\mathbb{P}[\boldsymbol{X}^{(s)} = z] = \Theta(3^{-n})$. I.e., each step of the random walk is approximately uniformly distributed over $\mathsf{Mid}(\ell)$. As a corollary, for any fixed edge $(u,v) \in E_\ell$, we have*

$$\mathbb{P}_{(\boldsymbol{X},\boldsymbol{Y})\sim\mathcal{D}}[(\boldsymbol{X}, \boldsymbol{Y}) = (u,v)] = \Theta\left(\frac{1}{n \cdot 3^n}\right).$$

The second claim is that the probability of $(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{D}$ being an influential edge is small.

Claim 13.1.8. $\mathbb{P}_{(\boldsymbol{X},\boldsymbol{Y})\sim\mathcal{D}}[S(\boldsymbol{X}) \neq S(\boldsymbol{Y})] \leq O\left(\frac{1}{\sqrt{m}}\right)$.

We defer the proof of both claims to the end of the section. We now prove Theorem 12.1.4 using Claim 13.1.7 and Claim 13.1.8 as follows. Let E denote the edges of the ternary hypercube and let $E_\ell = \{(u,v) \in E \colon u, v \in \mathsf{Mid}(\ell)\}$. By definition,

$$\mathbb{I}(S) = \frac{1}{3^n} \cdot \Big(|\{(u,v) \in E \setminus E_\ell \colon S(u) \neq S(v)\}| + |\{(u,v) \in E_\ell \colon S(u) \neq S(v)\}|\Big).$$

The first term is bounded using Fact 12.4.9 as

$$\frac{|\{(u,v) \in E \setminus E_\ell \colon S(u) \neq S(v)\}|}{3^n} \leq \frac{|E \setminus E_\ell|}{3^n} \leq 2n \cdot \frac{|\overline{\mathsf{Mid}(\ell)}|}{3^n} \leq 2n \cdot 2\exp(-\ell^2/2n) = O(1)$$

since every vertex has degree at most $2n$. The second term is bounded as

$$\frac{|\{(u,v)\in E_\ell\colon S(u)\neq S(v)\}|}{3^n} = \frac{|E_\ell|}{3^n}\cdot \mathbb{P}_{(u,v)\sim E_\ell}[S(u)\neq S(v)] \leq \frac{2n}{3}\cdot \mathbb{P}_{(u,v)\sim E_\ell}[S(u)\neq S(v)]$$
$$\leq Ln\cdot \mathbb{P}_{(\boldsymbol{X},\boldsymbol{Y})\sim\mathcal{D}}[S(\boldsymbol{X})\neq S(\boldsymbol{Y})] \leq L'n\cdot m^{-1/2} = L'\cdot n^{3/4}\log^{1/4} n\,,$$

where L, L' are absolute constants. The first inequality follows simply from $E_\ell \subset E$ and $|E| = 2n\cdot 3^{n-1}$. The second inequality follows from Claim 13.1.7 and the third inequality follows from Claim 13.1.8. This completes the proof of Theorem 12.1.4. $\square$

Let us now complete the deferred proofs of Claim 13.1.7 and Claim 13.1.8.

Proof of Claim 13.1.7. Let $\|z\|_1 = \frac{2n}{3} + r$ where $|r| = O(\sqrt{n\log n})$. In order for $X^{(s)} = z$ to occur we must have $\left\|X^{(0)}\right\|_1 = \frac{2n}{3} + r - s$. Thus, the probability is

$$\mathbb{P}[X^{(s)} = z] = \frac{1}{3^n}\left(\binom{n}{\frac{2n}{3}+r-s}\cdot 2^{\frac{2n}{3}+r-s}\right)\cdot\left(\binom{n}{\frac{2n}{3}+r}\cdot 2^{\frac{2n}{3}+r}\right)^{-1}$$
$$= \frac{1}{3^n}\cdot\frac{1}{2^s}\cdot\binom{n}{\frac{2n}{3}+r-s}\binom{n}{\frac{2n}{3}+r}^{-1} = \Theta(3^{-n})$$

where the last step is due to the following fact:

If $|r| \leq O(\sqrt{n\log n})$ and $s = O(\sqrt{\frac{n}{\log n}})$, then $\binom{n}{\frac{2n}{3}+r-s}\binom{n}{\frac{2n}{3}+r}^{-1} = \Theta(2^s)$. As a corollary, the number of points in the ternary cube with hamming weight $\frac{2n}{3}+r-s$ and $\frac{2n}{3}+r$ differ by at most a constant factor.

This is proved as follows.

$$\frac{\binom{n}{\frac{2n}{3}+r-s}}{\binom{n}{\frac{2n}{3}+r}} = \frac{\left(\frac{2n}{3}+r\right)!\left(\frac{n}{3}-r\right)!}{\left(\frac{2n}{3}+r-s\right)!\left(\frac{n}{3}-r+s\right)!} = \prod_{p=0}^{s-1}\frac{\frac{2n}{3}+r-p}{\frac{n}{3}-r+s-p} = 2^s\cdot\prod_{p=0}^{s-1}\frac{\frac{n}{3}+\frac{r}{2}-\frac{p}{2}}{\frac{n}{3}-r+s-p}$$
$$= 2^s\cdot\prod_{p=0}^{s-1}\frac{\frac{n}{3}-r+s-p+\left(\frac{3r}{2}+\frac{p}{2}-s\right)}{\frac{n}{3}-r+s-p} = 2^s\cdot\prod_{p=0}^{s-1}\left(1+\frac{\frac{3r}{2}+\frac{p}{2}-s}{\frac{n}{3}-r+s-p}\right)$$

Observe that the numerator inside the product is $\pm O(\sqrt{n\log n})$ since r is the dominating term and the denominator is $\Omega(n)$ since $n/3$ is the dominating term. Therefore, we have

$$\frac{\binom{n}{\frac{2n}{3}+r-s}}{\binom{n}{\frac{2n}{3}+r}} = 2^s \cdot \left(1 \pm O\left(\sqrt{\frac{\log n}{n}}\right)\right)^s = \Theta(1) \cdot 2^s$$

since $s = O(\sqrt{\frac{n}{\log n}})$. □

Proof of Claim 13.1.8. Let S be an intersection of halfspaces $S = \bigcap_{v \in V} H_v$, with thresholds $\tau : V \to \mathbb{R}$, in the form promised by Proposition 13.1.6. In particular, each vector $v \in V$ has the DSS property (Definition 13.1.3). Fix any value of $\boldsymbol{X}^{(0)} = X^{(0)}$ and fix any permutation σ and sign-vector $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m) \in \{\pm 1\}^m$ in the definition of $\mathcal{D}$, and consider the resulting fixed values of $X^{(0)}, X^{(1)}, \dots, X^{(m)}$. Define $a : V \to \mathbb{R}$ as $a(v) = \langle v, X^{(0)} \rangle - \tau(v)$.

For each $v \in V$, consider the sequences $W_v^{+a(v)} := W_v^{+a(v)}(\cdot\,; \sigma, \varepsilon)$. For each $X^{(s)}$, observe that $X^{(s)} \in H_v$ if and only if $\langle v, X^{(s)} \rangle < \tau(v)$, which is equivalent to the condition $W_v(s) < 0$, since

$$\begin{aligned} W_v^{+a(v)}(s) = a(v) + \sum_{j=1}^{s} \varepsilon_j \cdot v_{\sigma(j)} &= \left(\sum_{i : X_i^{(0)} \neq 0} v_i X_i^{(0)} \right) - t(v) + \sum_{j=1}^{s} X_{\sigma(j)}^{(s)} v_{\sigma(j)} \\ &= \left(\sum_{j : X_j^{(s)} \neq 0} X_j^{(s)} v_j \right) - t(v) = \langle X^{(s)}, v \rangle - t(v) \,. \end{aligned}$$

Therefore $X^{(s)} \in S$ if and only if $W_v^{+a(v)}(s) < 0$ for all $v \in V$, which is equivalent to $M_V^{+a}(s) < 0$ where M_V^{+a} is the max-walk (recall Definition 13.1.4). Then for fixed sequence $X^{(0)}, \dots, X^{(m)}$ and uniformly random $\boldsymbol{s} \sim [m]$, the probability that $(X^{(\boldsymbol{s}-1)}, X^{(\boldsymbol{s})})$ is an influential edge is equal to $\frac{C(M_V^{+a})}{m}$. Therefore, taking $\boldsymbol{\sigma}$ and $\boldsymbol{\varepsilon}$ to be random, we have

$$\mathbb{P}_{(\boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{D}} \left[S(\boldsymbol{X}) \neq S(\boldsymbol{Y}) \right] = \frac{1}{m} \cdot \mathbb{E}\left[C(\boldsymbol{M}_V^{+\boldsymbol{a}}) \right] = O(\sqrt{m}) \,,$$

where the final bound is due to Lemma 13.1.5, since each vector in V was assumed to have the DSS property. This concludes the proof of the claim. $\square$

13.1.3 Crossing Bound for the Max-Walk: Proof of Lemma 13.1.5

We prove an upper bound on the number of times the maximum of a set of one-dimensional random walks can change sign. Let us define certain special events in a random walk.

Fix any walk time m and let $W : \{0\} \cup [m] \to \mathbb{R}$. We define:

- A *downcrossing* of W is a time $t \in [m]$ such that $W(t) < 0 \leq W(t-1)$. $C_{\downarrow}(W)$ is the number of downcrossings of W.
- An *upcrossing* of W is a time $t \in [m]$ such that $W(t) \geq 0 > W(t)$. $C_{\uparrow}(W)$ is the number of upcrossings of W.
- A *downwards level return* of W is any time t such that either:
  - If $W(0) \geq 0$ then the smallest time $t \in [m]$ such that $W(t) < W(0)$ is a downwards level return.
  - For any upcrossing s of W, the first time $t > s$ such that $W(t) < W(s)$ is a downwards level return.

  We write $L_{\downarrow}(W)$ for the number of downwards level returns of W.
- The *downwards level decrease* times of W is the unique sequence $s_1 < s_2 < \cdots$ defined inductively as follows.
  - If $W(0) \geq 0$ then s_1 is the first time such that $W(s_1) < W(0)$. Otherwise let t be the first upcrossing of W. Then s_1 is the first time such that $W(s_1) < W(t)$.
  - For $i > 1$, if $W(s_{i-1}) \geq 0$ then $s_i \in [m]$ is the smallest time such that $W(s_i) < W(s_{i-1})$. Otherwise, if $W(s_{i-1}) < 0$, then let t be the first upcrossing $t > s_{i-1}$ and define s_i as the first time $s_i > t$ such that $W(s_i) < W(t)$.

  We write $S_{\downarrow}(W)$ for the number of downwards level decreases of W.

- The *upwards level increase* times of W is the unique sequence $t_1 < t_2 < \cdots$ defined inductively as follows.
  - If $W(0) < 0$ then t_1 is the first time such that $W(t_1) > W(0)$. Otherwise let s be the first downcrossing of W. Then t_1 is the first time such that $W(t_1) > W(t)$.
  - For $i > 1$, if $W(t_{i-1}) < 0$ then t_i is the first time such that $W(t_i) > W(t_{i-1})$. Otherwise if $W(t_{i-1}) \geq 0$, then let s be the first downcrossing $s > t_{i-1}$ and define t_i as the first time $t_i > s$ such that $W(t_i) > W(s)$.

  We write $S_{\uparrow}(W)$ for the number of upwards level increases of W.

The main technical tool in our analysis is the following version of Sparre Andersen's fluctuation theorem [Spa54], as found in [BB23, Prop. 4.1]. Recall the definition of $\boldsymbol{W}_x$ from Definition 13.1.1 and the DSS property from Definition 13.1.3.

Theorem 13.1.9 (Sparre Anderson; see [BB23], Proposition 4.1)**.** *For every $m \in \mathbb{N}$, if $x \in \mathbb{R}^m$ has the DSS property, then the random walk $\boldsymbol{W}_x$ satisfies*

$$\mathbb{P}\left[\forall t \in [m] : \boldsymbol{W}_x(t) > 0\right] = g(m) := \frac{1}{4^m}\binom{2m}{m}.$$

We define a random variable $\boldsymbol{R}$ on the positive integers with

$$\forall t \in \mathbb{N}\,,\ \mathbb{P}\left[\boldsymbol{R} = t\right] := g(t-1) - g(t) = \frac{1}{4^{t-1}}\binom{2(t-1)}{t-1} - \frac{1}{4^t}\binom{2t}{t},$$

where we define $g(0) := 1$. For each $m \in \mathbb{N}$, we also define a random variable $\boldsymbol{Q}^{(m)}$ by the following process. Set $q = 0$ and $X = 0$; while $X < m$, increment q and set $X \leftarrow X + \boldsymbol{R}$ where $\boldsymbol{R}$ is a new independent copy of the random variable defined above. Then set $\boldsymbol{Q}^{(m)} = q$ once this process terminates; note that $\boldsymbol{Q}^{(0)} = 0$. Observe that for every $k \in \mathbb{N}$,

$$\mathbb{P}\left[\boldsymbol{Q}^{(m)} \geq k\right] = \mathbb{P}\left[\boldsymbol{R}_1 + \boldsymbol{R}_2 + \cdots + \boldsymbol{R}_k \leq m\right]$$

where each $\boldsymbol{R}_i$ is an independent copy of $\boldsymbol{R}$, and

$$\mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] = \sum_{t=1}^{m} \mathbb{P}[\boldsymbol{R} = t] \cdot \left(1 + \mathbb{E}\left[\boldsymbol{Q}^{(m-t)}\right]\right) .$$

The following holds due to Theorem 13.1.9.

Proposition 13.1.10. *Let $x \in \mathbb{R}^m$ have the DSS property, let $a \in \mathbb{R}$, and let $\boldsymbol{s}_1$, $\boldsymbol{t}_1$ denote the first downwards level decrease time and upwards level increase times of $\boldsymbol{W}_x^{+a}$, respectively. Then for all $z \in [m]$,*

1. *If $a \geq 0$ then $\mathbb{P}[\boldsymbol{s}_1 = z] = \mathbb{P}[\boldsymbol{R} = z]$; and,*
2. *If $a < 0$ then $\mathbb{P}[\boldsymbol{t}_1 = z] = \mathbb{P}[\boldsymbol{R} = z]$.*

Proposition 13.1.11. *Let $x \in \mathbb{R}^m$ have the DSS property and let $a \in \mathbb{R}$. Then*

$$\mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] = \mathbb{E}\left[S_{\downarrow}(\boldsymbol{W}_x^{+a}) + S_{\uparrow}(\boldsymbol{W}_x^{+a})\right] .$$

Proof. By induction on m. For $m = 1$ we have $\mathbb{E}\left[\boldsymbol{Q}^{(1)}\right] = \mathbb{P}[\boldsymbol{R} = 1] = 1/2$ and $\mathbb{E}\left[S_{\downarrow}(\boldsymbol{W}_x^{+a}) + S_{\uparrow}(\boldsymbol{W}_x^{+a})\right] = 1/2$ since the random walk has probability $1/2$ of increasing or decreasing in the first step; if $a \geq 0$ then the walk must decrease to create a downwards level decrease, while if $a > 0$ then the walk must increase to create an upwards level increase.

Now let $m > 1$. Suppose $a \geq 0$ without loss of generality. Then the first level increase or decrease is a downwards level decrease. Let $\boldsymbol{s}_1$ be the first downwards level decrease and let $\boldsymbol{y}$ denote the random subsequence of x that remains after removing the first $\boldsymbol{s}_1$ elements according to the random permutation $\boldsymbol{\sigma}$. Then by induction and Proposition 13.1.10,

$$\begin{aligned}
\mathbb{E}\left[S_{\downarrow}(\boldsymbol{W}_x^{+a}) + S_{\uparrow}(\boldsymbol{W}_x^{+a})\right] &= \sum_{t=1}^{m} \mathbb{P}[\boldsymbol{s}_1 = t] \cdot \left(1 + \mathbb{E}\left[S_{\downarrow}\left(\boldsymbol{W}_{\boldsymbol{y}}^{+\boldsymbol{W}_x(\boldsymbol{s}_1)}\right) + S_{\uparrow}\left(\boldsymbol{W}_{\boldsymbol{y}}^{+\boldsymbol{W}_x(\boldsymbol{s}_1)}\right) \;\middle|\; \boldsymbol{s}_1 = t\right]\right) \\
&= \sum_{t=1}^{m} \mathbb{P}[\boldsymbol{R} = t] \cdot \left(1 + \mathbb{E}\left[\boldsymbol{Q}^{(m-t)}\right]\right) = \mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] .
\end{aligned}$$

$\square$

For a sequence $W : \{0\} \cup [m] \to \mathbb{R}$, write $Z(W) = \sum_{t=1}^{m} \mathbb{1}\left[W(t) \in \{0, \pm 1\}\right]$.

Lemma 13.1.12. *For any m, $\mathbb{E}\left[Z(\boldsymbol{W}_{\vec{1}})\right] = O(\sqrt{m})$.*

Proof. We first bound the number of times t such that $\boldsymbol{W}_{\vec{1}}(t) = 0$. If t is odd then $\mathbb{P}[\boldsymbol{W}_{\vec{1}}(t) = 0] = 0$. If t is even then there is a universal constant C such that

$$\mathbb{P}[\boldsymbol{W}_{\vec{1}}(t) = 0] = \frac{1}{2^t}\binom{t}{t/2} \leq C \cdot \frac{1}{\sqrt{t}}\,.$$

Therefore the expected number of times t with $\boldsymbol{W}_{\vec{1}}(t) = 0$ is at most

$$\sum_{t \text{ even}} \mathbb{P}[\boldsymbol{W}_{\vec{1}}(t) = 0] \leq C \cdot \sum_{t=1}^{m} \frac{1}{\sqrt{t}} = O(\sqrt{m})\,.$$

Now observe that the expected number of times t where $\boldsymbol{W}_{\vec{1}}(t) = 1$ is the average of the expected number of times where the shifted walks $\boldsymbol{W}_{\vec{1}}^{+a}$ is 0 on domain $[m-1]$, where $a = \pm 1$, and the same holds for the number of times t where $\boldsymbol{W}_{\vec{1}}(t) = -1$. □

Proposition 13.1.13. *There exists $x \in \mathbb{R}^m$ with the DSS property such that*

$$\mathbb{E}\left[S_{\downarrow}(\boldsymbol{W}_x) + S_{\uparrow}(\boldsymbol{W}_x)\right] \leq \mathbb{E}\left[Z(\boldsymbol{W}_{\vec{1}})\right].$$

As a consequence, $\mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] = O(\sqrt{m})$.

Proof. Let $\delta := \frac{1}{3m}$. Let $\boldsymbol{x} := \vec{1} + \boldsymbol{z}$ where $\boldsymbol{z} \sim [-\delta, \delta]^m$ uniformly at random. Note that $\boldsymbol{x}$ has the DSS property with probability 1. For any fixed $z \in [-\delta, \delta]^m$, any permutation σ, and any $r \in \{\pm 1\}^n$, write $W_x(t) := W_x(t; \sigma, \varepsilon)$ for $x = \vec{1} + z$. Then we have $W_x(t; \sigma, \varepsilon) \in [W_{\vec{1}}(t; \sigma, \varepsilon) - 1/3, W_{\vec{1}}(t; \sigma, \varepsilon) + 1/3]$. Now fix any $z \in [-\delta, \delta]^m$ such that $x = \vec{1} + z$ has the DSS property; we show that it satisfies the required condition.

Let $s_1 < s_2 < \cdots < s_k$ be the downwards level decreasing or upwards level increasing points for W_x, let $s_0 = 0$, and observe that a point cannot be both downwards level decreasing and upwards level increasing. We show by induction on i that $|W_x(s_i)| \leq 1+1/3$ and therefore that $W_{\vec{1}}(s_i) \in \{0, \pm 1\}$. Since $W_x(0) = 0$ it must be that s_1 is downwards level decreasing and $W_x(s_1) < 0 = W_x(0) \le W_x(s_1 - 1)$ and therefore $W_x(s_1) \ge W_x(s_1 - 1) - 1 - 1/3 \ge -4/3$ so it must be that $W_{\vec{1}}(s_1) \in \{-1, 0\}$. For $i > 1$, suppose that s_i is a downwards level decreasing point. If there exists an upcrossing point $a > s_{i-1}$ such that $W_x(s_i) < W_x(a)$, then we observe that $W_x(a-1) < 0 \le W_x(a)$ and therefore $W_x(a) < 1 + 1/3$ so $W_{\vec{1}}(a) \in \{0, 1\}$. Now $W_x(s_i) < W_x(a) \le W_x(s_i - 1)$ so it must be that $-1 - 1/3 \le W_x(s_i) < 1 + 1/3$ so $W_{\vec{1}}(s_i) \in \{0, \pm 1\}$. On the other hand, if $W_x(s_{i-1}) \ge 0$ and $W_x(s_i) < W_x(s_{i-1})$ then by induction we have $W_{\vec{1}}(s_{i-1}) \in \{0, 1\}$, and also $W_x(s_i - 1) > W_x(s_{i-1})$, so again we have $-1 - 1/3 \le W_x(s_i) < 1 + 1/3$ and therefore $W_{\vec{1}}(s_i) \in \{0, \pm 1\}$. A similar argument holds for the upwards level increasing points.

The conclusion now follows from Proposition 13.1.11 and Lemma 13.1.12, since for the $x \in \mathbb{R}^m$ defined in the current proof,

$$\mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] = \mathbb{E}\left[S_\downarrow(\boldsymbol{W}_x) + S_\uparrow(\boldsymbol{W}_x)\right] \le \mathbb{E}\left[Z(\boldsymbol{W}_{\vec{1}})\right] = O(\sqrt{m})\,. \qquad \square$$

Proposition 13.1.14. *Let X be a set of sequences $x \in \mathbb{R}^m$ each having the DSS property, and let $a : X \to \mathbb{R}$ be arbitrary. Then*

$$\mathbb{E}\left[L_\downarrow(\boldsymbol{M}_X^{+a})\right] \le \mathbb{E}\left[\boldsymbol{Q}^{(m)}\right]\,.$$

Proof. By induction on m. For $m = 1$, the probability that $\boldsymbol{M}_X^{+a}$ has a downwards level return is at most $1/2$, because if $M_X^{+a}(0) \ge 0$, all of the maximizing constituent walks $x \in X$ satisfying $W_x(0)^{+a(x)} = M_X^{+a}(0)$ must decrease. If $M_X^{+a}(0) < 0$ then there is no downwards level return for $m = 1$.

Let $m > 1$ and consider two cases. First assume that $M_X^{+a}(0) \ge 0$ and let $x \in X$ be an arbitrary constituent walk satisfying $W_x^{+a(x)}(0) = M_X^{+a}(0)$. For fixed permutation σ and sign vector ε, let s_1 be the first downwards level return point of $M_X^{+a}(\cdot\,; \sigma, \varepsilon)$ and let s_1' be the first downwards level return point of $W_x^{+a(x)}(\cdot\,; \sigma, \varepsilon)$. Note that $s_1' \le s_1$ since M_X^{+a} is the maximum of its constituents.

Now we may write

$$\mathbb{E}\left[L_\downarrow(\boldsymbol{M}_X^{+a})\right] = \sum_{s=1}^{m} \mathbb{P}[\boldsymbol{s}_1 = s] \left(1 + \mathbb{E}\left[L_\downarrow\left(\boldsymbol{M}_{\boldsymbol{Y}}^{+\boldsymbol{b}}\right) \;\middle|\; \boldsymbol{s}_1 = s\right]\right) ,$$

where $\boldsymbol{Y}$ denotes the set of vectors X after removing the first t coordinates according to the random permutation $\boldsymbol{\sigma}$ and $\boldsymbol{b} : \boldsymbol{Y} \to \mathbb{R}$ is the starting point $\boldsymbol{b}(\boldsymbol{y}) = \boldsymbol{W}_x^{+a(x)}(t)$ of each walk $\boldsymbol{y} \in \boldsymbol{Y}$ obtained from the original vector $x \in X$ by removing the first t coordinates according to $\boldsymbol{\sigma}$. By induction, this is

$$\mathbb{E}\left[L_\downarrow\left(\boldsymbol{M}_{\boldsymbol{Y}}^{+\boldsymbol{b}}\right) \;\middle|\; \boldsymbol{s}_1 = s\right] \leq \mathbb{E}\left[\boldsymbol{Q}^{(m-s)}\right] .$$

Now we write $\boldsymbol{s}_1 = \boldsymbol{s}_1' + (\boldsymbol{s}_1 - \boldsymbol{s}_1')$ where the second term is non-negative. Then

$$\begin{aligned}
\mathbb{E}\left[L_\downarrow(\boldsymbol{M}_X^{+a})\right] &\leq \sum_{s=1}^{m} \mathbb{P}\left[\boldsymbol{s}_1' + (\boldsymbol{s}_1 - \boldsymbol{s}_1') = s\right] \cdot \left(1 + \mathbb{E}\left[\boldsymbol{Q}^{(m-s)}\right]\right) \\
&= \sum_{t=1}^{m} \mathbb{P}[\boldsymbol{s}_1' = t] \sum_{s=0}^{m-t} \mathbb{P}\left[\boldsymbol{s}_1 - \boldsymbol{s}_1' = s \mid \boldsymbol{s}_1' = t\right] \left(1 + \mathbb{E}\left[\boldsymbol{Q}^{(m-(s+t))}\right]\right) .
\end{aligned}$$

The inner sum is a convex sum of terms $1 + \mathbb{E}\left[\boldsymbol{Q}^{(m-(s+t))}\right]$ which are each bounded by $1+\mathbb{E}\left[\boldsymbol{Q}^{(m-t)}\right]$ because $\mathbb{E}\left[\boldsymbol{Q}^{(k)}\right] \geq \mathbb{E}\left[\boldsymbol{Q}^{(k')}\right]$ when $k \geq k'$. We also have $\mathbb{P}[\boldsymbol{s}_1' = t] = \mathbb{P}[\boldsymbol{R} = t]$ due to Proposition 13.1.10. Therefore,

$$\mathbb{E}\left[L_\downarrow(\boldsymbol{M}_X^{+a})\right] \leq \sum_{t=1}^{m} \mathbb{P}[\boldsymbol{R} = t] \left(1 + \mathbb{E}\left[\boldsymbol{Q}^{(m-t)}\right]\right) = \mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] .$$

We must now handle the case where $M_X^{+a}(0) < 0$. Let $\boldsymbol{t}$ be the smallest time where $\boldsymbol{M}_X^{+a}(\boldsymbol{t}) \geq 0$. Then

$$\mathbb{E}\left[L_\downarrow(\boldsymbol{M}_X^{+a})\right] = \sum_{t=1}^{m} \mathbb{P}[\boldsymbol{t} = t] \cdot \mathbb{E}\left[L_\downarrow\left(\boldsymbol{M}_{\boldsymbol{Y}}^{+\boldsymbol{b}}\right) \;\middle|\; \boldsymbol{t} = t\right] ,$$

where $\boldsymbol{Y}$ and $\boldsymbol{b}$ are defined similarly as before, as the sequences X after removing the first

t coordinates according to the random permutation $\boldsymbol{\sigma}$ and $\boldsymbol{b}(\boldsymbol{y}) = \boldsymbol{W}_x^{+a(x)}(t)$ is where the x walk ended up at time t. This new walk starts above 0 so the above argument applies and the conclusion holds. $\square$

We can now prove Lemma 13.1.5.

Proof of Lemma 13.1.5. First observe that $C(\boldsymbol{M}_X^{+a}) \leq 2\, C_\downarrow(\boldsymbol{M}_X^{+a}) + 1$ so it suffices to bound $C_\downarrow(\boldsymbol{M}_X^{+a})$. By definition it holds that $C_\downarrow(\boldsymbol{M}_X^{+a}) \leq L_\downarrow(\boldsymbol{M}_X^{+a})$, so by Proposition 13.1.14 and Proposition 13.1.13, we have

$$\mathbb{E}\left[C_\downarrow(\boldsymbol{M}_X^{+a})\right] \leq \mathbb{E}\left[\boldsymbol{Q}^{(m)}\right] = O(\sqrt{m})\,. \qquad \square$$

13.2 Lower Bound

Recall the definition of the *influence* of a set in the ternary hypercube given in eq. (12.2). In this section, we show that there exists a convex set whose influence is $\Omega(n^{3/4})$, nearly matching the upper bound given in Theorem 12.1.4.

The construction of the high-influence set is obtained by considering the intersection of $2^{\sqrt{n}}$ random halfspaces whose distance from the origin is $\Theta(n^{3/4})$. This approach is inspired by [Kan14, Theorem 2], who showed that in the Boolean hypercube, an intersection of k random halfspaces with an appropriately chosen distance from the origin will have expected influence $\Omega(\sqrt{n \log k})$. In the ternary hypercube, this type of argument still works as long as $k \leq 2^{O(\sqrt{n})}$. This type of construction was also used by Nazarov [Naz03] to show the existence of convex sets in $\mathbb{R}^n$ whose Gaussian surface area is $\Omega(n^{1/4})$, which matches the $O(n^{1/4})$ upper bound proven by Ball [Bal93].

13.2.1 A Convex Set with Large Influence

We prove the following theorem.

Theorem 12.1.3. *There exists a convex set $S \subseteq \{0, \pm 1\}^n$ with influence $\mathbb{I}(S) = \Omega(n^{3/4})$.*

Proof. Recall that the edges of $\{0, \pm 1\}^n$ are the directed pairs of points (x, y) such that there exists $i \in [n]$ for which $x_i = 0$, $y_i \in \{\pm 1\}$ and $x_j = y_j$ for all $j \neq i$. We will use the following claim which is also used by Kane (see [Kan14, Lemma 7] and its proof).

Claim 13.2.1. *For any n and $\varepsilon \in [2^{-n}, 1/2]$, there exists $\tau = \Theta(\sqrt{n \log 1/\varepsilon})$ such that*

$$\mathbb{P}_{x \sim \{\pm 1\}^n} \left[\sum_{i=1}^{n} x_i > \tau \right] \geq \varepsilon.$$

Let $\varepsilon = 2^{-\sqrt{n}}$ and choose $\tau = \Theta(\sqrt{n \log 1/\varepsilon}) = \Theta(n^{3/4})$ so that

$$\rho := \mathbb{P}_{z \sim \{\pm 1\}^{2n/3}} \left[\sum_{i} z_i > \tau \right] \geq \varepsilon \tag{13.1}$$

as guaranteed by Claim 13.2.1. The main technical lemma that allows our construction to work is the following, which we prove in Section 13.2.2. Note that this lemma crucially uses the assumption that $\tau = O(n^{3/4})$ and this is where the structure of the ternary hypercube prevents this construction from obtaining sets with influence $\gg n^{3/4}$.

Lemma 13.2.2. *For all n, all $\Omega(\sqrt{n}) \leq \tau \leq O(n^{3/4})$, and all $\ell = O(\sqrt{n})$,*

$$\underset{x \sim \{\pm 1\}^{n+\ell}}{\mathbb{P}} \left[\sum_{i=1}^{n+\ell} x_i > \tau \right] \leq O\left(\underset{x \sim \{\pm 1\}^{n}}{\mathbb{P}} \left[\sum_{i=1}^{n} x_i > \tau \right] \right).$$

By Lemma 13.2.2 there are constants $C_0 < 1 < C_1$ such that for all $\ell \in [-\sqrt{n}, \sqrt{n}]$, we have

$$C_0 \rho \leq \underset{z \sim \{\pm 1\}^{2n/3+\ell}}{\mathbb{P}} \left[\sum_{i} z_i > \tau \right] \leq C_1 \rho. \tag{13.2}$$

Define $H := \{x \in \{0, \pm 1\}^n \colon \sum_{i=1}^{n} x_i > \tau\}$. Let $L_m = \{x \in \{0, \pm 1\}^n \colon \|x\|_1 = m\}$ and note that

$$\mathop{\mathbb{P}}_{z\sim\{\pm 1\}^{2n/3+\ell}}\left[\sum_i z_i > \tau\right] = \mathop{\mathbb{P}}_{z\sim L_{2n/3+\ell}}\left[\sum_i z_i > \tau\right]$$

and so eq. (13.2) tells us that the density of H in $L_{2n/3+\ell}$ only differs by a constant multiplicative factor for any $\ell \in [-\sqrt{n}, \sqrt{n}]$. We abuse notation and write $H(x) = \mathbf{1}(x \in H)$. Now, let $E_{\sqrt{n}}$ denote the set of edges in $\{0, \pm 1\}^n$ which have both endpoints in $\mathsf{Mid}(\sqrt{n})$ and let

$$\mathbb{I}_{\mathsf{mid}}(H) := \frac{1}{3^n} \cdot |\{(x, y) \in E_{\sqrt{n}} \colon H(x) \neq H(y)\}| \tag{13.3}$$

denote the influence of H restricted to $\mathsf{Mid}(\sqrt{n})$. We prove the following lower bound on this quantity.

Claim 13.2.3. $\mathbb{I}_{\mathsf{mid}}(H) = \Omega(\rho \cdot \tau)$.

Proof. Let $\mathcal{I} = [2n/3 - \sqrt{n}, 2n/3 + \sqrt{n} - 1]$ and observe that we can write

$$\begin{aligned}
\mathbb{I}_{\mathsf{mid}}(H) &= \frac{1}{3}\sum_{i=1}^{n} \mathop{\mathbb{E}}_{x\sim\{0,\pm 1\}^n}\left[\mathbf{1}(\|x\|_1 \in \mathcal{I}) \cdot \left(|H(x^{i\leftarrow 1}) - H(x^{i\leftarrow 0})| + |H(x^{i\leftarrow 0}) - H(x^{i\leftarrow -1})|\right)\right] \\
&= \frac{1}{3}\sum_{i=1}^{n} \mathop{\mathbb{E}}_{x\sim\{0,\pm 1\}^n}\left[\mathbf{1}(\|x\|_1 \in \mathcal{I}) \cdot (H(x^{i\leftarrow 1}) - H(x^{i\leftarrow -1}))\right] \qquad (13.4) \\
&= \frac{1}{3}\mathop{\mathbb{E}}_{x\sim\{0,\pm 1\}^n}\left[\sum_{i=1}^{n} \mathbf{1}(\|x\|_1 \in \mathcal{I}) \left(H(x^{i\leftarrow 1}) - H(x^{i\leftarrow -1})\right)\right] \\
&= \frac{1}{3}\mathop{\mathbb{E}}_{z\sim\{0,\pm 1\}^n}\left[\sum_{i:z_i=1} \mathbf{1}(\left\|z^{i\leftarrow 0}\right\|_1)H(z) - \sum_{i:z_i=-1} \mathbf{1}(\left\|z^{i\leftarrow 0}\right\|_1)H(z)\right] \qquad (13.5)
\end{aligned}$$

where eq. (13.4) holds because H is a monotone function with respect to the standard partial order, i.e., if $x_i \leq y_i$ for all $i \in [n]$, then $x \in H$ implies $y \in H$, and eq. (13.5) follows by the observation that for each i, $z \in \{0, \pm 1\}^n$ appears in the sum as $H(z)$ whenever $x = z^{i\leftarrow 0}$ and $z_i = 1$, and appears in the sum as $-H(z)$ whenever $x = z^{i\leftarrow 0}$ and $z_i = -1$. Let

$\mathcal{I}' = [2n/3 - \sqrt{n} + 1, 2n/3 + \sqrt{n} - 1]$. Thus, the above expression can be rewritten as

$$\begin{aligned}\mathbb{I}_{\mathsf{mid}}(H) &= \frac{1}{3} \mathop{\mathbb{E}}_{z \sim \{0,\pm1\}^n} \left[H(z) \sum_{i=1}^{n} z_i \cdot \mathbf{1}(\left\| z^{i\leftarrow 0} \right\|_1 \in \mathcal{I}) \right] \\ &\geq \frac{1}{3} \mathop{\mathbb{E}}_{z \sim \{0,\pm1\}^n} \left[H(z) \mathbf{1}(\|z\|_1 \in \mathcal{I}') \sum_{i=1}^{n} z_i \right] > \frac{\tau}{3} \mathop{\mathbb{E}}_{z \sim \{0,\pm1\}^n} \left[H(z) \mathbf{1}(\|z\|_1 \in \mathcal{I}') \right] \end{aligned} \tag{13.6}$$

where the first inequality holds since $\mathcal{I}' \subset \mathcal{I}$ and $\|z\|_1 \in \mathcal{I}'$ implies $\|z^{i\leftarrow 0}\|_1 \in \mathcal{I}$ for all $i \in [n]$. The second inequality holds since $H(z) = 1$ if and only if $\sum_i z_i > \tau$. Finally,

$$\begin{aligned}\mathop{\mathbb{E}}_{z \sim \{0,\pm1\}^n} \left[H(z) \mathbf{1}(\|z\|_1 \in \mathcal{I}') \right] &= \frac{1}{3^n} \cdot \sum_{\ell=-\sqrt{n}+1}^{\sqrt{n}-1} \sum_{z \in L_{2n/3+\ell}} H(z) \\ &= \sum_{\ell=-\sqrt{n}+1}^{\sqrt{n}-1} \frac{\binom{n}{2n/3+\ell} \cdot 2^{2n/3+\ell}}{3^n} \cdot \mathop{\mathbb{P}}_{z \sim \{\pm1\}^{2n/3+\ell}} \left[\sum_i z_i > \tau \right] \end{aligned} \tag{13.7}$$

and the quantity in the RHS is $\Omega(\rho)$ by the lower bound in eq. (13.2) and since $\binom{n}{2n/3+\ell} \cdot 2^{2n/3+\ell} = \Omega(3^n/\sqrt{n})$ for all $\ell \in [-\sqrt{n}, \sqrt{n}]$ by an application of Stirling's approximation. Combining this with eq. (13.6), we conclude that $\mathbb{I}_{\mathsf{mid}}(H) = \Omega(\rho \cdot \tau)$ as claimed. □

Let $k := \max\{\lfloor (4C_1\rho)^{-1} \rfloor, 1\} \leq \varepsilon^{-1}$. Choose $v^{(1)}, \ldots, v^{(k)} \in \{\pm1\}^n$ i.i.d. uniformly at random and for each $i \in [k]$ define $H_i = \{x \in \{0,\pm1\}^n \colon \langle x, v^{(i)} \rangle > \tau\}$. Let $S = \cap_{i=1}^k \overline{H_i}$ be the convex set formed by the intersection of the complements of the H_i's. Note that $\overline{S} = \cup_{i=1}^k H_i$ and $\mathbb{I}(S) = \mathbb{I}(\overline{S})$ and thus it suffices to give a lower bound on $\mathbb{I}(\overline{S})$. Observe that every edge (x,y) that is influential for H_i is guaranteed to be influential for $\overline{S}$ if $x, y \notin H_j$ for all $j \neq i \in [k]$. Moreover, if $\|x\|_1 = \frac{2n}{3} + \ell$ where $\ell \in [-\sqrt{n}, \sqrt{n}]$, then

$$\mathop{\mathbb{P}}_{v^{(j)} \sim \{\pm1\}^n} [x \in H_j] = \mathop{\mathbb{P}}_{v^{(j)} \sim \{\pm1\}^n} \left[\langle v^{(j)}, x \rangle > \tau \right] = \mathop{\mathbb{P}}_{z \sim \{\pm1\}^{2n/3+\ell}} \left[\sum_i z_i > \tau \right] \leq C_1 \cdot \rho.$$

Thus, by a union bound, the probability that a H_i-influential edge (x,y) with $x, y \in \mathsf{Mid}(\sqrt{n})$

remains influential for $\overline{S}$ is at least $1-2(k-1)\cdot C_1\rho \geq 1-2\cdot(4C_1\rho)^{-1}\cdot C_1\rho \geq 1/2$. Therefore,

$$\mathop{\mathbb{E}}_{v^{(1)},\ldots,v^{(m)}}\left[\mathbb{I}(\overline{S})\right] \geq \frac{1}{2}\sum_{i=1}^{k}\mathbb{I}_{\mathsf{mid}}(H_i) = \Omega(k\cdot\rho\cdot\tau) = \Omega(\tau) = \Omega\left(n^{3/4}\right) \tag{13.8}$$

and this completes the proof. □

13.2.2 Bounding the Density Increment of Halfspaces

In this section we prove Lemma 13.2.2. Our proof makes crucial use of the following tight bound on the binomial coefficient $\binom{n}{\frac{n-\tau}{2}}$ for any $\tau = O(n^{3/4})$. Importantly, the bound is tight up to constant multiplicative factors, as opposed to constant factors in the exponent.

Fact 13.2.4. *If $\tau = O(n^{3/4})$, then $\binom{n}{\frac{n-\tau}{2}} = \Theta\left(\frac{2^n}{\sqrt{n}}\cdot\exp(-\frac{\tau^2}{2n})\right)$.*

Fact 13.2.4 is a special case of a much more general approximation, Corollary 13.3.2, specifically the case of $s = 2$. The proof is relatively tedious and so we relegate it to Section 13.3. Using Fact 13.2.4 we are able to prove the following claim which is important for the proof of Lemma 13.2.2.

Claim 13.2.5. *For all n, all $\Omega(\sqrt{n}) \leq \tau \leq O(n^{3/4})$, and all $\ell = O((n/\tau)^2)$,*

$$2^{-(n+\ell)}\binom{n+\ell}{\frac{n+\ell-\tau}{2}} \leq O\left(2^{-n}\binom{n}{\frac{n-\tau}{2}}\right).$$

Proof. Since $\tau = O(n^{3/4})$, by Fact 13.2.4

$$2^{-(n+\ell)}\binom{n+\ell}{\frac{n+\ell-\tau}{2}} \leq \frac{1}{\Theta(\sqrt{n+\ell})}\exp\left(-\frac{\tau^2}{2(n+\ell)}\right) = \frac{1}{\Theta(\sqrt{n})}\exp\left(-\frac{\tau^2}{2n(1+\ell/n)}\right).$$

Observe that

$$-\frac{\tau^2}{2n(1+\ell/n)} = -\frac{\tau^2}{2n}\left(1-\frac{\ell}{n+\ell}\right) = -\frac{\tau^2}{2n}+\frac{\tau^2\ell}{2n(n+\ell)} = -\frac{\tau^2}{2n}+O(1)$$

since $\ell = O((n/\tau)^2)$. Therefore,

$$2^{-(n+\ell)}\binom{n+\ell}{\frac{n+\ell-\tau}{2}} \le \frac{1}{\Theta(\sqrt{n})}\exp\left(-\frac{\tau^2}{2n}\right) \le O\left(2^{-n}\binom{n}{\frac{n-\tau}{2}}\right)$$

where the second inequality is by Fact 13.2.4 since $\tau = O(n^{3/4})$. □

We are now set up to prove Lemma 13.2.2.

Proof of Lemma 13.2.2. First, write

$$\mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n+\ell}}\left[\sum_{i=1}^{n+\ell} x_i > \tau\right] = \mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n+\ell}}\left[\tau < \sum_{i=1}^{n+\ell} x_i \le 2\tau\right] + \mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n+\ell}}\left[\sum_{i=1}^{n+\ell} x_i > 2\tau\right]. \quad (13.9)$$

By Hoeffding's inequality, the second term is

$$\begin{aligned}\mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n+\ell}}\left[\sum_{i=1}^{n+\ell} x_i > 2\tau\right] &\le \exp\left(-\frac{2\cdot(2\tau)^2}{4(n+\ell)}\right) \\ &= \exp\left(-\frac{2\tau^2}{n}\left(1-\frac{1}{(n/\ell)+1}\right)\right) = O\left(\exp\left(-\frac{2\tau^2}{n}\right)\right) \quad (13.10)\end{aligned}$$

since $\frac{\tau^2\ell}{n^2} = O(1)$. Using Claim 13.2.5, the first term is

$$\begin{aligned}\mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n+\ell}}\left[\tau < \sum_{i=1}^{n+\ell} x_i \le 2\tau\right] &= \frac{1}{2^{n+\ell}}\sum_{\tau'\in(\tau,2\tau]\text{ even}}\binom{n+\ell}{\frac{n+\ell-\tau'}{2}} \\ &\le O\left(\frac{1}{2^n}\sum_{\tau'\in(\tau,2\tau]\text{ even}}\binom{n}{\frac{n-\tau'}{2}}\right) \\ &= O\left(\mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n}}\left[\tau < \sum_{i=1}^{n} x_i \le 2\tau\right]\right). \quad (13.11)\end{aligned}$$

We now just need to show that the first term dominates the second term. By Fact 13.2.4

$$\begin{aligned}
\mathop{\mathbb{P}}_{x\sim\{\pm1\}^n}\left[\tau < \sum_{i=1}^n x_i \le 2\tau\right] &= \frac{1}{2^n}\sum_{\tau'\in(\tau,2\tau]\text{ even}}\binom{n}{\frac{n-\tau'}{2}} \\
&\ge \Omega\left(\frac{\tau}{\sqrt{n}}\exp\left(-\frac{2\tau^2}{n}\right)\right) = \Omega\left(\exp\left(-\frac{2\tau^2}{n}\right)\right). \qquad (13.12)
\end{aligned}$$

Plugging the bounds from eq. (13.10), eq. (13.11), and eq. (13.12) back into eq. (13.9) yields

$$\mathop{\mathbb{P}}_{x\sim\{\pm1\}^{n+\ell}}\left[\sum_{i=1}^{n+\ell} x_i > \tau\right] \le O\left(\mathop{\mathbb{P}}_{x\sim\{\pm1\}^n}\left[\tau < \sum_{i=1}^n x_i \le n^{3/4}\right]\right). \qquad \square$$

13.3 Approximating Binomial Coefficients Near the Middle

In this section we prove the following approximation of the binomial coefficient $\binom{n}{\frac{n-\tau}{2}}$.

Theorem 13.3.1. *For all $0 \le \tau \le n(1-\Omega(1))$, we have*

$$\binom{n}{\frac{n-\tau}{2}} = \frac{2^n\cdot\sqrt{\frac{2n}{\pi(n-\tau)(n+\tau)}}}{\exp\left(\tau\cdot\sum_{k=1}^\infty\left(\frac{\tau}{n}\right)^{2k-1}\left(\frac{1}{2k-1}-\frac{1}{2k}\right)+\Theta\left(\frac{1}{n}\right)\right)}.$$

Theorem 13.3.1 implies the following corollary.

Corollary 13.3.2. *For every constant integer $s \ge 1$, when $\tau = O(n^{1-\frac{1}{2s}})$ then*

$$\binom{n}{\frac{n-\tau}{2}} = \Theta\left(\frac{2^n}{\sqrt{n}}\cdot\exp\left(-\sum_{k=1}^{s-1}\frac{\tau^{2k}}{n^{2k-1}}\left(\frac{1}{2k-1}-\frac{1}{2k}\right)\right)\right).$$

Proof. Since $\tau = o(n)$, the square-root term in the numerator of Theorem 13.3.1 becomes

$\Theta(1/\sqrt{n})$. Since $\tau = O(n^{1-1/2s})$ we have

$$\begin{aligned}\sum_{k=s}^{\infty} \frac{\tau^{2k}}{n^{2k-1}} \left(\frac{1}{2k-1} - \frac{1}{2k}\right) &\le \sum_{k=s}^{\infty} n^{1-k/s} \left(\frac{1}{2k-1} - \frac{1}{2k}\right) \\ &= \sum_{k=0}^{\infty} \left(\frac{1}{n^{1/s}}\right)^k \left(\frac{1}{2(k+s)-1} - \frac{1}{2(k+s)}\right) = O(1).\end{aligned}$$

I.e., the infinite summation converges when one ignores the first $s-1$ terms. □

13.3.1 Proof of Theorem 13.3.1

We use the following standard identities and approximations in the proof.

Fact 13.3.3. *If $N > 1$, then*

$$\left(1 + \frac{1}{N}\right)^N = \exp\left(1 - \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{N^k(k+1)}\right) \text{ and } \left(1 - \frac{1}{N}\right)^N = \exp\left(-1 - \sum_{k=1}^{\infty} \frac{1}{N^k(k+1)}\right).$$

Proof. Consider the Taylor series expansions $\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}x^k}{k}$ and $\ln(1-x) = -\sum_{k=1}^{\infty} \frac{x^k}{k}$ at $x \in (0,1)$. Since $N > 1$, we have $1/N \in (0,1)$. Thus,

$$\left(1 + \frac{1}{N}\right)^N = \exp\left(N \ln\left(1 + \frac{1}{N}\right)\right) = \exp\left(N \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{N^k k}\right) = \exp\left(1 - \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{N^k(k+1)}\right)$$

and

$$\left(1 - \frac{1}{N}\right)^N = \exp\left(N \ln\left(1 - \frac{1}{N}\right)\right) = \exp\left(-N \sum_{k=1}^{\infty} \frac{1}{N^k k}\right) = \exp\left(-1 - \sum_{k=1}^{\infty} \frac{1}{N^k(k+1)}\right).$$

□

Fact 13.3.4 (Stirling's Approximation)**.** *For all $n \ge 1$*

$$\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} \le n! \le \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

Therefore, for all $n \geq 1$ and $1 \leq k \leq n-1$,

$$\binom{n}{k} = \sqrt{\frac{n}{2\pi k(n-k)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} \cdot \exp\left(\frac{1}{12n+\Theta(1)} - \frac{1}{12k+\Theta(1)} - \frac{1}{12(n-k)+\Theta(1)}\right).$$

We are now ready to complete the proof of the approximation for binomial coefficients.

Proof of Theorem 13.3.1. Letting $k = \frac{n-\tau}{2}$ and substituting this value in Fact 13.3.4 yields

$$\begin{aligned}\binom{n}{\frac{n-\tau}{2}} &= \sqrt{\frac{2n}{\pi(n-\tau)(n+\tau)}} \cdot \left(\frac{n}{\frac{n-\tau}{2}}\right)^{\frac{n-\tau}{2}} \left(\frac{n}{\frac{n+\tau}{2}}\right)^{\frac{n+\tau}{2}} \cdot \exp(\Theta(1/n)) \\ &= 2^n \sqrt{\frac{2n}{\pi(n-\tau)(n+\tau)}} \cdot \left(\frac{1}{1-\frac{\tau}{n}}\right)^{\frac{n-\tau}{2}} \left(\frac{1}{1+\frac{\tau}{n}}\right)^{\frac{n+\tau}{2}} \cdot \exp(\Theta(1/n)). \qquad (13.13)\end{aligned}$$

We now apply Fact 13.3.3 and get the following bounds:

$$\begin{aligned}\left(1-\frac{\tau}{n}\right)^{\frac{1}{2}(n-\tau)} &= \left(1-\frac{\tau}{n}\right)^{\frac{n}{\tau}\cdot\frac{\tau}{2n}(n-\tau)} \\ &= \exp\left(\left(-1-\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^k \frac{1}{k+1}\right)\cdot\left(\frac{\tau}{2}-\frac{\tau^2}{2n}\right)\right) \\ &= \exp\left(-\frac{\tau}{2}+\frac{\tau^2}{2n}\right)\exp\left(\left(\frac{\tau^2}{2n}-\frac{\tau}{2}\right)\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^k \frac{1}{k+1}\right) \qquad (13.14)\end{aligned}$$

and

$$\begin{aligned}\left(1+\frac{\tau}{n}\right)^{\frac{1}{2}(n+\tau)} &= \left(1+\frac{\tau}{n}\right)^{\frac{n}{\tau}\cdot\frac{\tau}{2n}(n+\tau)} \\ &= \exp\left(\left(1-\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^k \frac{(-1)^{k+1}}{k+1}\right)\cdot\left(\frac{\tau}{2}+\frac{\tau^2}{2n}\right)\right) \\ &= \exp\left(\frac{\tau}{2}+\frac{\tau^2}{2n}\right)\exp\left(\left(-\frac{\tau^2}{2n}-\frac{\tau}{2}\right)\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^k \frac{(-1)^{k+1}}{k+1}\right). \qquad (13.15)\end{aligned}$$

Taking the product of eq. (13.14) and eq. (13.15) gives

$$
\begin{aligned}
&\left(1-\frac{\tau}{n}\right)^{\frac{1}{2}(n-\tau)}\left(1+\frac{\tau}{n}\right)^{\frac{1}{2}(n+\tau)} \\
&=\exp\left(\frac{\tau^2}{n}\right)\exp\left(\frac{\tau^2}{2n}\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^k\frac{1-(-1)^{k+1}}{k+1}\right)\exp\left(-\frac{\tau}{2}\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^k\frac{1+(-1)^{k+1}}{k+1}\right).
\end{aligned}
$$

In the first sum, the kth term cancels when k is odd and doubles when k is even. In the second sum, the kth term cancels when k is even and doubles when k is odd. Thus,

$$
\begin{aligned}
\left(1-\frac{\tau}{n}\right)^{\frac{1}{2}(n-\tau)}\left(1+\frac{\tau}{n}\right)^{\frac{1}{2}(n+\tau)} &= \exp\left(\frac{\tau^2}{n}+\left(\frac{\tau^2}{n}\sum_{k=2,\text{ even}}^{\infty}\left(\frac{\tau}{n}\right)^k\frac{1}{k+1}\right)-\left(\tau\sum_{k=1,\text{ odd}}^{\infty}\left(\frac{\tau}{n}\right)^k\frac{1}{k+1}\right)\right) \\
&= \exp\left(\frac{\tau^2}{n}+\left(\tau\sum_{k=2,\text{ even}}^{\infty}\left(\frac{\tau}{n}\right)^{k+1}\frac{1}{k+1}\right)-\left(\tau\sum_{k=2,\text{ even}}^{\infty}\left(\frac{\tau}{n}\right)^{k-1}\frac{1}{k}\right)\right) \\
&= \exp\left(\frac{\tau^2}{n}+\left(\tau\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^{2k+1}\frac{1}{2k+1}\right)-\left(\tau\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^{2k-1}\frac{1}{2k}\right)\right) \\
&= \exp\left(\left(\tau\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^{2k-1}\frac{1}{2k-1}\right)-\left(\tau\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^{2k-1}\frac{1}{2k}\right)\right) \\
&= \exp\left(\tau\sum_{k=1}^{\infty}\left(\frac{\tau}{n}\right)^{2k-1}\left(\frac{1}{2k-1}-\frac{1}{2k}\right)\right) \qquad (13.16)
\end{aligned}
$$

and plugging this quantity back into eq. (13.13) completes the proof of Theorem 13.3.1. □

CHAPTER 14

Sample-Based Testing and Learning

In this chapter we prove upper and lower bounds for testing and learning convex sets on $\{0, \pm 1\}^n$ with samples.

14.1 Upper Bound

Theorem 12.1.5. *There is a uniform-distribution learning algorithm for convex sets in $\{0, \pm 1\}^n$ which achieves error at most ε with time and sample complexity $3^{\widetilde{O}(n^{3/4}/\varepsilon)}$. The $\widetilde{O}(\cdot)$ hides a factor of $\log^{1/4} n$.*

Our proof of Theorem 12.1.5 uses the standard approach of showing that when S is convex, its low-degree Fourier coefficients contain most of the information about S. We can then learn S by estimating its low-degree Fourier coefficients. This learning approach was established by Linial, Mansour, and Nisan, and is referred to as the "Low-Degree Algorithm" [LMN93]. The section is organized as follows:

1. Section 14.1.1: Setup of the Fourier analysis over the ternary hypercube that will be necessary for the learning result.

2. Section 14.1.2: Bounds on the Fourier concentration of convex sets, using the influence bounds from Chapter 13.

3. Section 14.1.3: The low-degree learning algorithm and the proof of Theorem 12.1.5 and Corollary 12.1.6.

14.1.1 Fourier Analysis Setup over the Ternary Hypercube

This subsection uses mostly standard techniques, following Chapter 8 of [O'D14] which outlines how to generalize Fourier analysis of Boolean functions to arbitary product spaces.

Let $\pi_{1/3}$ and $\pi_{1/3}^{\otimes n}$ denote the uniform distribution over $\{0, \pm 1\}$ and $\{0, \pm 1\}^n$, respectively. Let $L^2(\{0, \pm 1\}^n, \pi_{1/3}^{\otimes n})$ denote the real inner product space of functions $f \colon \{0, \pm 1\}^n \to \mathbb{R}$ with inner product $\langle f, g \rangle = \mathbb{E}_x[f(x)g(x)]$.

Definition 14.1.1 (Fourier Basis)**.** *A* Fourier basis *for $L^2(\{0, \pm 1\}, \pi_{1/3})$ is an orthonormal basis $\phi_{-1}, \phi_0, \phi_{+1} \colon \{0, \pm 1\}^n \to \mathbb{R}$ with $\phi_0 \equiv 1$.*

An important message in Chapter 8 of [O'D14] is that the specific choice of Fourier basis does not matter. For concreteness, we can use the following basis given in Example 8.10 of [O'D14].

Definition 14.1.2. *Define the following Fourier basis for $L^2(\{0, \pm 1\}, \pi_{1/3})$: $\phi_0 \equiv 1$,*

$$(\phi_{-1}(-1), \phi_{-1}(0), \phi_{-1}(1)) = (-\sqrt{6}/2, 0, \sqrt{6}/2), \text{ and}$$
$$(\phi_1(-1), \phi_1(0), \phi_1(1)) = (-\sqrt{2}/2, \sqrt{2}, -\sqrt{2}/2).$$

It can be easily confirmed that the basis in Definition 14.1.2 is orthonormal and so is indeed a Fourier basis. Now, given $\alpha \in \{0, \pm 1\}^n$, we define $\phi_\alpha \in L^2(\{0, \pm 1\}^n, \pi_{1/3}^{\otimes n})$ as

$$\phi_\alpha(x) := \prod_{i=1}^{n} \phi_{\alpha_i}(x_i) \tag{14.1}$$

An immediate corollary of Proposition 8.13 from [O'D14] is that $(\phi_\alpha)_{\alpha \in \{0, \pm 1\}^n}$ is a Fourier basis for $L^2(\{0, \pm 1\}^n, \pi_{1/3}^{\otimes n})$. I.e., $\phi_{(0,0,\ldots,0)} \equiv 1$ and this basis is orthonormal. Definition 8.14 of [O'D14] now asserts that every function $f \colon \{0, \pm 1\}^n \to \mathbb{R}$ can be written as

$$f(x) = \sum_{\alpha \in \{0, \pm 1\}^n} \widehat{f}(\alpha)\phi_\alpha(x) \text{ where } \widehat{f} = \langle f, \phi_\alpha \rangle = \mathbb{E}_x[f(x)\phi_\alpha(x)] \tag{14.2}$$

is the Fourier coefficient of f on α.

14.1.2 Fourier Concentration for Convex Sets

We use the notation $\#\alpha := |\{i\colon \alpha_i \neq 0\}|$. Our goal is now to prove the following fact about the Fourier coefficients of convex sets. Here we abuse notation and use $S\colon \{0, \pm 1\}^n \to \{\pm 1\}$ defined as $S(x) = (-1)^{\mathbf{1}(x \notin S)}$ to denote membership in the set S.

Lemma 14.1.3 (Fourier Concentration for Convex Sets)**.** *There exists a constant $C > 0$ such that for any convex set $S \subseteq \{0, \pm 1\}^n$ and $\varepsilon > 0$,*

$$\sum_{\alpha\colon \#\alpha > \frac{C}{\varepsilon} n^{3/4} \log^{1/4} n} \widehat{S}(\alpha)^2 \leq \varepsilon.$$

Proof. The idea is to make use of our upper bound on the influence of convex sets from Theorem 12.1.4. We will need the following slightly different definition of the influence given by [O'D14]. We will refer to this slightly different notion as *Fourier influence* and will denote it by $\mathbb{I}^{\mathsf{Fourier}}(f)$ for clarity. We will show that these definitions are equivalent up to a constant factor. Below, for $x \in \{0, \pm 1\}^n$, $i \in [n]$, and $b \in \{0, \pm 1\}$, we write $x^{i \leftarrow b}$ for the vector obtained from x by setting x_i to b.

Definition 14.1.4 (Def. 8.17 and 8.22, [O'D14])**.** *For $f \in L^2(\{0, \pm 1\}^n, \pi_{1/3}^{\otimes n})$ and $i \in [n]$, the* projection of f onto i *is*

$$E_i f(x) = \mathbb{E}_{b \in \{0, \pm 1\}}[f(x^{i \leftarrow b})].$$

The i'th coordinate Laplacian operator *L_i is defined as $L_i f := f - E_i f$. The* Fourier influence of coordinate i on f is *$\mathbb{I}_i^{\mathsf{Fourier}}(f) = \langle f, L_i f \rangle$. The* total Fourier influence of f *is $\mathbb{I}^{\mathsf{Fourier}}(f) = \sum_{i=1}^n \mathbb{I}_i^{\mathsf{Fourier}}(f)$.*

We will need the two following identities.

Proposition 14.1.5 (Prop. 8.16 and Prop. 8.23, [O'D14]). *Every $f \in L^2(\{0, \pm 1\}^n, \pi_{1/3}^{\otimes n})$ satisfies the following two identities:*

1. $\sum_{\alpha \in \{0,\pm 1\}^n} \widehat{f}(\alpha)^2 = \mathbb{E}[f^2]$
2. $\mathbb{I}^{\mathsf{Fourier}}(f) = \sum_{\alpha \in \{0,\pm 1\}^n} \#\alpha \cdot \widehat{f}(\alpha)^2$

Lemma 14.1.6 (Fourier Concentration from Influence). *Let $\mathcal{F}\colon \{0, \pm 1\}^n \to \{\pm 1\}$ be a class of functions with Fourier influence upper bounded by $\mathbb{I}^{\mathsf{Fourier}}(f) \le B$ for all $f \in \mathcal{F}$. Then, for any $\varepsilon > 0$, we have*

$$\sum_{\alpha \in \{0,\pm 1\}^n :\, \#\alpha > B/\varepsilon} \widehat{f}(\alpha)^2 \le \varepsilon.$$

Proof. By item (1) of Proposition 14.1.5, we have $\sum_\alpha f(\alpha)^2 = 1$. Thus, $\sum_\alpha \#\alpha \cdot \widehat{f}(\alpha)^2$ is the expectation of $\#\alpha$ when α is sampled with probability $\widehat{f}(\alpha)^2$. By item (2) of Proposition 14.1.5 we have

$$\sum_{\alpha \in \{0,\pm 1\}^n} \#\alpha \cdot \widehat{f}(\alpha)^2 = \mathbb{I}^{\mathsf{Fourier}}(f) \le B$$

and now applying Markov's inequality yields the desired inequality. □

Fact 14.1.7 (Equivalence of Influence Definitions). *For every $f\colon \{0, \pm 1\}^n \to \{\pm 1\}$, we have*

$$\frac{3}{8} \cdot \mathbb{I}^{\mathsf{Fourier}}(f) \le \mathbb{I}(f) \le \frac{3}{4} \cdot \mathbb{I}^{\mathsf{Fourier}}(f).$$

Proof. Let $\Delta_i(f)$ denote the number of lines in the ternary cube of the form $(x^{i\leftarrow -1}, x^{i\leftarrow 0}, x^{i\leftarrow 1})$ such that $f_{\{x^{i\leftarrow -1}, x^{i\leftarrow 0}, x^{i\leftarrow 1}\}}$ is not constant, and let $\Delta(f) = \sum_i \Delta_i$ be the total number of such lines. Recall the definition of $\mathbb{I}(f)$ from eq. (12.2) and observe that since every such line contains either 1 or 2 influential edges, we have $\Delta(f) \cdot 3^{-n} \le \mathbb{I}(f) \le 2\Delta(f) \cdot 3^{-n}$. We will show that $\mathbb{I}^{\mathsf{Fourier}}(f) = \frac{8}{3}\Delta(f) \cdot 3^{-n}$ and combining these observations completes the proof. We have

$$\mathbb{I}_i^{\mathsf{Fourier}}(f) = \langle f, \mathrm{L}_i f \rangle = \mathbb{E}_x \left[f(x)(f(x) - \mathrm{E}_i f(x)) \right] = \mathbb{E}_x \left[1 - f(x) \mathbb{E}_{b \in \{0,\pm 1\}} [f(x^{i \leftarrow b})] \right]. \quad (14.3)$$

Now, for a fixed x, consider the line in dimension i, containing x: $\ell_i(x) := (x^{i\leftarrow -1}, x^{i\leftarrow 0}, x^{i\leftarrow 1})$. Observe that if f is constant on $\ell_i(x)$, then $1 - f(x)\mathbb{E}_{b\in\{0,\pm1\}}[f(x^{i\leftarrow b})] = 0$. If f is non-constant on $\ell_i(x)$, then either it contains two $+1$'s and one -1 or vice versa. In both cases we have

$$\mathbb{E}_{a\in\{0,\pm1\}}\left[1 - f(x^{i\leftarrow a})\mathbb{E}_{b\in\{0,\pm1\}}[f(x^{i\leftarrow b})]\right] = 1 - \mathbb{E}_{b\in\{0,\pm1\}}[f(x^{i\leftarrow b})]^2 = 8/9. \tag{14.4}$$

Therefore,

$$\begin{aligned}\mathbb{I}_i^{\mathsf{Fourier}}(f) &= \mathbb{E}_x\left[1 - f(x)\mathbb{E}_{b\in\{0,\pm1\}}[f(x^{i\leftarrow b})]\right] = \mathbb{E}_x\mathbb{E}_{a\in\{0,\pm1\}}\left[1 - f(x^{i\leftarrow a})\mathbb{E}_{b\in\{0,\pm1\}}[f(x^{i\leftarrow b})]\right]\\ &= \frac{8}{9}\mathbb{E}_x\left[\mathbf{1}(f|_{\ell_i(x)} \text{ is not constant})\right] = \frac{8}{3}\Delta_i(f)\cdot 3^{-n}\end{aligned}$$

and summing over all i completes the proof. □

Combining Theorem 12.1.4, Lemma 14.1.6, and Fact 14.1.7 completes the proof of Lemma 14.1.3. □

14.1.3 Low-Degree Learning Algorithm and Proof of Theorem 12.1.5

Recall that we are using the basis for the space of functions $f\colon \{0,\pm1\}^n \to \mathbb{R}$ given by

$$\phi_\alpha(x) := \prod_{i=1}^n \phi_{\alpha_i}(x_i) \quad \text{for every} \quad \alpha \in \{0,\pm1\}^n \tag{14.5}$$

where $\alpha_{-1}, \alpha_0, \alpha_1$ are a basis for the space of functions $f\colon \{0,\pm1\} \to \mathbb{R}$ defined as: $\phi_0 \equiv 1$,

$$\begin{aligned}&(\phi_{-1}(-1), \phi_{-1}(0), \phi_{-1}(1)) = (-\sqrt{6}/2, 0, \sqrt{6}/2), \text{ and}\\ &(\phi_1(-1), \phi_1(0), \phi_1(1)) = (-\sqrt{2}/2, \sqrt{2}, -\sqrt{2}/2).\end{aligned}$$

Recall that $\widehat{f}(\alpha) = \mathbb{E}_x[f(x)\phi_\alpha(x)]$. Our learning upper bound Theorem 12.1.5 follows immediately by combining Lemma 14.1.3 with the following theorem.

Theorem 14.1.8 (Low-Degree Algorithm over $\{0,\pm1\}^n$)**.** *Let* $\mathcal{F}\colon \{0,\pm1\}^n \to \{\pm1\}$ *be a* *class of functions such that for $\varepsilon > 0$ and $\tau = \tau(\varepsilon, n)$,*

$$\sum_{\alpha \in \{0,\pm 1\}^n \colon \#\alpha > \tau} \widehat{f}(\alpha) \leq \varepsilon.$$

Then $\mathcal{F}$ can be learned with time and sample complexity poly$(n^\tau, 1/\varepsilon)$.

Proof. Let $A := \{\alpha \in \{0, \pm 1\}^n \colon \#\alpha \leq \tau\}$. Note that $|A| = \sum_{\Delta=0}^{\tau} \binom{n}{\Delta} \cdot 2^\Delta = \text{poly}(n^\tau)$. We take s samples $x_1, \ldots, x_s \in \{0, \pm 1\}^n$ where s will be chosen later. For each α, we use the empirical estimate $Z_\alpha := \frac{1}{s} \sum_{i=1}^{s} f(x_i)\phi_\alpha(x_i)$ and return the hypothesis

$$h(x) = \text{sgn}\left(\sum_{\alpha \in A} Z_\alpha \phi_\alpha(x)\right). \tag{14.6}$$

Consider the event that $|Z_\alpha - \widehat{f}(\alpha)| \leq \sqrt{\varepsilon/|A|}$ for all α. We first show that this event occurs with high probability, and then show that if it occurs, then h is a good hypothesis.

Claim 14.1.9. *Set $s := \frac{3|A|^2}{\varepsilon} =$ poly$(n^\tau, 1/\varepsilon)$. Then*

$$\mathbb{P}_{x_1,\ldots,x_s}\left[|Z_\alpha - \widehat{f}(\alpha)| \leq \sqrt{\varepsilon/|A|} \textit{ for all } \alpha \in A\right].$$

Proof. Fix any $\alpha \in A$ and observe that $Z_\alpha = \frac{1}{s} \sum_{i=1}^{s} X_i$ where $X_1, \ldots, X_s$ are independent copies of $X = f(x)\phi_\alpha(x)$ for $x \sim \{0, \pm 1\}^n$ drawn uniformly at random. Note that in the setting of the Boolean hypercube when one uses the standard basis of parity functions, the random variable X always lies in $\{\pm 1\}$. In the ternary hypercube this is not the case and in fact $|\phi_\alpha(x)|$ can be exponentially large. For instance $\phi_{\vec{1}}(\vec{0}) = (\sqrt{2})^n$. However, since $\langle \phi_\alpha, \phi_\alpha \rangle = 1$ for all α, we're able to show that $\mathbf{Var}(X) \leq 1$ and this allows us to obtain good estimates for $\widehat{f}(\alpha)$. We have

$$\sigma^2 := \mathbf{Var}(X) = \mathbb{E}_x[(f(x)\phi_\alpha(x))^2] - \mathbb{E}_x[f(x)\phi_\alpha(x)]^2 = 1 - \widehat{f}(\alpha)^2$$

by definition of the Fourier coefficient $\widehat{f}(\alpha)$ and the fact that our basis is orthonormal and so in particular $\mathbb{E}_x[\phi_\alpha(x)^2] = \langle \phi_\alpha, \phi_\alpha \rangle = 1$. Therefore, $\mathbf{Var}(Z_\alpha) = \frac{\sigma^2}{s} = \frac{1-\widehat{f}(\alpha)^2}{s} \leq \frac{1}{s}$. Note also that $\mathbb{E}[Z_\alpha] = \widehat{f}(\alpha)$. Now, by Chebyshev's inequality, we have

$$\mathbb{P}\left[|Z_\alpha - \widehat{f}(\alpha)| \geq \frac{k}{\sqrt{s}}\right] \leq \mathbb{P}\left[|Z_\alpha - \widehat{f}(\alpha)| \geq k\sqrt{\mathbf{Var}(Z_\alpha)}\right] \leq \frac{1}{k^2}.$$

Setting $k = \sqrt{3|A|}$ and recalling $s = \frac{3|A|^2}{\varepsilon}$ yields

$$\mathbb{P}\left[|Z_\alpha - \widehat{f}(\alpha)| \geq \sqrt{\varepsilon/|A|}\right] = \mathbb{P}\left[|Z_\alpha - \widehat{f}(\alpha)| \geq \frac{k}{\sqrt{s}}\right] \leq \frac{1}{3|A|}$$

and taking a union bound over all $\alpha \in A$ completes the proof of the claim. $\square$

Claim 14.1.10. *Using the definition of h in eq. (14.6), if $|Z_\alpha - \widehat{f}(\alpha)| \leq \sqrt{\varepsilon/|A|}$ for all $\alpha \in A$, then $\mathbb{P}_{x\sim\{0,\pm1\}^n}[h(x) \neq f(x)] \leq \varepsilon$.*

Proof. First, observe that

$$\mathbb{P}_x\left[f(x) \neq h(x)\right] = \frac{1}{4}\mathbb{E}_x\left[(f(x) - h(x))^2\right]. \tag{14.7}$$

Now, if $f(x) \neq h(x)$, then $\left(f(x) - \sum_{\alpha\in A} Z_\alpha\phi_\alpha(x)\right)^2 \geq 1 = \frac{1}{4}(f(x) - h(x))^2$. Clearly if $f(x) = h(x)$, then this inequality also holds. Thus, for any $x \in \{0, \pm1\}^n$, this inequality holds. Combining this observation with eq. (14.7) yields

$$\mathbb{P}_x\left[f(x) \neq h(x)\right] \leq \mathbb{E}_x\left[\left(f(x) - \sum_{\alpha\in A} Z_\alpha\phi_\alpha(x)\right)^2\right]. \tag{14.8}$$

In the next calculation, for $\alpha \notin A$, let $Z_\alpha := 0$. Now, writing $f(x) = \sum_\alpha \widehat{f}(\alpha)\phi_\alpha(x)$, expanding the squared sum, applying linearity of expectation, and using the fact that

$\mathbb{E}_x[\phi_\alpha(x)\phi_{\alpha'}(x)] = \langle \phi_\alpha, \phi_{\alpha'} \rangle = 0$ for any $\alpha \neq \alpha'$, we get

$$\begin{aligned}
\mathbb{E}_x\left[\left(\sum_\alpha \phi_\alpha(x)\left(\widehat{f}(\alpha) - Z_\alpha\right)\right)^2\right] &= \mathbb{E}_x\left[\sum_{\alpha,\alpha'} \phi_\alpha(x)\phi_{\alpha'}(x)\left(\widehat{f}(\alpha) - Z_\alpha\right)\left(\widehat{f}(\alpha') - Z_{\alpha'}\right)\right] \\
&= \sum_{\alpha,\alpha'} \langle \phi_\alpha, \phi_{\alpha'} \rangle \left(\widehat{f}(\alpha) - Z_\alpha\right)\left(\widehat{f}(\alpha') - Z_{\alpha'}\right) \\
&= \sum_\alpha \left(\widehat{f}(\alpha) - Z_\alpha\right)^2. \qquad (14.9)
\end{aligned}$$

Finally, using eq. (14.8), eq. (14.9), and the fact that $|\widehat{f}(\alpha) - Z_\alpha| \leq \sqrt{\varepsilon/|A|}$ for $\alpha \in A$ and $\sum_{\alpha \notin A} \widehat{f}(\alpha)^2 \leq \varepsilon$, yields

$$\mathbb{P}_x\left[f(x) \neq h(x)\right] \leq \sum_\alpha \left(\widehat{f}(\alpha) - Z_\alpha\right)^2 = \sum_{\alpha \in A} \left(\widehat{f}(\alpha) - Z_\alpha\right)^2 + \sum_{\alpha \notin A} \widehat{f}(\alpha)^2 \leq |A| \cdot \frac{\varepsilon}{|A|} + \varepsilon = 2\varepsilon$$

and this completes the proof of the claim. □

Combining Claim 14.1.9 and Claim 14.1.10 completes the proof of Theorem 14.1.8. □

14.2 Lower Bound

In this section we prove the following lower bound on the *sample complexity* of convexity testing in the ternary hypercube.

Theorem 12.1.7. *For sufficiently small constant $\varepsilon > 0$, every sample-based convexity tester for sets in $\{0, \pm 1\}^n$ has sample complexity $3^{\Omega(\sqrt{n})}$.*

Our proof of Theorem 12.1.7 follows the standard approach of defining a pair of distributions $\mathcal{D}_{\mathtt{yes}}, \mathcal{D}_{\mathtt{no}}$ over subsets of $\{0, \pm 1\}^n$ such that the following hold:

- $\mathcal{D}_{\mathtt{yes}}$ is supported over convex sets.
- Sets drawn from $\mathcal{D}_{\mathtt{no}}$ are typically far from convex: $\mathbb{P}_{S \sim \mathcal{D}_{\mathtt{no}}}[\varepsilon(S) = \Omega(1)] = \Omega(1)$.

- The distributions over labeled examples from $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$ are close in TV-distance.

14.2.1 The Distributions $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$

Our construction uses a variant of random Talagrand DNFs adapted to the case of testing convexity in the ternary hypercube, $\{0, \pm 1\}^n$. In particular, our construction is inspired by the approach of [BB21] and [CWX17] to prove lower bounds for testing monotonicity of functions on the Boolean hypercube, $\{0, 1\}^n$.

Let $N = 3^{\sqrt{n}}$ and choose N *terms* $t^{(1)}, \dots, t^{(N)} \in \{0, \pm 1\}^n$ i.i.d. according the following distribution. For each $i \in [N]$:

1. Form a (multi)-set T_i by taking $\sqrt{n}$ independent uniform samples from $[n]$.
2. For each $a \in T_i$, set $t_a^{(i)} \in \{\pm 1\}$ uniformly at random. For each $a \notin T_i$, set $t_a^{(i)} = 0$.

Let $\boldsymbol{t} = (t^{(1)}, \dots, t^{(N)})$ denote the random sequence of terms. Recall the *outward-oriented poset* (Definition 12.4.1) over $\{0, \pm 1\}^n$. For each $i \in [N]$, let

$$U_i := \left\{x \in \mathsf{Mid}(\sqrt{n}) \colon x \succeq t^{(i)} \text{ and } x \not\succeq t^{(j)} \text{ for all } j \in [N] \setminus \{i\}\right\} \tag{14.10}$$

denote the set of points in the middle layers of the ternary hypercube which satisfy the i'th term, uniquely. Let $U = \cup_{i=1}^N U_i$ denote the set of points which satisfy a unique term.

Sets drawn from $\mathcal{D}_{\mathsf{yes}}$ are generated as follows. Choose a uniform random assignment $\boldsymbol{\phi} \colon [N] \to \{0, 1\}$. For every $x \in \mathsf{Mid}(\sqrt{n})$ define

$$S_{\boldsymbol{t},\boldsymbol{\phi}}(x) = \begin{cases} 1, & \text{if } \forall i \in [N],\ x \not\succeq t^{(i)} \\ 0, & \text{if } \exists i \neq j \in [N],\ x \succeq t^{(i)} \text{ and } x \succeq t^{(j)} \\ \boldsymbol{\phi}(i), & \text{if } x \in U_i. \end{cases}$$

Sets drawn from $\mathcal{D}_{\mathsf{no}}$ are generated as follows. Choose a uniform random function $\boldsymbol{r} \colon U \to$

$\{0,1\}$. For each $x \in \mathsf{Mid}(\sqrt{n})$ define

$$S_{\boldsymbol{t},\boldsymbol{r}}(x) = \begin{cases} 1, & \text{if } \forall i \in [N],\ x \not\succeq t^{(i)} \\ 0, & \text{if } \exists i \neq j \in [N],\ x \succeq t^{(i)} \text{ and } x \succeq t^{(j)} \\ \boldsymbol{r}(x), & \text{if } x \in U. \end{cases}$$

For $x \notin \mathsf{Mid}(\sqrt{n})$: if $x \in \mathsf{Inn}(\sqrt{n})$, then both the yes and no distributions assign value 1 and if $x \in \mathsf{Out}(\sqrt{n})$, then both the yes and no distributions assign value 0.

Theorem 12.1.7 follows immediately by combining the following three lemmas.

Lemma 14.2.1. *Every set in the support of $\mathcal{D}_{\mathsf{yes}}$ is convex.*

Proof. Let $S_{\boldsymbol{t},\boldsymbol{\phi}} \subseteq \{0,\pm 1\}^n$ be any set drawn from $\mathcal{D}_{\mathsf{yes}}$. We observe that $S_{\boldsymbol{t},\boldsymbol{\phi}}$ is non-increasing with respect to the *outward-oriented poset* (recall Definition 12.4.1). Suppose $y \notin S_{\boldsymbol{t},\boldsymbol{\phi}}$ and let $x \in \mathsf{Up}(y)$ (recall Definition 12.4.2). We have three cases depending on where y lies.

- $y \in \mathsf{Out}(\sqrt{n})$: in this case $x \in \mathsf{Out}(\sqrt{n})$ as well and so $x \notin S_{\boldsymbol{t},\boldsymbol{\phi}}$.
- $y \succeq t^{(i)}, t^{(j)}$ for two terms $i \neq j \in [N]$: in this case we have $x \succeq y \succeq t^{(i)}, t^{(j)}$ and so $x \notin S_{\boldsymbol{t},\boldsymbol{\phi}}$.
- $y \in U_i$ for some $i \in [N]$ and $\boldsymbol{\phi}(i) = 0$: in this case we have $x \succeq y \succeq t^{(i)}$ and so either (i) $x \in U_i$, (ii) there exists $j \neq i \in [N]$ for which $x \succeq t^{(j)}$, or (iii) $x \in \mathsf{Out}(\sqrt{n})$. In all cases $x \notin S_{\boldsymbol{t},\boldsymbol{\phi}}$.

Since $S_{\boldsymbol{t},\boldsymbol{\phi}}$ is non-increasing we have $\mathsf{Up}(y) \subset \overline{S_{\boldsymbol{t},\boldsymbol{\phi}}}$ and so by Fact 12.4.5 any minimal set of points X such that $y \in \mathsf{Conv}(X)$ satisfies $X \subset \overline{S_{\boldsymbol{t},\boldsymbol{\phi}}}$. Thus $S_{\boldsymbol{t},\boldsymbol{\phi}}$ is convex by Fact 12.4.6. $\square$

Lemma 14.2.2. *For $S_{\boldsymbol{t},\boldsymbol{r}} \sim \mathcal{D}_{\mathsf{no}}$, we have $\mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}[\varepsilon(S_{\boldsymbol{t},\boldsymbol{r}}) \geq \Omega(1)] \geq \Omega(1)$.*

We prove Lemma 14.2.2 in Section 14.2.2.

Lemma 14.2.3. *Given a collection of points $\boldsymbol{x} = (x_1, \ldots, x_s) \in (\{0, \pm 1\}^n)^s$ and a set $S \subseteq \{0, \pm 1\}^n$, let $(\boldsymbol{x}, S(\boldsymbol{x})) := ((x_1, S(x_1)), \ldots, (x_s, S(x_s)))$ denote the corresponding collection of labelled examples. Let $\mathcal{E}_{\mathsf{yes}}$ and $\mathcal{E}_{\mathsf{no}}$ denote the distributions over $(\boldsymbol{x}, S(\boldsymbol{x}))$ when $\boldsymbol{x}$ consists of s i.i.d. uniform samples and $S \sim \mathcal{D}_{yes}$ and $S \sim \mathcal{D}_{no}$, respectively. If $s \leq 3^{\sqrt{n}/3}$, then the total variation distance between $\mathcal{E}_{yes}$ and $\mathcal{E}_{no}$ is $o(1)$.*

We prove Lemma 14.2.3 in Section 14.2.3.

14.2.2 Sets Drawn from $\mathcal{D}_{\mathsf{no}}$ are Far from Convex: Proof of Lemma 14.2.2

Proof. Recall the definition of the set U in eq. (14.10). We prove Lemma 14.2.2 by showing that with constant probability over the terms $\boldsymbol{t}$ and the random function $\boldsymbol{r} \colon U \to \{0, 1\}$, there exists a collection L of $\Omega(3^n)$ disjoint co-linear triples (x, y, z) such that $x, z \in S_{\boldsymbol{t},\boldsymbol{r}}$, $y \notin S_{\boldsymbol{t},\boldsymbol{r}}$, and $y = \frac{1}{2}(x + z)$. The existence of such a set implies that $\varepsilon(S_{\boldsymbol{t},\boldsymbol{r}}) \geq \frac{1}{3}|L| \cdot 3^{-n} = \Omega(1)$ since the membership of at least one point from each of these triples would need to changed in order to make the set convex.

We first show that there is a large collection T of disjoint co-linear triples lying in $\mathsf{Mid}(\sqrt{n})$. Then, by Claim 14.2.6 and fact that each point in U is included in the set $S_{\boldsymbol{t},\boldsymbol{r}}$ with probability $1/2$, we can argue that with constant probability, a constant fraction of the triples in T will be violations of convexity.

Claim 14.2.4. *There exists a set T of $\Omega(3^n)$ disjoint co-linear triples (x, y, z) such that (a) $x, y, z \in \mathsf{Mid}(\sqrt{n})$, and (b) $y = \frac{1}{2}(x + z)$.*

Proof. Given $z \in \{0, \pm 1\}^{n-1}$ and $b \in \{0, \pm 1\}$, let $(b, z) \in \{0, \pm 1\}^n$ denote the point whose first coordinate is b and the rest of the coordinates are given by z. Consider the set of disjoint triples

$$T := \left\{ ((-1, z), (0, z), (+1, z)) \colon z \in \{0, \pm 1\}^{n-1} \text{ such that } \|z\|_1 \in \left[\frac{2n}{3} - \sqrt{n}, \frac{2n}{3} + \sqrt{n} - 1\right] \right\}.$$

Observe that every triple (x, y, z) is contained in $\mathsf{Mid}(\sqrt{n})$ and clearly $y = \frac{1}{2}(x+z)$. We use the following fact to lower bound $|T|$. This fact follows from an application of Stirling's approximation.

Fact 14.2.5. *For any N and $\ell \in [-O(\sqrt{N}), O(\sqrt{N})]$, we have $\binom{N}{\frac{2N}{3}+\ell} = \Theta\left(\frac{1}{\sqrt{N}} \cdot \frac{3^N}{2^{2N/3+\ell}}\right)$.*

By the above fact,

$$|T| = \sum_{\ell=-\sqrt{n}}^{\sqrt{n}-1} \binom{n-1}{\frac{2n}{3}+\ell} 2^{2n/3+\ell} = \sum_{\ell=-\sqrt{n}}^{\sqrt{n}-1} \Omega\left(\frac{1}{\sqrt{n-1}} \cdot \frac{3^{n-1}}{2^{2(n-1)/3+\ell}}\right) 2^{2n/3+\ell} = \Omega(3^n)$$

and this completes the proof of the claim. □

Let T denote the set of $\Omega(3^n)$ disjoint co-linear triples in $\mathsf{Mid}(\sqrt{n})$ given by Claim 14.2.4. We will need the following claim which shows that triples in T are contained in U with constant probability.

Claim 14.2.6. *For any $(x, y, z) \in T$, we have $\mathbb{P}_{\boldsymbol{t}}[x, y, z \in U] \geq \frac{1}{1{,}000{,}000}$.*

Proof. By definition of T we have $x, y, z \in \mathsf{Mid}(\sqrt{n})$ and $x_1 = +1$, $y_1 = 0$, $z_1 = -1$ and $x_j = y_j = z_j$ for all $j \in [2, n]$. Recall the distribution over the terms $\mathbf{t} = (t^{(1)}, \ldots, t^{(N)})$ defined in Section 14.2.1. Note that $\mathbb{P}_{t^{(i)}}[t^{(i)} \preceq y] = (\frac{\|y\|_1}{2n})^{\sqrt{n}}$ since $t^{(i)} \preceq y$ if and only if each of the $\sqrt{n}$ non-zero coordinates a of $t^{(i)}$ (a) is chosen as one of the non-zero coordinates of y which happens with probability $\|y\|/n$ and (b) $t_a^{(i)}$ is set to y_a which happens with probability $1/2$. Also note that for a term $t^{(i)}$, we have $t^{(i)} \prec y$ implies $t^{(i)} \prec x, z$. Therefore,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{t}}[x, y, z \in U_i] &= \mathbb{P}_{t^{(i)}}[t^{(i)} \preceq x, y, z] \cdot \prod_{j \neq i} \mathbb{P}_{t^{(j)}}[t^{(j)} \not\preceq x, y, z] \\
&= \mathbb{P}_{t^{(i)}}[t^{(i)} \preceq y] \cdot \prod_{j \neq i} \mathbb{P}_{t^{(j)}}[t^{(j)} \not\preceq x, z] \\
&= \mathbb{P}_{t^{(i)}}[t^{(i)} \preceq y] \cdot \prod_{j \neq i} \left(1 - \mathbb{P}_{t^{(j)}}[(t^{(i)} \preceq x) \vee (t^{(i)} \preceq z)]\right). \qquad (14.11)
\end{aligned}$$

The first term is lower bounded by

$$\begin{aligned}
\mathbb{P}_{t^{(i)}}[t^{(i)} \preceq y] &= \left(\frac{\|y\|_1}{2n}\right)^{\sqrt{n}} \geq \left(\frac{\frac{2n}{3} - \sqrt{n}}{2n}\right)^{\sqrt{n}} \\
&= \frac{1}{3^{\sqrt{n}}}\left(1 - \frac{3}{2\sqrt{n}}\right)^{\sqrt{n}} \geq \frac{e^{-3/2-o(1)}}{N} \geq \frac{1}{5N}. \qquad (14.12)
\end{aligned}$$

To lower bound the second term, observe that

$$\mathbb{P}_{t^{(i)}}[t^{(i)} \preceq x] = \left(\frac{\|x\|_1}{2n}\right)^{\sqrt{n}} \leq \left(\frac{\frac{2n}{3} + \sqrt{n}}{2n}\right)^{\sqrt{n}} = \frac{1}{3^{\sqrt{n}}}\left(1 + \frac{3}{2\sqrt{n}}\right)^{\sqrt{n}} \leq \frac{e^{3/2}}{N} \leq \frac{5}{N} \qquad (14.13)$$

and the same bound holds for the point z. Therefore, by a union bound $\mathbb{P}_{t^{(i)}}[(t^{(i)} \preceq x) \vee (t^{(i)} \preceq z)] \leq 10/N$. Plugging this bound along with eq. (14.12) into eq. (14.11) and summing over all $i \in [N]$ yields

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{t}}[x, y, z \in U] &= \sum_{i=1}^{N} \mathbb{P}_{t^{(i)}}[t^{(i)} \preceq y] \cdot \prod_{j \neq i} \left(1 - \mathbb{P}_{t^{(j)}}[(t^{(i)} \preceq x) \vee (t^{(i)} \preceq z)]\right) \\
&\geq N \cdot \frac{1}{5N} \cdot \left(1 - \frac{10}{N}\right)^{N}
\end{aligned}$$

which is at least $\frac{1}{1{,}000{,}000}$ and this completes the proof. $\square$

Now, for a set $S_{\boldsymbol{t},\boldsymbol{r}} \sim \mathcal{D}_{\mathsf{no}}$, let

$$T_{\mathsf{viol}} = \{(x, y, z) \colon x, z \in S_{\boldsymbol{t},\boldsymbol{r}} \text{ and } y \notin S_{\boldsymbol{t},\boldsymbol{r}}\}$$

denote the set of triples in T that are violations of convexity for $S_{\boldsymbol{t},\boldsymbol{r}}$. By definition of $\mathcal{D}_{\mathsf{no}}$ and using Claim 14.2.6, for any fixed $(x, y, z) \in T$, we have

$$\mathbb{P}[(x, y, z) \in T_{\mathsf{viol}}] = \mathbb{P}_{\boldsymbol{t}}[x, y, z \in U] \cdot \mathbb{P}_{\boldsymbol{r}}[\boldsymbol{r}(x) = \boldsymbol{r}(z) = 1, \boldsymbol{r}(y) = 0 \mid x, y, z \in U] \geq \frac{1}{8,000,000}.$$

Therefore, $\mathbb{E}_{\boldsymbol{t},\boldsymbol{r}}[|T \setminus T_{\mathsf{viol}}|] \le |T|(1 - \frac{1}{8{,}000{,}000})$ and so by Markov's inequality

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}\left[|T_{\mathsf{viol}}| \le \frac{|T|}{8,000,000^2}\right] &\le \mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}\left[|T \setminus T_{\mathsf{viol}}| \ge |T|\left(1 - \frac{1}{8,000,000^2}\right)\right] \\
&= \mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}\left[|T \setminus T_{\mathsf{viol}}| \ge |T|\left(1 - \frac{1}{8,000,000}\right)\left(1 + \frac{1}{8,000,000}\right)\right] \\
&\le \mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}\left[|T \setminus T_{\mathsf{viol}}| \ge \mathbb{E}_{\boldsymbol{t},\boldsymbol{r}}[|T \setminus T_{\mathsf{viol}}|]\left(1 + \frac{1}{8,000,000}\right)\right] \\
&\le \frac{1}{1 + \frac{1}{8{,}000{,}000}} = 1 - \frac{1}{8,000,001}.
\end{aligned}$$

Finally, since $|T| = \Omega(3^n)$, this gives us

$$\mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}\left[\varepsilon(S_{\boldsymbol{t},\boldsymbol{r}}) \ge \Omega(1)\right] \ge \mathbb{P}_{\boldsymbol{t},\boldsymbol{r}}\left[|T_{\mathsf{viol}}| \ge \frac{|T|}{8,000,000^2}\right] \ge \frac{1}{8,000,001}$$

and this completes the proof of Lemma 14.2.2. $\square$

14.2.3 $\mathcal{D}_{\mathsf{yes}}$ and $\mathcal{D}_{\mathsf{no}}$ are Hard to Distinguish: Proof of Lemma 14.2.3

Proof. Recall the definition of the set U_i in eq. (14.10). For $a \ne b \in [s]$, let E_{ab} denote the event that x_a and x_b belong to the same U_i for some $i \in [N]$. Observe that conditioned on $\overline{\vee_{a,b} E_{ab}}$, the distributions $\mathcal{E}_{\mathsf{yes}}$ and $\mathcal{E}_{\mathsf{no}}$ are identical.

Let $x, y \in \{0, \pm 1\}^n$ denote two independent uniform samples. We have

$$\begin{aligned}
\mathbb{P}[E_{ab}] &= \mathbb{P}_{x,y,\boldsymbol{t}}\left[\bigvee_{i=1}^{n}(x \in U_i \wedge y \in U_i)\right] \\
&= \sum_{i=1}^{n} \mathbb{P}_{x,y,\boldsymbol{t}}\left[x \in U_i \wedge y \in U_i\right] = \sum_{i=1}^{N} \mathbb{P}_{y,\boldsymbol{t}}[y \in U_i]^2 \qquad (14.14)
\end{aligned}$$

where the second equality holds since the U_i's are disjoint and the third equality holds by independence of x and y. Now, for a fixed $i \in [N]$, if $y \notin \mathsf{Mid}(\sqrt{n})$ observe that $\mathbb{P}_{\boldsymbol{t}}[y \in U_i] = 0$ and if $y \in \mathsf{Mid}(\sqrt{n})$, we have $\mathbb{P}_{\boldsymbol{t}}[y \in U_i] \le 5/N$ by eq. (14.13). Thus, $\mathbb{P}_{y,\boldsymbol{t}}[y \in U_i] \le 5/N$ and combining this with eq. (14.14) yields $\mathbb{P}[E_{ab}] \le 25/N$. Finally, by a union bound, we have

$$d_{\mathrm{TV}}(\mathcal{E}_{\mathsf{yes}}, \mathcal{E}_{\mathsf{no}}) \le \mathbb{P}_{\boldsymbol{x},\boldsymbol{t}}\left[\bigvee_{a \neq b \in [s]} E_{ab}\right] \le s^2 \cdot \frac{25}{N} = o(1)$$

since $N = 3^{\sqrt{n}} = \omega(s^2)$. $\square$

CHAPTER 15

Non-Adaptive Testing with One-Sided Error

In this chapter we prove upper and lower bounds for non-adaptive convexity testing with one-sided error in the ternary hypercube.

15.1 Non-Adaptive Upper Bound

We complete the proof of Theorem 12.1.9 in this section. The upper bound on the query complexity for testing convexity non-adaptively with one-sided error is achieved by Algorithm 6. (As a reminder, the notions of upward shadow $\mathsf{Up}(y)$ and middle layers $\mathsf{Mid}(\ell)$ in the algorithm are introduced in Definition 12.4.2 and eq. (12.3), respectively.)

Algorithm 6 Convexity tester for sets in $\{0, \pm 1\}^n$.

Input: A set $S \subseteq \{0, \pm 1\}^n$ and a parameter $\varepsilon \in (0, 1)$.
Set $\ell := \sqrt{2n \ln 8/\varepsilon}$ and repeat $\frac{4}{\varepsilon}$ times:

1. Query $y \in \{0, \pm 1\}^n$ uniformly at random.
2. If $y \in \overline{S} \cap \mathsf{Mid}(\ell)$, then query all points in $\mathsf{Up}(y) \cap \mathsf{Mid}(\ell)$.
3. If there exists $X \subseteq S \cap \mathsf{Up}(y) \cap \mathsf{Mid}(\ell)$ such that $y \in \mathsf{Conv}(X)$, then reject.

Accept.

The analysis of Algorithm 6 relies on the following lemma regarding sets that are far from convex.

Lemma 15.1.1. *Let* $S \subseteq \{0, \pm 1\}^n$ *and* $\varepsilon \leq \varepsilon(S)$*. If* $\ell = \sqrt{2n \ln 8/\varepsilon}$*, then*

$$|\mathsf{Conv}\big(S \cap \mathsf{Mid}(\ell)\big) \cap \big(\overline{S} \cap \mathsf{Mid}(\ell)\big)| \geq \frac{\varepsilon}{2} \cdot 3^n.$$

In words, there are at least $\frac{\varepsilon}{2} \cdot 3^n$ points in the middle layers $\mathsf{Mid}(\ell)$ that are not in S but that lie in the convex hull of the portion of S in the middle layers.

Proof. Let $T := \mathsf{Conv}(S \cap \mathsf{Mid}(\ell)) \cap \{0, \pm 1\}^n$. Clearly, T is convex and so

$$\varepsilon(S) \cdot 3^n \leq \Delta(S, T) = |T \cap \overline{S}| + |\overline{T} \cap S|.$$

Now observe that $S \cap \mathsf{Mid}(\ell) \subseteq T$ and so $\overline{T} \cap S \subseteq \overline{\mathsf{Mid}(\ell)}$. Thus, $|\overline{T} \cap S| \leq |\overline{\mathsf{Mid}(\ell)}|$. Next, we have

$$|T \cap \overline{S}| = |T \cap (\overline{S} \cap \mathsf{Mid}(\ell))| + |T \cap (\overline{S} \cap \overline{\mathsf{Mid}(\ell)})| \leq |T \cap (\overline{S} \cap \mathsf{Mid}(\ell))| + |\overline{\mathsf{Mid}(\ell)}|.$$

By Fact 12.4.9, $|\overline{\mathsf{Mid}(\ell)}| \leq 2\exp(-\ln(8/\varepsilon)) \cdot 3^n = \frac{\varepsilon}{4} \cdot 3^n$. Therefore, combining the above yields

$$|T \cap (\overline{S} \cap \mathsf{Mid}(\ell))| \geq \varepsilon(S) \cdot 3^n - 2|\overline{\mathsf{Mid}(\ell)}| \geq \left(\varepsilon(S) - \frac{\varepsilon}{2}\right) \cdot 3^n \geq \frac{\varepsilon(S)}{2} \cdot 3^n$$

where the last step holds since $\varepsilon \leq \varepsilon(S)$. ∎

We now prove the correctness of Algorithm 6. The tester always accepts when S is convex, since in this case $\mathsf{Conv}(S \cap \mathsf{Mid}(\ell)) \subseteq S$. Now suppose $\varepsilon(S) \geq \varepsilon$. If $y \in \mathsf{Conv}(S \cap \mathsf{Mid}(\ell)) \cap (\overline{S} \cap \mathsf{Mid}(\ell))$, then there exists some $X \subseteq S \cap \mathsf{Mid}(\ell)$ such that (X, y) is a minimal violating pair. Crucially, Fact 12.4.5 guarantees that $X \subseteq \mathsf{Up}(y)$. Thus, if the tester picks such a y in step (1), then it is guaranteed to reject S since step (2) queries all points in $\mathsf{Up}(y) \cap \mathsf{Mid}(\ell)$. Therefore, using Lemma 15.1.1, the probability that the tester rejects S

after one iteration of steps 1-3 is at least

$$\underset{y \in \{0, \pm 1\}^n}{\mathbb{P}} \left[y \in \mathsf{Conv}(S \cap \mathsf{Mid}(\ell)) \cap (\overline{S} \cap \mathsf{Mid}(\ell)) \right] \geq \varepsilon/2.$$

Thus, the tester rejects S with probability at least $1 - (1-\varepsilon/2)^{4/\varepsilon} \geq 2/3$.

We now bound the number of queries. I.e., we need to bound the size of $\mathsf{Up}(y) \cap \mathsf{Mid}(\ell)$ when $y \in \mathsf{Mid}(\ell)$. Note that each point in this set can be obtained by choosing a set of 2ℓ coordinates where y has a zero, and then flipping each of these coordinates to a value in $\{0, \pm 1\}$. Therefore, when $y \in \mathsf{Mid}(\ell)$, we have

$$|\mathsf{Up}(y) \cap \mathsf{Mid}(\ell)| \leq \binom{n}{2\ell} \cdot 3^{2\ell} \leq n^{4\ell} = n^{\sqrt{32n \ln 8/\varepsilon}}$$

and so the total number of queries made by the tester is at most $\frac{4}{\varepsilon} \cdot n^{\sqrt{32n \ln 8/\varepsilon}}$. This completes the proof of Theorem 12.1.9.

15.2 Non-Adaptive Lower Bound

We complete the proof of Theorem 12.1.10 establishing the lower bound on the query complexity of non-adaptive convexity testers with one-sided error in this section. The starting point for the lower bound is the notion of an *anti-slab* in $\{0, \pm 1\}^n$.

Definition 15.2.1 (Slab). *Fix $\tau \geq 1$ and $v \in \{0, \pm 1\}^n$. The (τ, v)-*slab *is defined as*

$$\mathsf{Slab}_{\tau,v} = \{x \in \{0, \pm 1\}^n \colon |\langle v, x \rangle| \leq \tau\}.$$

*We refer to $\overline{\mathsf{Slab}_{\tau,v}}$ as the (τ, v)-*anti-slab.

Note that a slab is an intersection of two parallel halfspaces and so an anti-slab is a union of two parallel and disjoint halfspaces. Anti-slabs are clearly non-convex, and the following claim establishes an important property of any certificate of non-convexity for the anti-slab.

In particular, it shows that if a set of queries contains a witness of non-convexity for the (τ, v)-anti-slab, then it must contain two points $x \in \overline{\mathsf{Slab}_{\tau,v}}$ and $y \in \mathsf{Slab}_{\tau,v}$ whose projections onto v are separated by at least distance τ.

Claim 15.2.2 (The Structure of Violating Pairs for Anti-slabs)**.** *Suppose (X, y) is a violating pair for the (τ, v)-anti-slab, $\overline{\mathsf{Slab}_{\tau,v}}$. Then there exists a point $x \in X$ for which $|\langle v, x-y\rangle| > \tau$.*

Proof. We have $y \in \mathsf{Conv}(X)$ and so $\sum_{x\in X} \lambda_x x = y$ where $\sum_{x\in X} \lambda_x = 1$. Moreover, we have $y \in \mathsf{Slab}_{\tau,v}$ and so

$$\sum_{x\in X} \lambda_x \langle v, x\rangle = \left\langle v, \sum_{x\in X} \lambda_x x \right\rangle = \langle v, y\rangle \in [-\tau, \tau]. \tag{15.1}$$

We also have $X \subseteq \overline{\mathsf{Slab}_{\tau,v}}$, which implies $|\langle v, x\rangle| > \tau$ for all $x \in X$. Therefore, by eq. (15.1) there clearly has to be some $x \in X$ where $\langle v, x\rangle$ is positive and some $x' \in X$ where $\langle v, x'\rangle$ is negative, for otherwise the LHS would be outside the interval $[-\tau, \tau]$. In particular, this implies $\langle v, x\rangle > \tau$ and $\langle v, x'\rangle < -\tau$ and so

$$\langle v, x'\rangle < -\tau \le \langle v, y\rangle \le \tau < \langle v, x\rangle.$$

Thus, if $\langle v, y\rangle \le 0$, then $|\langle v, x - y\rangle| > \tau$, and if $\langle v, y\rangle \ge 0$, then $|\langle v, x' - y\rangle| > \tau$. □

We now introduce our hard family of sets: truncated anti-slabs. (As a reminder, the sets $\mathsf{Inn}(t)$ and $\mathsf{Out}(t)$ are defined in eq. (12.3).)

Definition 15.2.3 (Truncated Anti-slab)**.** *Fix $\tau \ge 1$, $v \in \{0, \pm 1\}^n$, and $t \ge 1$. The t-truncated (τ, v)-anti-slab is defined as follows:*

$$\mathsf{TAS}_{\tau,v,t} = \left(\overline{\mathsf{Slab}_{\tau,v}} \cup \mathsf{Inn}(t)\right) \setminus \mathsf{Out}(t).$$

In particular, we fix $\tau = \sqrt{n}$, $t = 0.7\sqrt{n}$, and consider vectors $v \in \{0, \pm 1\}^n$ for which $\|v\|_1 = n/2$. Thus, henceforth we will drop the subscripts τ, t and abbreviate $\mathsf{TAS}_v :=$

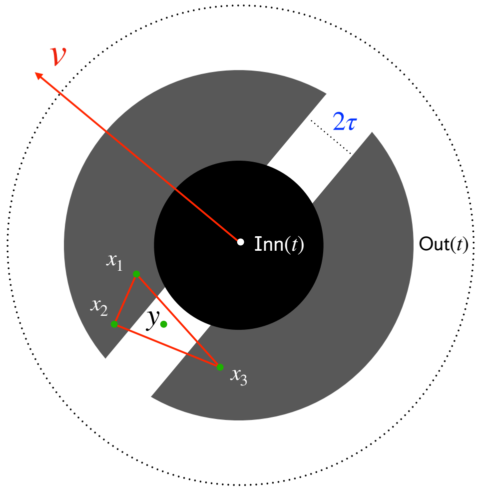


Figure 15.1: *An illustration of the t-truncated (τ, v)-anti-slab. The dotted circle represents $\{\pm 1\}^n$ and everything within it is $\{0, \pm 1\}^n$. The dark shaded regions are TAS_v. The pair $(\{x_1, x_2, x_3\}, y)$ is a violation for the set.*

$\mathsf{TAS}_{\sqrt{n}, v, 0.7\sqrt{n}}$.

In other words, TAS_v is the set obtained by taking the $(\sqrt{n}, v)$-anti-slab, adding in all points with fewer than $\frac{2}{3}n - 0.7\sqrt{n}$ non-zero entries, and removing all points with more than $\frac{2}{3}n + 0.7\sqrt{n}$ non-zero entries. The intuition for why these sets are hard to test (for non-adaptive testers with one-sided error) is as follows. Suppose a one-sided error tester T has queried a set $Q \subset \{0, \pm 1\}^n$ and rejects TAS_v. By Corollary 12.4.8, Q must contain a minimal violating pair (X, y) for TAS_v. Note that $X \subset \mathsf{TAS}_v$, $y \notin \mathsf{TAS}_v$, and $y \prec x$ for all $x \in X$ by Fact 12.4.5. By Claim 15.2.2, there is some $x \in X$ such that $|\langle v, x - y \rangle| > \sqrt{n}$. Additionally, by the truncation, we have $x \notin \mathsf{Out}(0.7\sqrt{n})$ and $y \notin \mathsf{Inn}(0.7\sqrt{n})$. Since $y \prec x$, this implies $\|x - y\|_1 \leq 1.4\sqrt{n}$. In summary, for T to reject TAS_v after querying Q, there must be some $y \prec x \in Q$ for which $|\langle v, x - y \rangle| > \sqrt{n}$, but also $\|x - y\|_1 \leq 1.4\sqrt{n}$.

We consider the family of sets $F = \{\mathsf{TAS}_v \colon \|v\|_1 = n/2\}$. By the above argument the lower bound boils down to the following question: given $y \prec x$ such that $\|x - y\|_1 \leq 1.4\sqrt{n}$, how many vectors $v \in \{0, \pm 1\}^n$ (with $\|v\|_1 = n/2$) exist for which $|\langle v, x - y \rangle| > \sqrt{n}$? We show that this number is upper bounded by $|F| \cdot \exp(-\Omega(\sqrt{n}))$ and so, by a union bound, the number of sets in F that T can reject after querying Q is bounded by $|Q|^2 \cdot |F| \cdot \exp(-\Omega(\sqrt{n}))$. Therefore, for T to be a valid non-adaptive tester with one-sided error, we must have $|Q|^2 = \exp(\Omega(\sqrt{n}))$ and this gives the result. This argument is formalized in Section 15.2.1.

Of course, for the above argument to prove Theorem 12.1.10, we need to show that truncated anti-slabs are $\Omega(1)$-far from convex.

Lemma 15.2.4. *Consider $v \in \{0, \pm 1\}^n$ where $\|v\|_2^2 = n/2$. We have $\mathsf{dist}(\mathsf{TAS}_v, \mathbf{convex}) = \Omega(1)$.*

The above Lemma 15.2.4 is a corollary of the following Lemma 15.2.5.

Lemma 15.2.5. *Consider $v \in \{0, \pm 1\}^n$ where $\|v\|_1 = n/2$. There exists a set $L \subset (\{0, \pm 1\}^n)^3$ of $\Omega(3^n)$ disjoint colinear triples such that for every $(x, y, z) \in L$ the following hold.*

1. $y = \frac{x+z}{2}$ *and* $y \in \mathsf{Slab}_{\sqrt{n},v}$, $x, z \in \overline{\mathsf{Slab}_{\sqrt{n},v}}$.
2. $x, y, z \in \mathsf{Mid}(0.7\sqrt{n})$.

In Section 15.2.1 we prove Theorem 12.1.10 using Claim 15.2.2 and Lemma 15.2.4. In Section 15.2.2 we prove Lemma 15.2.5, which immediately implies Lemma 15.2.4.

15.2.1 Proof of the Lower Bound

Recall the definition of $\mathsf{Inn}(t)$, $\mathsf{Mid}(t)$, and $\mathsf{Out}(t)$ in eq. (12.3). Given $v \in \{0, \pm 1\}^n$, recall that

$$\mathsf{TAS}_v = \left(\overline{\mathsf{Slab}_{\sqrt{n},v}} \cup \mathsf{Inn}(0.7\sqrt{n})\right) \setminus \mathsf{Out}(0.7\sqrt{n})$$

is the $0.7\sqrt{n}$-truncated $(\sqrt{n}, v)$-anti-slab (Definition 15.2.3). Let V denote the set of all vectors $v \in \{0, \pm 1\}^n$ where $\|v\|_2^2 = n/2$. By Lemma 15.2.4, we have $\mathsf{dist}(\mathsf{TAS}_v, \mathbf{convex}) = \Omega(1)$ for all $v \in V$. Also note that $|V| = \binom{n}{n/2} \cdot 2^{n/2} = 2^{3n/2}/\Theta(\sqrt{n})$.

Given $x, y \in \{0, \pm 1\}^n$, let $\Delta(x, y) = \{i \in [n] \colon x_i \neq y_i\}$. For $v \in \{0, \pm 1\}^n$, let $\mathsf{NZ}_v = \{i \colon v_i \neq 0\}$. Let T be a one-sided error, non-adaptive tester for convex sets in $\{0, \pm 1\}^n$.

Claim 15.2.6. *Fix $v \in V$ and suppose that T rejects TAS_v after querying a set $Q \subseteq \{0, \pm 1\}^n$. Then there exists $x \neq y \in Q$ such that (a) $|\Delta(x, y)| \leq 1.4\sqrt{n}$ and (b) $|\Delta(x, y) \cap \mathsf{NZ}_v| > \sqrt{n}$.*

Proof. By Corollary 12.4.8, Q must contain a minimal violating pair (X, y) for TAS_v. By Claim 15.2.2, there exists $x \in X$ for which $|\langle y - x, v\rangle| > \sqrt{n}$. Observe that $|\Delta(x, y) \cap \mathsf{NZ}_v| \geq |\langle y - x, v\rangle|$ and so (b) holds.

Now, since (X, y) is a violating pair we have $x \in \mathsf{TAS}_v$ and $y \notin \mathsf{TAS}_v$ and since (X, y) is minimal, Fact 12.4.5 implies that $y \prec x$. By construction, we have $\mathsf{Inn}(0.7\sqrt{n}) \subseteq \mathsf{TAS}_v$ and $\mathsf{Out}(0.7\sqrt{n}) \subseteq \overline{\mathsf{TAS}_v}$ and so it must be the case that $x, y \in \mathsf{Mid}(0.7\sqrt{n})$. In summary, x can be obtained by changing at most $1.4\sqrt{n}$ zero values in y to non-zero values. Thus, (a) holds. $\square$

Now, given $x, y \in \{0, \pm 1\}^n$, let

$$V(x, y) = \left\{v \in V \colon |\Delta(x, y) \cap \mathsf{NZ}_v| > \sqrt{n}\right\} \subseteq V. \tag{15.2}$$

Claim 15.2.7. *If $|\Delta(x, y)| \leq 1.4\sqrt{n}$, then $|V(x, y)| \leq O(|V|) \cdot 2^{-0.08\sqrt{n}}$.*

Proof. Note that $|\mathsf{NZ}_v| = n/2$ for all $v \in V$ and so $|V(x, y)|$ can be bounded as follows. In the following calculation, a vector $v \in V(x, y)$ is chosen by picking $\ell > \sqrt{n}$ coordinates from $\Delta(x, y)$ and $n/2 - \ell$ coordinates from $[n] \setminus \Delta(x, y)$ to be non-zero. Then each of these coordinates is fixed to a value in $\{\pm 1\}$.

$$
\begin{aligned}
|V(x,y)| &= \sum_{\ell > \sqrt{n}} \binom{|\Delta(x,y)|}{\ell} \binom{n - |\Delta(x,y)|}{n/2 - \ell} 2^{n/2} \\
&\leq 2^{n/2} \binom{n - |\Delta(x,y)|}{n/2 - \sqrt{n}} \sum_{\ell > \sqrt{n}} \binom{|\Delta(x,y)|}{\ell} \\
&= \frac{2^{3n/2}}{\Theta(\sqrt{n})} \cdot 2^{-|\Delta(x,y)|} \sum_{\ell > \sqrt{n}} \binom{|\Delta(x,y)|}{\ell} = O(|V|) \cdot 2^{-|\Delta(x,y)|} \sum_{\ell > \sqrt{n}} \binom{|\Delta(x,y)|}{\ell}
\end{aligned} \tag{15.3}
$$

To bound the RHS, observe that $2^{-k} \cdot \sum_{\ell > \sqrt{n}} \binom{k}{\ell}$ is precisely the probability that a random subset $S \subseteq [k]$ has size $|S| > \sqrt{n}$, which is a monotone increasing function of k. Thus, the RHS of eq. (15.3) is a monotone increasing function of $|\Delta(x,y)|$ and so is maximized by setting $\Delta(x,y) = 1.4\sqrt{n}$. Thus,

$$
\begin{aligned}
|V(x,y)| &\leq O(|V|) \cdot 2^{-1.4\sqrt{n}} \sum_{\ell > \sqrt{n}} \binom{1.4\sqrt{n}}{\ell} \\
&\leq O(|V|) \cdot 2^{-1.4\sqrt{n}} \cdot \sqrt{n} \cdot \binom{1.4\sqrt{n}}{\sqrt{n}} \leq O(|V|) \cdot 2^{-0.08\sqrt{n}}.
\end{aligned}
$$

The last inequality holds by the well known bound $\binom{m}{k} \leq \left(\frac{em}{k}\right)^k$ as follows. We have $\binom{1.4\sqrt{n}}{\sqrt{n}} = \binom{1.4\sqrt{n}}{0.4\sqrt{n}} \leq (\frac{e \cdot 1.4}{0.4})^{0.4\sqrt{n}} = 2^{0.4 \log_2(1.4e/0.4)\sqrt{n}} < 2^{1.31\sqrt{n}}$. $\square$

Now, given a set of queries $Q \subseteq \{0, \pm 1\}^n$, let

$$
V(Q) = \left\{ v \in V \colon \exists x \neq y \in Q \text{ such that } |\Delta(x,y) \cap \mathsf{NZ}_v| > \sqrt{n} \text{ and } |\Delta(x,y)| \leq 1.4\sqrt{n} \right\}. \tag{15.4}
$$

By Claim 15.2.6, if T rejects TAS_v after querying the set Q, then $v \in V(Q)$. Informally, $V(Q)$ contains all v for which Q can contain a witness of non-convexity for the set TAS_v.

Moreover, by Claim 15.2.7 and the union bound, we have

$$|V(Q)| \leq \sum_{x,y\in Q} |V(x,y)| \leq |Q|^2 \cdot O(|V|) \cdot 2^{-0.08\sqrt{n}}. \tag{15.5}$$

Now, let Q be the set of q queries sampled according to the distribution defined by the non-adaptive, one-sided error tester T. Then, using linearity of expectation and the bound from eq. (15.5) we obtain

$$\begin{aligned}\sum_{v\in V} \mathbb{P}_Q\left[T \text{ rejects } \mathsf{TAS}_v \text{ after querying } Q\right] &\leq \sum_{v\in V} \mathbb{P}_Q\left[v \in V(Q)\right] \\ &= \mathbb{E}_Q\left[|V(Q)|\right] \leq q^2 \cdot O(|V|) \cdot 2^{-0.08\sqrt{n}}\end{aligned}$$

and therefore, by averaging over V, there exists $v \in V$ such that

$$\frac{2}{3} \leq \mathbb{P}_Q\left[T \text{ rejects } \mathsf{TAS}_v \text{ after querying } Q\right] \leq O(1) \cdot q^2 \cdot 2^{-0.08\sqrt{n}} \tag{15.6}$$

where the first inequality is due to the fact that T rejects any S_v with probability at least $2/3$. Therefore, it follows that $q \geq \Omega(1) \cdot 2^{0.04\sqrt{n}}$.

15.2.2 Truncated Anti-slabs are Far from Convex

We complete the proof of Lemma 15.2.5 in this section, restated below for ease of reading.

Lemma 15.2.8. *Consider $v \in \{0, \pm 1\}^n$ where $\|v\|_1 = n/2$. There exists a set $L \subset (\{0, \pm 1\}^n)^3$ of $\Omega(3^n)$ disjoint colinear triples such that for every $(x, y, z) \in L$ the following hold.*

1. $y = \frac{x+z}{2}$ *and* $y \in \mathsf{Slab}_{\sqrt{n},v}$, $x, z \in \overline{\mathsf{Slab}_{\sqrt{n},v}}$.
2. $x, y, z \in \mathsf{Mid}(0.7\sqrt{n})$.

Proof. Let $J = \{j \in [n] \colon v_j \neq 0\}$. Without loss of generality, by a rotation, we may assume that $v_j = 1$ for all $j \in J$. Note that under this assumption, we have $\langle v, x\rangle = \sum_{j\in J} x_j$ for all $x \in \{0, \pm 1\}^n$.

To construct our set L of disjoint colinear triples we start by constructing a *matching* of $\Omega(3^n)$ pairs (x, y) such that (a) $y \in \mathsf{Slab}_{\sqrt{n},v} \cap \mathsf{Mid}(0.7\sqrt{n})$, (b) $x \in \overline{\mathsf{Slab}_{\sqrt{n},v}} \cap \mathsf{Mid}(0.7\sqrt{n})$, and (c) y can be obtained from x by changing a subset of x's $+1$ coordinates in J to 0. A third point z is obtained by reflecting x across y, i.e., this same set of coordinates is changed to -1 to obtain z. By symmetry we have $\|z\|_1 = \|x\|_1$, (x, y, z) are colinear, and the resulting set of triples are disjoint. We also choose the original matching so that we will always have $z \in \overline{\mathsf{Slab}_{\sqrt{n},v}} \cap \mathsf{Mid}(0.7\sqrt{n})$ and so the resulting triple satisfies item (1) and (2) of the lemma, i.e., it is a violation of convexity for the $0.7\sqrt{n}$-truncated $(\sqrt{n}, v)$-anti-slab, $\mathsf{TAS}_{\sqrt{n},v,0.7\sqrt{n}}$.

To construct our matching we use the following simple claim.

Claim 15.2.9. *Let (U, V, E) be a bipartite graph and $\Delta > 0$ be such that (a) each vertex $x \in U$ has degree exactly Δ and (b) each vertex $y \in V$ has degree at least Δ. Then there exists a matching $M \subseteq E$ in (U, V, E) of size $|M| \geq (1 - 1/e)|V|$.*

Proof. We construct a random map $\phi\colon U \to V$ as follows. For each $x \in U$ let $\phi(x)$ be a uniform random neighbor of x. Observe that $\phi^{-1}(y) \cap \phi^{-1}(y') = \emptyset$ for all $y \neq y' \in V$. Thus, given ϕ, we can obtain a matching M_ϕ as follows: for each $y \in V$, if $\phi^{-1}(y) \neq \emptyset$, then add (x, y) to M_ϕ for some arbitrary $x \in \phi^{-1}(y)$. To lower bound the size of M_ϕ, observe that

$$\mathbb{E}_\phi\left[|M_\phi|\right] = |V| - \mathbb{E}_\phi\left[\sum_{y\in V} \mathbf{1}(\phi^{-1}(y) = \emptyset)\right] = |V| - \sum_{y\in V} \mathop{\mathbb{P}}_{\phi}\left[\phi^{-1}(y) = \emptyset\right].$$

Now, if $\phi^{-1}(y) = \emptyset$, this means that all $\deg(y) \geq \Delta$ neighbors of y were mapped to some neighbor other than y, of which there are exactly Δ in total. Therefore,

$$\mathop{\mathbb{P}}_{\phi}\left[\phi^{-1}(y) = \emptyset\right] = \left(1 - \frac{1}{\Delta}\right)^{\deg(y)} \leq 1/e$$

since $\deg(y) \geq \Delta$. Thus, $\mathbb{E}_\phi\left[|M_\phi|\right] \geq |V| \cdot (1 - 1/e)$ and so there exists a matching M

satisfying the claim. $\square$

Given $x \in \{0, \pm 1\}^n$ and $b \in \{0, \pm 1\}$, let $|x|_{b,J} = |\{j \in J : x_j = b\}|$ and similarly for $\overline{J}$. Let $I = [n/6 + 0.6\sqrt{n}, n/6 + 0.8\sqrt{n}]$. We define the following sets.

$$X = \left\{ x \in \{0, \pm 1\}^n \colon \sqrt{n} < \sum_{j \in J} x_j < 1.2\sqrt{n},\ |x|_{1,J} \geq |x|_{0,J} + 1.1\sqrt{n}, \text{ and } |x|_{0,\overline{J}} \in I \right\} \tag{15.7}$$

$$Y = \left\{ y \in \{0, \pm 1\}^n \colon -0.1\sqrt{n} < \sum_{j \in J} y_j < 0.1\sqrt{n},\ |y|_{1,J} \geq |y|_{0,J} - 1.1\sqrt{n}, \text{ and } |y|_{0,\overline{J}} \in I \right\} \tag{15.8}$$

Observe that $X \subset \overline{\mathsf{Slab}_{\sqrt{n},v}}$ and $Y \subset \mathsf{Slab}_{\sqrt{n},v}$. We now partition X and Y as follows. For each $\ell \in \mathbb{N}$, let

$$X_\ell = \{x \in X \colon |x|_{0,J} = \ell\} \text{ and } Y_\ell = \left\{y \in Y \colon |y|_{0,J} = \ell + 1.1\sqrt{n}\right\}. \tag{15.9}$$

For each such ℓ we consider the bipartite graph (Y_ℓ, X_ℓ, E_ℓ) where there is an edge $(y, x) \in E_\ell$ if x can be obtained from y by choosing a set of $1.1\sqrt{n}$ coordinates from J where y has a 0 and flipping all of these bits to $+1$. Formally, $(y, x) \in E$ iff $\exists A \subseteq J$ of size $|A| = 1.1\sqrt{n}$ such that (a) for all $j \in A$, $y_j = 0$, $x_j = +1$, and (b) for all $j \in [n] \setminus A$, $y_j = x_j$. Observe now that (a) every vertex in Y_ℓ has degree exactly $\Delta := \binom{\ell + 1.1\sqrt{n}}{1.1\sqrt{n}}$, and (b) each vertex $x \in X_\ell$ has degree

$$\deg(x) = \binom{|x|_{1,J}}{1.1\sqrt{n}} \geq \binom{|x|_{0,J} + 1.1\sqrt{n}}{1.1\sqrt{n}} = \binom{\ell + 1.1\sqrt{n}}{1.1\sqrt{n}} = \Delta$$

where the inequality is by definition of X and the second to last equality is by definition of X_ℓ. Thus, by Claim 15.2.9, there exists a matching M_ℓ in (Y_ℓ, X_ℓ, E_ℓ) of size $|M_\ell| \geq \Omega(|X_\ell|)$.

Now, we obtain a set of disjoint colinear triples by taking

$$L = \left\{ (x, y, 2y - x) \colon (y, x) \in \bigcup_{\ell = \frac{n}{6} - 1.3\sqrt{n}}^{\frac{n}{6} - 1.2\sqrt{n}} M_\ell \right\}.$$

Note that by construction every $(x, y, z) \in L$ is a colinear triple in $\{0, \pm 1\}^n$.

Proof of items (1) and (2) of Lemma 15.2.5: By definition of the sets X and Y, we have $x \in \overline{\mathsf{Slab}_{\sqrt{n},v}}$ and $y \in \mathsf{Slab}_{\sqrt{n},v}$. Note that z is obtained from x by flipping a set of $1.1\sqrt{n}$ coordinates in J where x is $+1$ to -1. Therefore, we have $z \in \overline{\mathsf{Slab}_{\sqrt{n},v}}$ since

$$\sum_{j \in J} z_j = \sum_{j \in J} x_j - 2.2\sqrt{n} < 1.2\sqrt{n} - 2.2\sqrt{n} = -\sqrt{n}$$

where the inequality used the definition of the set X. Thus, item (1) of the lemma is satisfied.

Now, for every $(x, y, z) \in L$, we have

$$\frac{n}{6} - 1.3\sqrt{n} \leq |x|_{0,J} = |z|_{0,J} = |y|_{0,J} - 1.1\sqrt{n} \leq \frac{n}{6} - 1.2\sqrt{n}$$

and so

$$|x|_{0,J}, |z|_{0,J}, |y|_{0,J} \in \left[\frac{n}{6} - 1.3\sqrt{n}, \frac{n}{6} - 0.1\sqrt{n}\right].$$

Now, recalling that $I = [n/6 + 0.6\sqrt{n}, n/6 + 0.8\sqrt{n}]$ and the the definition of X and Y, we have

$$|y|_{0,\overline{J}}, |z|_{0,\overline{J}}, |x|_{0,\overline{J}} \in \left[\frac{n}{6} + 0.6\sqrt{n}, \frac{n}{6} + 0.8\sqrt{n}\right].$$

Combining the two bounds above we get that the number of 0-coordinates of x, y, and z are all in the range $[n/3 - 0.7\sqrt{n}, n/3 + 0.7\sqrt{n}]$. Therefore, we have $\|x\|_1, \|y\|_1, \|z\|_1 \in [2n/3 - 0.7\sqrt{n}, 2n/3 + 0.7\sqrt{n}]$, i.e., item (2) of the lemma is satisfied.

Proof that $|L| \geq \Omega(3^n)$: It remains to lower bound the size of L. Towards this, recall that

$$|L| = \sum_{r=1.2\sqrt{n}}^{1.3\sqrt{n}} |M_{n/6-r}| = \sum_{r=1.2\sqrt{n}}^{1.3\sqrt{n}} \Omega(|X_{n/6-r}|). \tag{15.10}$$

We use the following claim to simplify our calculation of $|X_{n/6-r}|$.

Claim 15.2.10. *If $|x|_{0,J} < n/6 - 1.2\sqrt{n}$ and $\sum_{j\in J} x_j > \sqrt{n}$, then $|x|_{1,J} \geq |x|_{0,J} + 1.1\sqrt{n}$.*

Proof. Note that $|x|_{1,J} - |x|_{-1,J} = \sum_{j\in J} x_j > \sqrt{n}$ and

$$|x|_{1,J} + |x|_{-1,J} = n/2 - |x|_{0,J} > n/3 + 1.2\sqrt{n} > 2|x|_{0,J} + 3.6\sqrt{n}.$$

Adding these inequalities and dividing by 2 yields $|x|_{1,J} > |x|_{0,J} + 1.85\sqrt{n} > |x|_{0,J} + 1.1\sqrt{n}$. □

In particular, recalling the definition of X in eq. (15.7), using Claim 15.2.10, we get that for $\ell \in [n/6 - 1.3\sqrt{n}, n/6 - 1.2\sqrt{n}]$, we can write

$$X_\ell = \left\{ x \in \{0, \pm 1\}^n \colon \sqrt{n} < \sum_{j\in J} x_j < 1.2\sqrt{n},\ |x|_{0,J} = \ell \text{ and } |x|_{0,\overline{J}} \in I \right\}.$$

I.e., the condition $|x|_{1,J} \geq |x|_{0,J} + 1.1\sqrt{n}$ in the definition of X is not needed to describe X_ℓ for the values of ℓ that we consider.

Claim 15.2.11. $\sum_{r=1.2\sqrt{n}}^{1.3\sqrt{n}} |X_{n/6-r}| = \Omega(3^n)$.

Proof. For simplicity let us assume that $\sqrt{n}$ is an integer. Note that $\sum_{r=1.2\sqrt{n}}^{1.3\sqrt{n}} |X_{n/6-r}|$ is equal to

$$\left(\sum_{0.6\sqrt{n} \leq q \leq 0.8\sqrt{n}} \binom{\frac{n}{2}}{\frac{n}{3} - q} 2^{\frac{n}{3} - q} \right) \left(\sum_{1.2\sqrt{n} \leq k \leq 1.3\sqrt{n}} \binom{\frac{n}{2}}{\frac{n}{3} + k} \sum_{0.5\sqrt{n} < s < 0.6\sqrt{n}} \binom{\frac{n}{3} + k}{\frac{n}{6} + \frac{k}{2} + s} \right). \tag{15.11}$$

The first term in the product in eq. (15.11) comes from the fact that the bits in $\overline{J}$ can be set to anything, as long as the number of zero bits is in the interval $I = [n/6+0.6\sqrt{n}, n/6+0.8\sqrt{n}]$. Equivalently, the number of non-zero entries is in the interval $[n/3-0.8\sqrt{n}, n/3-0.6\sqrt{n}]$.

Now consider the second term. The first sum is over the number of non-zero coordinates in J, which is in the interval $[\frac{n}{3}+1.2\sqrt{n}, \frac{n}{3}+1.3\sqrt{n}]$. The second sum is over all ways to set the non-zero coordinates in J so that they're sum is in the interval $(\sqrt{n}, 1.2\sqrt{n})$. Notice that if the number of non-zero coordinates is $\frac{n}{3}+k$, then the sum of the non-zero coordinates is in the interval $(\sqrt{n}, 1.2\sqrt{n})$ iff the number of $+1$'s is in the interval $(\frac{n}{6}+\frac{k}{2}+0.5\sqrt{n}, \frac{n}{6}+\frac{k}{2}+0.6\sqrt{n})$. This explains the second sum in the term.

To bound the RHS of eq. (15.11), we use the following fact, which follows readily from Stirling's formula.

Fact 15.2.12. *Let $N \in \mathbb{N}$, $t \in \mathbb{Z}$ be such that $|t| \le c\sqrt{N}$ for some constant $c > 0$. Then,*

$$\textit{(a)} \binom{N}{N/2+t} = \Theta\left(\frac{1}{\sqrt{N}} \cdot 2^N\right) \textit{ and (b)} \binom{N}{2N/3+t} = \Theta\left(\frac{1}{\sqrt{N}} \cdot \frac{3^N}{2^{2N/3+t}}\right).$$

By part (b) of Fact 15.2.12 we can bound the first term of eq. (15.11) as

$$\sum_{0.6\sqrt{n}\le q\le 0.8\sqrt{n}} \binom{\frac{n}{2}}{\frac{n}{3}-q} 2^{\frac{n}{3}-q} = \sum_{0.6\sqrt{n}\le q\le 0.8\sqrt{n}} \Omega\left(\frac{1}{\sqrt{n}} \cdot 3^{n/2}\right) = \Omega(3^{n/2}). \tag{15.12}$$

For the second term in eq. (15.11), we have $k, s = \Theta(\sqrt{n})$. Thus, by part (a) of Fact 15.2.12 we have

$$\binom{\frac{n}{3}+k}{\frac{n}{6}+\frac{k}{2}+s} \ge \Omega\left(\frac{1}{\sqrt{n}} \cdot 2^{\frac{n}{3}+k}\right) \implies \sum_{0.5\sqrt{n}<s<0.6\sqrt{n}} \binom{\frac{n}{3}+k}{\frac{n}{6}+\frac{k}{2}+s} \ge \Omega\left(2^{\frac{n}{3}+k}\right)$$

and so the second term of eq. (15.11) is

$$\sum_{1.2\sqrt{n}\le k\le 1.3\sqrt{n}} \binom{\frac{n}{2}}{\frac{n}{3}+k} \sum_{0.5\sqrt{n}<s<0.6\sqrt{n}} \binom{\frac{n}{3}+k}{\frac{n}{6}+\frac{k}{2}+s} \geq \sum_{1.2\sqrt{n}\le k\le 1.3\sqrt{n}} \binom{\frac{n}{2}}{\frac{n}{3}+k} \cdot \Omega\left(2^{\frac{n}{3}+k}\right) \tag{15.13}$$

which is at least $\Omega(3^{n/2})$ by part (b) of Fact 15.2.12 since $k = \Theta(\sqrt{n})$.

To summarize, the LHS in the claim statement is equal to the quantity in eq. (15.11), which is a product of two terms, each of which is at least $\Omega(3^{n/2})$ (eq. (15.12) and eq. (15.13)). Therefore, $\sum_{r=1.2\sqrt{n}}^{1.3\sqrt{n}} |X_{n/6-r}| = \Omega(3^n)$ as claimed. □

Using the bound from Claim 15.2.11 in eq. (15.10) finishes the proof of Lemma 15.2.5. □

15.3 Sample-Based One-Sided Error Lower Bound

There is an upper bound of $O(n3^n)$ samples required for exactly learning any set $S \subseteq \{0, \pm 1\}^n$, due to the coupon-collector argument, and therefore there is an upper bound of $3^{O(n)}$ on one-sided error testing of convex sets with samples. For large enough $\varepsilon > 0$, there is a slightly improved bound of $O(3^n \cdot \frac{1}{\varepsilon}\log(1/\varepsilon))$ for one-sided sample-based testers for any property of sets on $\{0, \pm 1\}^n$ (even in the distribution-free setting where the distribution over $\{0, \pm 1\}^n$ is arbitrary and unknown to the algorithm), due to the general upper bound of $O(\mathrm{VC}(\mathcal{H}) \cdot \frac{1}{\varepsilon}\log(1/\varepsilon))$ on one-sided sample-based testing, where $\mathrm{VC}(\mathcal{H})$ is the VC dimension of the property $\mathcal{H}$ [BFH21]. We show that the exponent $O(n)$ is optimal for one-sided sample-based testers.

Theorem 12.1.11. *For sufficiently small constant $\varepsilon > 0$, sample-based convexity testing in $\{0, \pm 1\}^n$ with one-sided error requires $3^{\Theta(n)}$ samples.*

Proof. It suffices to prove the lower bound, due to the discussion above. Suppose that T is a one-sided sample-based tester and let $Q \subseteq \{0, \pm 1\}^n$ denote a random set of s samples made by T. If T is given a non-convex set $S \subseteq \{0, \pm 1\}^n$, then it must reject S with probability at least $2/3$. Moreover, by Corollary 12.4.8, for T to reject S it must be that Q contains a minimal violating pair (X, y) for S and, by Fact 12.4.5, $X \subseteq \mathsf{Up}(y)$. Thus, in particular, there must exist two points $x, y \in Q$ such that $x \in \mathsf{Up}(y)$. Thus, by the union bound over all pairs in Q, we have

$$2/3 \leq \mathop{\mathbb{P}}_{Q}[T \text{ rejects } S] \leq \mathop{\mathbb{P}}_{Q}[\exists x, y \in Q \colon x \in \mathsf{Up}(y)] \leq s^2 \cdot \mathop{\mathbb{P}}_{x,y \in \{0,\pm 1\}^n}[x \in \mathsf{Up}(y)] \tag{15.14}$$

To compute this probability, notice that

$$x \in \mathsf{Up}(y) \text{ if and only if } \forall i \in [n] \colon (y_i = 0) \vee (x_i = y_i = 1) \vee (x_i = y_i = -1)$$

and so

$$\mathop{\mathbb{P}}_{x,y \sim \{0,\pm 1\}^n}[x \in \mathsf{Up}(y)] = (5/9)^n. \tag{15.15}$$

Thus, combining eq. (15.14) and eq. (15.15), we have $s \geq \sqrt{\frac{2}{3}(\frac{9}{5})^n} = 3^{\Omega(n)}$. $\square$

[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. 230

[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. Prelim. version in Proc., RANDOM 2004. 230

[Bal93] Keith Ball. The reverse isoperimetric problem for gaussian measure. *Discrete Comput. Geom.*, 10(4):411–420, dec 1993. 293

[BB20] Eric Blais and Abhinav Bommireddi. On testing and robust characterizations of convexity. In *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2020. 262

[BB21] Aleksandrs Belovs and Eric Blais. A polynomial lower bound for testing monotonicity. *SIAM Journal on Computing (SICOMP)*, 50(3):406–433, 2021. Prelim. version in Proc., STOC 2016. 226, 230, 236, 266, 270, 311

[BB23] Quentin Berger and Loïc Béthencourt. An application of sparre andersen's fluctuation theorem for exchangeable and sign-invariant random variables. *arXiv preprint arXiv:2304.09031*, 2023. 269, 288

[BBBY12] Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS*, 2012. 220, 231

[BBH23] Hadley Black, Eric Blais, and Nathaniel Harms. Testing convexity of discrete sets in high dimensions. *CoRR*, abs/2305.03194, 2023. 6, 232, 242

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012. Prelim. version in Proc., CCC 2011. 8, 210, 230

[BCO+15] Eric Blais, Clément L. Canonne, Igor Carboni Oliveira, Rocco A. Servedio, and Li-Yang Tan. Learning circuits with few negations. In *RANDOM*, 2015. 222, 223, 224, 228, 231, 232, 250

[BCP+20] Roksana Baleshzar, Deeparnab Chakrabarty, Ramesh Krishnan S. Pallavoor, Sofya Raskhodnikova, and C. Seshadhri. Optimal unateness testers for real-valued functions: Adaptivity helps. *Theory of Computing*, 16(3):1–36, 2020. 211

[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. 8, 9, 21, 22, 132, 190, 230

[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. 5, 8, 19, 22, 36, 134, 135, 136, 228, 230

[BCS23a] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. A $d^{1/2+o(1)}$ monotonicity tester for boolean functions on d-dimensional hypergrids. In *Proceedings, IEEE Symposium on Foundations of Computer Science (FOCS)*, 2023. 4, 5, 8, 9, 24, 25, 26, 27, 31, 33, 131, 230

[BCS23b] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Directed isoperimetric theorems for boolean functions on the hypergrid and an $\widetilde{O}(n\sqrt{d})$ monotonicity tester. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC*, 2023. 5, 8, 9, 22, 23, 24, 32, 65, 137, 138, 153, 154, 230

[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. 230

[BFH21] Eric Blais, Renato Ferreira Pinto Jr, and Nathaniel Harms. VC dimension and distribution-free sample-based testing. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2021. 231, 267, 332

[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. Prelim. version in Proc., RANDOM 2010. 230

[Bha08] Arnab Bhattacharyya. A note on the distance to monotonicity of boolean functions. Technical Report 012, Electronic Colloquium on Computational Complexity (ECCC), 2008. 230

[BKKM23] Mark Braverman, Subhash Khot, Guy Kindler, and Dor Minzer. Improved monotonicity testers via hypercube embeddings. In *14th Innovations in Theoretical Computer Science Conference, ITCS*, 2023. 8, 9, 137, 230

[BKR23] Hadley Black, Iden Kalemaj, and Sofya Raskhodnikova. Isoperimetric inequalities for real-valued functions with applications to monotonicity testing. In *Proceedings, International Colloquium on Automata, Languages and Programming (ICALP)*, 2023. xviii, 5, 8, 17, 24, 32, 33, 187, 200, 213, 230

[BL91] Béla Bollobás and Imre Leader. Edge-isoperimetric inequalities in the grid. *Combinatorica*, 11(4):299–314, 1991. 275

[Bla23] Hadley Black. Nearly optimal bounds for sample-based testing and learning of k-monotone functions. 2023. 6

[BMR19a] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. The power and limitations of uniform samples in testing properties of figures. *Algorithmica*, 81(3):1247–1266, 2019. 232, 262

[BMR19b] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Testing convexity of figures under the uniform distribution. *Random Struct. Algorithms*, 54(3):413–443, 2019. 232, 262

[BMR22] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Tolerant testers of image properties. *ACM Trans. Algorithms*, 18(4):37:1–37:39, 2022. 262

[BRY14a] Piotr Berman, Sofya Raskhodnikova, and Grigory Yaroslavtsev. L_p-testing. In *Proceedings, ACM Symposium on Theory of Computing (STOC)*, 2014. 8, 19, 21, 38, 40, 42, 51, 135, 230

[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. 230

[BT96] Nader H. Bshouty and Christino Tamon. On the fourier spectrum of monotone functions. *J. ACM*, 43(4):747–770, 1996. 221, 222, 223, 231

[BY19] Eric Blais and Yuichi Yoshida. A characterization of constant-sample testable properties. *Random Struct. Algorithms*, 55(1):73–88, 2019. 231

[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. 67

[CDJS17] Deeparnab Chakrabarty, Kashyap Dixit, Madhav Jha, and C Seshadhri. Property testing on product distributions: Optimal testers for bounded derivative properties. *ACM Trans. on Algorithms (TALG)*, 13(2):1–30, 2017. Prelim. version in Proc., SODA 2015. 230

[CDL+23] Xi Chen, Anindya De, Yuhao Li, Shivam Nadimpalli, and Rocco A. Servedio. Mildly exponential lower bounds on tolerant testers for monotonicity, unateness, and juntas, 2023. 32, 226, 230

[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. 230

[CFSS17a] Xi Chen, Adam Freilich, Rocco A. Servedio, and Timothy Sun. Sample-based high-dimensional convexity testing. *CoRR*, abs/1706.09362, 2017. 232

[CFSS17b] Xi Chen, Adam Freilich, Rocco A. Servedio, and Timothy Sun. Sample-based high-dimensional convexity testing. In *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2017. 262, 273, 274

[CGG+19] Clément L. Canonne, Elena Grigorescu, Siyao Guo, Akash Kumar, and Karl Wimmer. Testing k-monotonicity: The rise and fall of boolean functions. *Theory Comput.*, 15:1–55, 2019. 230, 231

[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. 8, 21, 230

[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. Prelim. version in Proc., STOC 2013. 8, 10, 12, 13, 230

[CS14b] Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory of Computing*, 10:453–464, 2014. Prelim. version in Proc., RANDOM 2013. 8, 18

[CS19] Deeparnab Chakrabarty and C. Seshadhri. Adaptive boolean monotonicity testing in total influence time. In *Innovations in Theoretical Computer Science (ITCS)*, pages 20:1–20:7, 2019. 8, 31

[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. 8, 13, 230

[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. 4, 7, 8, 31, 34, 224, 226, 230, 236, 266, 270, 311

[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. 4, 7, 8, 9, 20, 21, 38, 68, 71, 80, 135, 230

[DNS22] Anindya De, Shivam Nadimpalli, and Rocco A Servedio. Convex influences. In *13th Innovations in Theoretical Computer Science Conference (ITCS)*, 2022. 266, 274

[EKK+00] Funda Ergun, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. *J. Comput. System Sci.*, 60(3):717–751, 2000. Prelim. version in Proc., STOC 1998. 101, 230

[Eng97] Konrad Engel. *Sperner Theory*. Cambridge University Press, 1997. 275

[Fer23] Renato Ferreira Pinto Jr. Directed poincaré inequalities and l^1 monotonicity testing of lipschitz functions. In *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2023. 33

[FGL14] Eldar Fischer, Yonatan Goldhirsh, and Oded Lachish. Partial tests, universal tests and decomposability. In *Innovations in Theoretical Computer Science, ITCS*. ACM, 2014. 231

[FH23] Renato Ferreira Pinto Jr and Nathaniel Harms. Distribution testing under the parity trace, 2023. 220, 231

[Fis04] Eldar Fischer. On the strength of comparisons in property testing. *Information and Computation*, 189(1):107–116, 2004. 230

[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. 8, 12, 33, 38, 195, 210, 230

[FLV15] Eldar Fischer, Oded Lachish, and Yadu Vasudev. Trading query complexity for sample-based testing and multi-testing scalability. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS*. IEEE Computer Society, 2015. 220, 231

[FR10] Shahar Fattal and Dana Ron. Approximating the distance to monotonicity in high dimensions. *ACM Trans. on Algorithms (TALG)*, 6(3), 2010. 81, 230

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samordinsky. Testing monotonicity. *Combinatorica*, 20:301–337, 2000. Prelim. version in Proc., FOCS 1998. 4, 6, 7, 8, 10, 11, 68, 221, 223, 230

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998. 3, 220, 231, 273

[GKW19] Elena Grigorescu, Akash Kumar, and Karl Wimmer. Flipping out with many flips: Hardness of testing k-monotonicity. *SIAM J. Discret. Math.*, 33(4):2111–2125, 2019. 228, 230

[GR16] Oded Goldreich and Dana Ron. On sample-based testers. *ACM Trans. Comput. Theory*, 8(2):7:1–7:54, 2016. 220, 231

[HK03] Shirley Halevy and Eyal Kushilevitz. Distribution-free property testing. *Proceedings, International Workshop on Randomization and Computation (RANDOM)*, 2003. 230

[HK08] Shirley Halevy and Eyal Kushilevitz. Testing monotonicity over graph products. *Random Structures Algorithms*, 33(1):44–67, 2008. Prelim. version in Proc., ICALP 2004. 230

[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. 8, 19, 131, 134, 225, 228, 230, 231, 247, 248, 262, 274

[Kan14] Daniel M. Kane. The average sensitivity of an intersection of half spaces. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 437–440. ACM, 2014. 265, 269, 293, 294

[Kea98] Michael J. Kearns. Efficient noise-tolerant learning from statistical queries. *J. ACM*, 45(6):983–1006, 1998. 228, 250, 253, 254

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric-type theorems. *SIAM Journal on Computing*, 47(6):2238–2276, 2018. Prelim. version in Proc., FOCS 2015. 4, 7, 8, 13, 14, 15, 16, 17, 18, 22, 31, 32, 65, 66, 67, 68, 70, 73, 74, 79, 81, 82, 86, 96, 138, 142, 143, 153, 181, 182, 188, 200, 201, 202, 203, 204, 206, 216, 224, 230

[KOS08] Adam R. Klivans, Ryan O'Donnell, and Rocco A. Servedio. Learning geometric concepts via gaussian surface area. In *Proceedings, IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, 2008. 262

[KR00] Michael J. Kearns and Dana Ron. Testing problems with sublearning sample complexity. *J. Comput. Syst. Sci.*, 61(3):428–456, 2000. 220, 231

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. 228, 250, 253, 265, 270, 303

[LR01] Eric Lehman and Dana Ron. On disjoint chains of subsets. *Journal of Combinatorial Theory, Series A*, 94(2):399–404, 2001. 230, 242

[LRV22] Jane Lange, Ronitt Rubinfeld, and Arsen Vasilyan. Properly learning monotone functions via local correction. In *63rd IEEE Annual Symposium on Foundations* *of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 75–86. IEEE, 2022. 221, 222, 231

[LV23] Jane Lange and Arsen Vasilyan. Agnostic proper learning of monotone functions: beyond the black-box correction barrier. *CoRR*, abs/2304.02700, 2023. 32, 222, 231

[Mar74] Grigory A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy Peredachi Informatsii*, 10(2):101–108, 1974. 13

[MOA11] Albert W Marshall, Ingram Olkin, and Barry C Arnold. *Inequalities: theory of majorization and its applications*, volume 143. Springer, 2nd edition, 2011. 84

[MR78] Nicholas Metropolis and Gian-Carlo Rota. Combinatorial structure of the faces of the n-cube. *SIAM J. Appl. Math.*, 35(4):689–694, December 1978. 275

[Naz03] Fedor Nazarov. On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a gaussian measure. In *Geometric Aspects of Functional Analysis: Israel Seminar 2001-2002*, pages 169–187, 2003. 265, 293

[Nel15] Gail S. Nelson. *A user-friendly introduction to Lebesgue measure and integration*, volume 78. American Mathematical Soc., 2015. 54

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. 4, 270, 304, 305, 306

[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. 33, 230

[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 Proc., SODA 2020. 14, 32, 70, 78, 79, 86, 96, 99, 213, 214, 215, 216, 217, 218

[Ras99] Sofya Raskhodnikova. Monotonicity testing. *Masters Thesis, MIT*, 1999. 8, 230

[Ras03] Sofya Raskhodnikova. Approximate testing of visual properties. In *Proceedings, International Workshop on Randomization and Computation (RANDOM)*. Springer, 2003. 262

[RG09] Luis Rademacher and Navin Goyal. Learning convex bodies is hard. 2009. 262

[RRS+12] Dana Ron, Ronitt Rubinfeld, Muli Safra, Alex Samorodnitsky, and Omri Weinstein. Approximating the influence of monotone boolean functions in $O(\sqrt{n})$ query complexity. *ACM Trans. Comput. Theory*, 4(4):11:1–11:12, 2012. Prelim. version in Proc., RANDOM 2011. 230

[RS96] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal of Computing*, 25:647–668, 1996. 3, 220

[RV04] Luis Rademacher and Santosh S. Vempala. Testing geometric convexity. In *Proceedings, Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. Springer, 2004. 262

[Sch92] Bernd Schmeltz. Learning convex sets under uniform distribution. In *Data Structures and Efficient Algorithms, Final Report on the DFG Special Joint Initiative*, Lecture Notes in Computer Science, pages 204–213. Springer, 1992. 262

[Spa54] Erik Sparre Andersen. On the fluctuations of sums of random variables II. 2(2):195–223, 1954. 269, 288

[SS05] Elias Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2005. 54, 56, 57

[SS08] Michael E. Saks and C. Seshadhri. Parallel monotonicity reconstruction. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008. 230

[SWC+22] Ben Stiller, Nicholas Weinstock, Jackie Cohn, Mark Friedman, Dan Erickson, Chris Black Andrew Colville, and John Cameron (Executive Producers). Severance, 2022. 1

[Tal93] Michel Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis' graph connectivity theorem. *Geom. Func. Anal.*, 3(3):295–314, 1993. 14, 82

[Tal96] Michel Talagrand. How much are increasing sets positively correlated? *Comb.*, 16(2):243–258, 1996. 226

[Val84] Leslie G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984. 3