On Monotonicity Testing and the 2-to-2 Games Conjecture

A thesis submitted for the degree of Doctor of Philosophy by Dor Minzer

Thesis supervisor: Prof. Shmuel Safra

Submitted to the Senate of Tel Aviv University September 2018
# Contents

<table>
<thead>
<tr>
<th>Abstract</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>vii</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Monotonicity Testing (Chapter 2)</td>
<td>1</td>
</tr>
<tr>
<td>1.2 NP-Hardness of 2-to-2 Games (Chapter 3)</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Pseudo-random Sets in the Grassmann Graph Have Near-Perfect Expansion (Chapter 4)</td>
<td>8</td>
</tr>
<tr>
<td>1.4 The Biased Long Code and Hardness of Vertex-Cover (Chapter 5)</td>
<td>10</td>
</tr>
<tr>
<td>1.5 Conclusions and Open Problems (Chapter 6)</td>
<td>10</td>
</tr>
<tr>
<td><strong>2 Monotonicity Testing and Directed Isoperimetric Inequalities</strong></td>
<td>11</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>11</td>
</tr>
<tr>
<td>2.1.1 Boolean Isoperimetric Type Theorems</td>
<td>11</td>
</tr>
<tr>
<td>2.1.2 Directed Analogues of Boolean Isoperimetric Type Theorems</td>
<td>12</td>
</tr>
<tr>
<td>2.1.3 Robust version of Talagrand’s Theorem</td>
<td>14</td>
</tr>
<tr>
<td>2.1.4 A Robust and Directed Analogue of Talagrand’s Theorem</td>
<td>14</td>
</tr>
<tr>
<td>2.1.5 Monotonicity Testing</td>
<td>14</td>
</tr>
<tr>
<td>2.1.6 Followup works</td>
<td>17</td>
</tr>
<tr>
<td>2.2 Proof of Theorem 2.1.8</td>
<td>17</td>
</tr>
<tr>
<td>2.3 The Switch and the Split Operators</td>
<td>19</td>
</tr>
<tr>
<td>2.3.1 The Switch Operator</td>
<td>19</td>
</tr>
<tr>
<td>2.3.2 The Split Operator</td>
<td>22</td>
</tr>
<tr>
<td>2.3.3 Pure Functions</td>
<td>24</td>
</tr>
<tr>
<td>2.3.4 Splitting only Decreases Talagrand Objective</td>
<td>26</td>
</tr>
<tr>
<td>2.4 Proof of Theorem 2.1.6</td>
<td>31</td>
</tr>
<tr>
<td>2.4.1 An Overview</td>
<td>31</td>
</tr>
<tr>
<td>2.4.2 The Formal Proof</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Proof of Theorem 2.1.9</td>
<td>37</td>
</tr>
<tr>
<td>2.6 Some Useful Results</td>
<td>38</td>
</tr>
<tr>
<td>2.6.1 A Combinatorial Fact</td>
<td>38</td>
</tr>
<tr>
<td>2.6.2 Bound on Total Influence</td>
<td>39</td>
</tr>
<tr>
<td>2.6.3 Bound on Fraction of Non-persistent Inputs</td>
<td>40</td>
</tr>
<tr>
<td>2.7 Algorithm for Monotonicity Testing</td>
<td>42</td>
</tr>
<tr>
<td>2.7.1 Existence of a Good Subgraph</td>
<td>42</td>
</tr>
<tr>
<td>2.7.2 The Tester</td>
<td>42</td>
</tr>
<tr>
<td>2.7.3 The Analysis of the Tester</td>
<td>43</td>
</tr>
</tbody>
</table>
# Table of Contents

2.8 A Lower Bound for Monotonicity Testing .................................................. 45
2.9 Additional Observations ........................................................................... 46
  2.9.1 Theorem 2.1.9 \(\Rightarrow\) Theorem 2.1.7 .............................................. 46
  2.9.2 Undirected Theorems from Directed Theorems ................................. 46

3 NP-hardness of 2-to-2 Games ........................................................................ 49
  3.1 Introduction ............................................................................................... 49
  3.1.1 Informal Overview of the Reduction ..................................................... 50
  3.2 The Inner PCP: Grassmann Graph and Encoding, 2-to-2 Linearity Test, Zooming In and Out ................................................................. 53
    3.2.1 Subspace Example and Zooming-in .................................................... 54
    3.2.2 Hyperplane Example and Zooming-Out ........................................... 56
  3.3 The Outer PCP Game ............................................................................... 59
    3.3.1 Equation vs Variable Game ................................................................. 60
    3.3.2 Smooth Equation vs Variable Game .................................................. 60
    3.3.3 Smooth Equation vs Variable Game with Advice .............................. 60
    3.3.4 The Final Game (Outer PCP) ............................................................... 61
  3.4 The Reduction to 2-to-2 Games ............................................................... 61
    3.4.1 Construction of the Transitive Game \(G_{2;2}\) .................................... 61
    3.4.2 The Game \(G_{2;2}\) is a Transitive Game ............................................ 64
    3.4.3 Construction of the Final 2-to-2 Games Instance ......................... 67
    3.4.4 Completeness ..................................................................................... 68
    3.4.5 Setting of the Parameters ................................................................. 68
  3.5 Preliminary steps towards Soundness Analysis ........................................ 68
    3.5.1 Covering Property ............................................................................ 69
    3.5.2 The Number of Successful Zoom-outs of Minimal Co-dimension is Bounded ................................................................. 70
    3.5.3 Incorporating Side Condition ............................................................. 75
  3.6 Soundness Analysis ................................................................................. 77
    3.6.1 Clique Consistent Assignment ......................................................... 77
    3.6.2 A Strategy for the First (Larger) Prover .......................................... 78
    3.6.3 A Strategy for the Second (Smaller) Prover .................................... 79
    3.6.4 The Success Probability of the Provers ......................................... 80
    3.6.5 Auxiliary Lemmas ............................................................................ 82
  3.7 “\(f\)-space vs \(b\)-space” Linearity Test ..................................................... 85
  3.8 Missing Proofs ......................................................................................... 89
    3.8.1 Covering Properties ..................................................................... 89
    3.8.2 Proof of Lemma 3.5.8 ...................................................................... 91
    3.8.3 Proof of the Upper Bound in Lemma 3.5.9 .................................... 91
    3.8.4 Auxiliary Lemmas and Facts ............................................................ 92

4 Pseudo-random Sets in the Grassmann Graph Have Near-Perfect Expansion ......................................................................................... 95
  4.1 Introduction ............................................................................................. 95
  4.2 Preliminaries ........................................................................................... 97
    4.2.1 The Barak-Kothari-Steurer Reduction ........................................... 97
    4.2.2 High Level Proof of Theorem 4.1.6 ............................................... 98
    4.2.3 Switching to the Graph \(H_{k,\ell}\) ......................................................... 99
    4.2.4 It Sufﬁces to Work with \(H_{k,\ell}\) ...................................................... 100
    4.2.5 The Eigenvectors and Eigenvalues of \(H_{k,\ell}\) and Fourier Levels .... 101
    4.2.6 Pseudorandomness \(\Rightarrow\) Low Weight at Low Levels \(\Rightarrow\) Near-Perfect Expansion ................................................................. 103
4.2.7 Lower-bounding the Fourth Moment of \( F_i \) .................................................. 104
4.2.8 Upper-bounding the Fourth Moment of \( F_i \) .................................................. 104
4.3 Analytic Machinery ................................................................. 105
  4.3.1 High-level Perspective .......................................................... 106
  4.3.2 Zoom-out Restriction Lemma ................................................... 106
  4.3.3 Defining \( f_x \) and Relating \( F_{xy}, f_x \) and Zoom-in Densities .................... 107
  4.3.4 Relating \( F_{xy} \) and \( f_x \) in the Fourier Domain ........................................... 114
  4.3.5 Bounding Restricted Fourier Sums of \( f_x \) .................................................. 117
4.4 Pair-wise and Three-wise Correlations of \( f_i \) ..................................... 120
  4.4.1 Pairwise Correlations ............................................................... 120
  4.4.2 Three-wise Correlations ............................................................. 121
4.5 Four-wise Correlations: Transforming the Matrices into Form ..................... 124
  4.5.1 Removing 4-wise and 3-wise Intersections of Rowspaces ......................... 124
  4.5.2 Getting \( M_1, M_2, M_3 \) into Form .................................................... 125
  4.5.3 Getting \( M_4 \) into Form: Part I ....................................................... 125
  4.5.4 Getting \( M_4 \) into Form: Part II ..................................................... 126
  4.5.5 Getting \( M_4 \) into Form: Part III .................................................... 127
4.6 Four-wise Correlations: A (somewhat) Simplified Case ................................ 130
4.7 Four-wise Correlations: the General Case ......................................... 134
4.8 Appendix: Auxiliary Lemmas ...................................................... 140

5 The Biased Long Code and Hardness of Vertex-Cover ..................................... 146
  5.1 Biased Long Code ................................................................. 146
    5.1.1 Independent Sets in the Kneser Graph ............................................. 146
  5.2 The Reduction ................................................................. 148
    5.2.1 Completeness ................................................................. 148
    5.2.2 Soundness ................................................................. 149

6 Conclusions and Open Problems .................................................... 154

Bibliography .................................................. 156

תקציר והפתעות בעברית
Abstract

This thesis discusses two questions in Complexity Theory: Monotonicity Testing and the 2-to-2 Games Conjecture.

Monotonicity Testing is a problem first considered by Goldreich et al. [GGL00], who gave a $O(n/\varepsilon)$ query tester for testing if a function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ is monotone or $\varepsilon$-far from monotone. The problem has since received considerable attention [CS13a, CST14], and it was shown that monotonicity can be tested using $\tilde{O}(n^{5/6} \varepsilon^{-4})$ queries. Our main results for Monotonicity Testing is a $\tilde{O}(n^{5/6} \varepsilon^{-2})$ query tester, which is optimal up to poly-logarithmic factors.

The analysis of the algorithm relies heavily on a directed and robust analogue of a Boolean isoperimetric inequality of Talagrand [Tal93].

The PCP Theorem is one of the cornerstones of modern Theoretical Computer Science. One area in which PCPs are essential, is the area of hardness of approximation. Therein, the goal is to prove that some optimization problems are hard to solve, even in approximately. Many hardness of approximation results were proven using the PCP theorem, however for some problems, optimal results were not obtained. This thesis touches some of this problems, and in particular the 2-to-2 Games problem and the Vertex-Cover problem. Our main result for PCP’s is a proof of the 2-to-2 Games Conjecture (with imperfect completeness). This result has few additional consequence on the hardness of Unique-Games ($1/2$ vs $\varepsilon$), Vertex-Cover ($\sqrt{2} - \varepsilon$), Almost-4-Coloring, Max-Cut-Gain and more.

At the core of the analysis of the reduction, is a characterization of small sets of vertices in Grassmann Graphs whose edge expansion is bounded away from 1.

This thesis contains material from the papers [KMS15, KMS17, DKK+18a, DKK+18b, KMS18].
Acknowledgements

First and foremost, I would like to express my deepest gratitude for my advisor Muli Safra. Muli and I began working together shortly after a conference held in December 2013 at Weizmann Institute in celebration of Goldwasser’s and Micali’s Turing award. Working with Muli has been: challenging, enjoyable, frustrating and rewarding, all at the same time; I truly couldn’t ask for more from this experience. Muli knew exactly when to be more involved, and when to be less. His insights, knowledge and intuition about scientific and non-scientific topics are truly remarkable, and I’m confident they have influenced me significantly over the last few years. I wouldn’t be half the researcher I am today if it weren’t for him.

I wish to thank Subhash Khot, who co-authored all of the work herein. I first met Subhash in the summer of 2015 when we started working together on the Vertex-Cover problem. Our collaboration that has started even before that and has continued since.

I would like to thank Irit Dinur and Guy Kindler for their collaboration and enthusiasm. To this day I do not know a person who is more enthusiastic about PCP’s than Irit.

I would like to thank Noga Alon and Gil Kalai for serving in my masters and research proposal committee. Though indirectly, they both had an impact on my research.

Vadim Grinshtein gave me very helpful advice during my years at the Open-University. His teaching and his love of mathematics have had great influence on me, and is mostly the reason I continued to study mathematics during my Bachelors. If it weren’t for him, I would have gone abroad for Bachelors studies.

During my Bachelors studies at Tel-Aviv, Mikhail Sodin has a mathematical mentor for me. He taught me a great deal about analysis, and if it weren’t for him I might have done my Ph.D abroad. I wish to thank him for encouragement in various stages; once, while consulting with him about doing a Ph.D, I told him that I just discovered a (possibly fatal) error in one of my proofs that possibly undermines the entire idea. Mikhail calmly told him that this has happened to him many times, and except for one case, his idea was sound and he was able to fix the error.

I would like to thank the administrative staff of in Tel-Aviv University, and especially Pnina Neira Brazilay, for helping me in various topics (especially housing).

Throughout the years I met several people who made the academic life more enjoyable. I would like to thank my office mates throughout the years: Orr Fischer, Ben-Lee Volk, Esty Kelman, Shelly Grossman. I learned from Ben-Lee about the greatness of polynomials, tex and all I know about teaching. I would like to thank Saleet Klein, Talya Eden, Amit Levi for enjoyable times.

Last and not least, I would like to thank my parents, Rudi and Dorit, and my sister Tal, for making it all possible. Ph.D life was not always easy for me, and I wish to thank them for putting me into perspective in rough times. I would especially like to thank my father Rudi for expressing so much interest in my life and work; he is always interested in hearing about math, computation and the Unique-Games Conjecture.
Chapter 1

Introduction

Boolean functions are central in the study of computation. All of the basic computational models (such as Turing Machines and Boolean circuits) attempt to capture the computational feasibility of Boolean functions. Analyzing Boolean plays crucial role in many exciting developments in theoretical computer science, such as Probabilistically Checkable Proofs, Property Testing, Learning Theory, Communication Complexity.

This thesis largely studies Boolean functions, from several aspects. The questions that motivate this study are the Monotonicity Testing problem and the 2-to-2 Games Conjecture. Interestingly, while both questions are very much about computation and efficiency, their resolution in this thesis is reduced to purely mathematical questions regarding Boolean Functions. In the case of monotonicity testing, the question reduces to directed isoperimetric-type inequalities on the hypercube, which is studied using (mostly) combinatorial techniques. In the case of 2-to-2 Games Conjecture, the question reduces to structural results on sets in the Grassmann Graph whose expansion is far from 1, which is studied using analytic techniques.

This introductory section presents the questions of interest in this thesis, the main result proven and their implication.

1.1 Monotonicity Testing (Chapter 2)

Property testing is a field that studies what properties of large objects, such as graphs and functions, can be checked very efficiently. Here, efficiently means within a sub-linear time, namely without even looking at the entire object! In particular, since the algorithm is unable to look at the entire object, some relaxation has to made. Therefore, a tester is usually allowed to use randomness – but this by itself is not sufficient: if two objects \( O, O' \) are extremely close to each other, and only one of them (say \( O' \)) has the property, the tester will probably fail to distinguish the two objects and thus fail in the task. Therefore, to have more meaningful questions one allows a further relaxation: the tester must accept objects that have the property with high probability, and reject objects that are far from having the property with high probability. The notion of “farness” may deviate from problem to problem, and is usually some \( \ell_p \) distance (\( 0 \leq p < \infty \), the case \( p = 0 \) is also known as the “hamming distance”).

The Blum-Luby-Rubinfeld Linearity Tester [BLR93] is a prime example of such algorithm, and perhaps the first one. Since then, many properties of functions, graphs and distributions were studied (see the book of Goldreich [Gol17]). Originally, the study of property testing was motivated by the study of Probabilistically Checkable Proofs (PCP’s), in which “local to global” tests are an essential components. However, today property testing is a rich and independent field of study.

Chapter 4 is concerned with the problem of Monotonicity Testing posed by Goldreich et al [GGL+00], defined next. Consider the Boolean hypercube \( \{0, 1\}^n \), and define a partial ordering \( \leq \) on it: we say \( x, y \in \{0, 1\}^n \) obey \( x \preceq y \) if for every \( i = 1, \ldots, n \) we have \( x_i \leq y_i \). A function \( f : \{0, 1\}^n \to \{0, 1\} \) is called
monotone if whenever \( x \preceq y \), we have \( f(x) \leq f(y) \). For example, the function \( f(x) = x_1 \) is monotone, as well as \( g(x) = \text{maj}(x_1, \ldots, x_n) \) (i.e. that function which is 1 if and only if more than half the bits \( x_1, \ldots, x_n \) are 1).

The distance between two functions \( f, g \) is the (normalized) Hamming distance:

\[
\Delta(f, g) = \frac{1}{2^n} |\{ x \mid f(x) \neq g(x)\}| .
\]

The distance of \( f \) from the class of monotone functions is defined to be \( \min_g \Delta(f, g) \). For a proximity parameter \( \varepsilon > 0 \), a tester \( T \) is called a (one-sided error) monotonicity tester, if given a black-box access to a function \( f \), makes at most \( q \) queries to it and:

- If \( f \) is monotone, \( T \) must accept with probability 1.
- If \( f \) is \( \varepsilon \)-far from monotone, then \( T \) must accept with probability at most \( \frac{1}{3} \).

The interesting question here is what is the minimum number of queries \( q \) for which there is a monotonicity tester \( T \). The parameter \( \varepsilon \) may depend on \( n \) in general, however for the discussion here it is best to think of it as a small constant.

The first tester for this problem was given by Goldreich et al. \cite{GGL+00} and has \( q = O(n/\varepsilon) \) queries. The tester is very simple: pick \( x \in \{0, 1\}^n \) and pick \( i \in [n] \), both uniformly at random, and query \( f(x), f(x \oplus e_i) \) where \( x \oplus e_i \) is the point \( x \) with the \( i\)th bit flipped. Check if \( f(x), f(x \oplus e_i) \) respect the order between the two points (note that any two points that differ on one coordinate are comparable according to \( \preceq \)). It is clear that if \( f \) is monotone, then it passes the test with probability 1, and the main result of \cite{GGL+00} showed that if \( f \) is \( \varepsilon \)-far from monotone, then the tester rejects \( f \) with probability at least \( \Omega(\varepsilon/n) \). Thus, repeating this test \( O(n/\varepsilon) \) times gives a tester as desired, with \( O(n/\varepsilon) \) queries.

Shortly thereafter, it was shown by Fischer et al. \cite{FLN+02} that any monotonicity tester must make at least \( \Omega(\sqrt{n}/\varepsilon) \) queries, leaving a quadratic gap between the lower and the upper bound for the problem.

It took over a decade until an improved tester for the problem was found by Chakrabarty and Seshadhri \cite{CS13}, who showed a \( \tilde{O}(n^{7/8} \varepsilon^{-3/2}) \) query tester for the problem. Roughly speaking, their tester sampled \( x \preceq y \) that differ either on a single coordinate (in which case the tester is the \cite{GGL+00} tester), or on \( \Theta(n^{3/8} \sqrt{\varepsilon}) \) coordinates. The analysis of this tester is more complicated than the analysis of the \cite{GGL+00} tester, and exhibited for the first time a deep connection between directed isoperimetric inequalities on the hypercube – and more precisely a directed version of an isoperimetric inequality of Margulis \cite{Mar74} and the problem of Monotonicity Testing.

An improved tester and a more refined analysis were later given by \cite{CST14} who showed a tester with \( \tilde{O}(n^{5/6} \varepsilon^{-4}) \) queries, however the key ingredient in the analysis remained the same, namely the Chakrabarty-Seshadhri directed version of the Margulis isoperimetric inequality.

The main result of this chapter is an optimal (up to poly-logarithmic factors) \( \tilde{O}(\sqrt{n} \varepsilon^{-2}) \) query tester for monotonicity. The tester is very natural: sample \( \tau \in \{ 2^t \mid t = 0, 1, \ldots, \frac{1}{2} \log n \} \), sample \( x \preceq y \) that are distance \( \tau \) apart (according to Hamming distance), query \( f(x), f(y) \) and reject if and only if \( f(x) > f(y) \). The analysis of the tester, however, is more challenging, and the key new ingredient is a directed version of an isoperimetric inequality of Talagrand \cite{Tal93} (which by itself is a strengthening of the isoperimetric inequality of Margulis). The proof of the direction version is significantly more difficult than the proof of undirected version, and uses a new operator on functions called the “Split Operator”.

**Source Material.** This chapter is based on the paper \cite{KMS15}.
1.2 NP-Hardness of 2-to-2 Games (Chapter 3)

The PCP theorem [FGL+96, AS98, ALM+98] is perhaps one of the greatest discoveries in theoretical computer science. It has applications in various topics such as cryptography and verification, and most prominently in the field of hardness of approximation.

Papadimitriou and Yannakakis [PY91] showed that if 3SAT is hard to approximate within some factor (which is equivalent to the PCP Theorem), then a large class of computational problems are hard to approximate within some constant factor. While very general, the hardness guarantee it promises is often weak, yielding only hardness for factors close to 1. Proving optimal inapproximability results for specific problems requires significantly more work.

Subsequent works developed many techniques for this goal: the Bellare-Goldreich-Sudan [BGS98] Long Code, Raz’s Parallel Repetition Theorem [Raz98] and the Fourier Analytic framework of Håstad [Hås01]. This set of techniques has since yielded a general recipe for proving hardness of approximation results, and a very successful one (for instance [Hås99, Hås01, ABH+05, Fei98] among many other works). However for some problems, this methodology failed to obtain optimal inapproximability results, and especially for 2CPS’s $^1$.

One notable problem for which this set of techniques failed, is the Vertex Cover problem. Given a graph $G = (V, E)$, the goal is to find the smallest set of vertices $C \subseteq V$ that contains at least one vertex from each edge of $G$. The problem admits an easy 2-approximation, however the best hardness factor obtained by the previously described techniques is Håstad’s $7/6 - o(1)$ [Hås01]. Dinur and Safra [DS05] improved this result, showing that Vertex Cover is hard to approximate within factor $10\sqrt{5} - 21 \approx 1.36$. Their reduction used a new property of PCP constructions, called $d$-to-$d$-ness, which we define next.

**Definition 1.2.1.** A 2-Prover-1-Round Game$^2$ $G = (V, E, \Phi, \Sigma)$ consists of a set of vertices $V$, a set of colors $\Sigma$, and a constraint $\Phi(u, v)$ for each (directed) edge $(u, v) \in E$. The goal is to assign colors to the vertices, say $A : V \to \Sigma$, so as to satisfy the maximum fraction of constraints. A constraint $\Phi(u, v)$ is satisfied if $(A(u), A(v)) \in \Phi(u, v)$, where $\Phi(u, v) \subseteq \Sigma \times \Sigma$ denotes the subset of color-pairs that are deemed satisfactory. The value of a game, denoted by $\text{val}(G)$, is the maximum fraction of constraints that can be satisfied by any assignments $A : V \to \Sigma$.

**Definition 1.2.2.** Let $d \geq 1$ be an integer. A constraint $\Phi(u, v) \subseteq \Sigma \times \Sigma$ is said to be a $d$-to-$d$ constraint if there are partitions $A_1, \ldots, A_r$ and $B_1, \ldots, B_r$ of $\Sigma$ into sets of size $d$ such that

$$\Phi(u, v) = \bigcup_{i=1}^r A_i \times B_i.$$ 

A 2-Prover-1-Round Game $G = (V, E, \Phi, \Sigma)$ is said to be a $d$-to-$d$ Game if every constraint $\Phi(u, v)$ is a $d$-to-$d$ constraint. A 1-to-1 Game is also called a Unique Game. In this case, $\Phi(u, v)$ is simply a perfect matching on $\Sigma \times \Sigma$.

$^1$A kCSP is composed of a predicate $P : \Sigma^k \to \{0, 1\}$, a set of variables $X$ and a set of constraints of the form $P(x_1, \ldots, x_k) = 1$. The goal is to choose an assignment $A : X \to \Sigma$ that satisfies maximum number of constraints.

$^2$The more conventional definition is slightly different. A 2-Prover-1-Round Game has a bipartite structure: the set of variables is partitioned into the left and the right side and every constraint is between two variables, one on either side. Instead of as an optimization problem, it can be equivalently viewed as a “game” between two provers and a verifier: the verifier picks a constraint $(u, v)$ at random, asks the “question” $u$ to the left prover $P_1$, the “question” $v$ to the right prover $P_2$, receives “answers” $P_1(u), P_2(v)$ respectively from the provers, and accepts if and only if $(P_1(u), P_2(v)) \in \Phi(u, v)$. The maximum acceptance probability of the verifier over all prover “strategies” is then the same as the maximum fraction of the constraints that can be satisfied by a coloring to the game.
Dinur and Safra considered a non-standard optimization parameter for \(d\)-to-\(d\) Games: find the largest \(V' \subset V\) and a partial assignment, i.e. \(A: V' \to \Sigma\), that satisfies all constraints inside \(V'\). Using their result for 2-to-2 games and their Biased Long Code, they showed new hardness gaps for the Independent Set and Vertex Cover problems.

Motivated by their work, Khot [Kho02] conjectured the hardness of an extreme case of \(d\)-to-\(d\) Games, known as the Unique-Games Conjecture. For an optimization problem \(A\) and \(0 \leq s < c \leq 1\), let \(\text{Gap}A[c, s]\) be the promise problem of distinguishing whether a given instance has value at least \(c\), or at most \(s\).

**Conjecture 1.2.3. Unique Games Conjecture:** For every constant \(\delta > 0\), for sufficiently large constant \(k = |\Sigma|\), given an instance \(G = (V, E, \Phi, \Sigma)\) of a Unique Game, it is NP-hard to distinguish between

- YES case: there is a coloring satisfying \(1 - \delta\) fraction of the constraints of \(G\).
- NO case: no coloring satisfies more than \(\delta\) fraction of the constraints of \(G\).

In gap notions, \(\text{GapUG}[1 - \delta, \delta]\) is NP-hard for instances with alphabet size \(k\).

Note that given a satisfiable Unique-Games instance, it is easy to find an assignment that satisfies all the constraints and thus the completeness, namely the YES case, i.e. \(\text{GapUG}[1, \delta]\) is in P for every \(\delta < 1\). In the same paper, Khot has also conjectured the following hardness for \(d\)-to-\(d\) games:

**Conjecture 1.2.4. \(d\)-to-\(d\) Games Conjecture:** Let \(d \geq 2\) be an integer. For every constant \(\delta > 0\), for sufficiently large constant \(|\Sigma|\), given an instance \(G = (V, E, \Phi, \Sigma)\) of a \(d\)-to-\(d\) Games, it is NP-hard to distinguish between:

- YES case: there is a coloring satisfying all the constraints of \(G\).
- NO case: no coloring satisfies more than \(\delta\) fraction of the constraints of \(G\).

We stress that a key point in the \(d\)-to-\(d\)-Games Conjecture is that the soundness is vanishing while \(d\) is a fixed constant. Using the PCP Theorem [FGL+96, AS98, ALM+98] and Raz’s Parallel Repetition Theorem [Raz98] one can achieve vanishing soundness \(\delta\), however \(d\) grows polynomially in \(1/\delta\) in the process.

These conjectures—and especially the Unique-Games Conjecture—were shown to have significant implications to hardness of approximation. Assuming the Unique-Games Conjecture, many new tight inapproximability results were achieved (e.g. [KKMO07, KR08, CKK+06, DMR09], see [Kho10, Tre12] for a more comprehensive list). Remarkably, the Unique-Games Conjecture also implies that a large class of problems, the best approximation ratio possible is achievable via a generic SDP (Semi-Definite Programming) algorithm [Rag08].

Despite considerable efforts, the Unique-Games Conjecture is still poorly understood. Moreover, in contrast to popular beliefs such as P \(\neq\) NP, researchers are divided regarding the validity of the conjecture, and it was attacked both from the algorithmic and hardness sides. On the algorithmic side, Khot [Kho02] analyzed a natural SDP based algorithm for Unique-Games, and showed guarantee of finding a solution satisfying \(c(k, \epsilon) > 0\) of the constraints given a \((1 - \epsilon)\)-satisfiable instance with alphabet size \(k\). Since \(c(k, \epsilon)\) approaches 0 as \(k\) tends to infinity, it follows that the alphabet size has to be sufficiently large in \(\delta\) in the conjecture. Since then, many SDP based algorithms have been proposed and analyzed [GT06, CMM06, Tre08], all falling short of refuting the conjecture. Using spectral techniques, it was shown that finding good assignments for Unique-Games whose graph is an expander can be done efficiently [AKK+08, Kol11]. These ideas, along with a “partitioning into expanders” procedure were later used by Arora, Barak and Steurer [ABS15] to obtain a sub-exponential time algorithm for Unique-Games. This algorithm runs in time

---

\(^3\)More precisely, they allow \(j\)-partial assignments that to each vertex give (up-to) \(j\) colors. \(j\)-assignments on \(u, v\) are said to satisfy the constraint on \((u, v)\) if there is a pair of colors \(a_u, a_v\) in the \(j\) assignment of \(u, v\) respectively that satisfies the constraint.
exp(\(n^{\alpha}\)) for some absolute constant \(\alpha > 0\), that given an \(1 - \varepsilon\) satisfiable instance of Unique-Games, it produces an assignment satisfying \(1 - \varepsilon^\alpha\) fraction of the constraints. This result was later reproduced using Sum-of-Squares algorithms \([\text{BBH} + 12]\).

On the hardness side, it is known that approximating the value of Unique-Games within any constant factor is NP-hard \([\text{FR04}]\). However, if one wishes to have a constant (non-vanishing) completeness, only partial hardness gaps are known \([\text{OW12}]\) – namely, it is NP-hard to distinguish whether a given Unique-Games has value at least \(1/2\), or at most \(3/8 + \varepsilon\). One of the main results of this thesis is the following theorem.

**Theorem 1.2.5.** For every \(\delta > 0\) there exists \(k\), such that given a Unique-Games instance \(G\) with alphabet size \(k\), it is NP-hard to distinguish between:

- **YES case:** there is a coloring satisfying \(\frac{1}{2}\) of the constraints of \(G\).
- **NO case:** no coloring satisfies more than \(\delta\) fraction of the constraints of \(G\).

This result is in fact a rather straightforward corollary of the following theorem, which is one of the main results of this thesis.

**Theorem 1.2.6.** For every constant \(\delta > 0\) there exists \(k \in \mathbb{N}\) such that given a 2-to-2-Game instance \(G = (V, E, \Phi, \Sigma)\) with alphabet size \(k\), it is NP-hard to distinguish between:

- **YES case:** there is a coloring satisfying \(1 - \delta\) fraction of the constraints of \(G\).
- **NO case:** no coloring satisfies more than \(\delta\) fraction of the constraints of \(G\).

In fact, one has the following additional properties: (a) In the YES case: there is a set \(X\) of \(1 - \delta\) fraction of the vertices such that all constraints inside \(X\) are satisfied, and (b) in both cases, any set containing \(\delta\) fraction of the vertices contains \(\Omega(\delta^2)\) fraction of the edges of the graph. These conditions are necessary towards certain applications, e.g. to the Vertex Cover and the Independent Set problems.

In words, Conjecture 1.2.4 holds with imperfect completeness. The proof appears in Chapters 3 and 4.

### Hardness of 2-to-2 Games as Evidence Towards the Unique Games Conjecture

The sub-exponential time algorithms for Unique Games \([\text{ABS15}]\) carried with it the (real) possibility that the Unique-Games Conjecture is false, and moreover that it is just a matter of time until a “slightly faster” algorithm is found, say quasi-polynomial time (thereby essentially refuting the Unique-Games Conjecture). Our results (Theorems 1.2.5, 1.2.6) imply that the existing set of algorithmic techniques refute the Unique-Games Conjecture (unless \(P = \text{NP}\)), since most of them (and especially the Spectral/Sum-of-Squares algorithms) work equally well for 2-to-2 games or Unique-Games with constant completeness (that are shown to be NP-hard herein). This can be viewed as compelling evidence for the Unique Games Conjecture, or more modestly that any future refutation attempts must utilize the near perfect completeness of the problem, and in particular must use new algorithmic ideas.

### Implications

#### Hardness of Approximation

Theorem 1.2.6 implies several new hardness of approximation results for various problems that we discuss next. The first implication is for the hardness of approximating the maximum Independent Set.

**Theorem 1.2.7.** For every \(\varepsilon > 0\), given an \(n\)-vertex graph \(G\) it is NP-hard to distinguish between:

- **YES case:** \(G\) has an independent set of size at least \((1 - \frac{1}{\sqrt{2}}) n\).
• **NO case**: no independent set in $G$ has size $\varepsilon n$.

Stated as a promise problem, given a graph $G$ that has an independent set of size $(1 - \frac{1}{\sqrt{2}} - \varepsilon)n$, it is NP-hard to find an independent set of size $\varepsilon n$. In the language of [BGS98], it implies that NP has a Probabilistically Checkable Proof with zero free bit complexity, constant completeness and vanishing soundness, see [BGS98, Proposition 5.6, Theorem 8.2].

Theorem 1.2.7 follows from Theorem 1.2.6 using the Biased Long Code of Dinur and Safra [DS05].

Using the fact that the complement of a maximum independent set is a minimum vertex cover, one gets the following hardness result for the Vertex Cover problem.

**Corollary 1.2.8.** For every $\varepsilon > 0$, given an $n$-vertex graph $G$ it is NP-hard to distinguish between:

- **YES case**: $G$ has a vertex cover of size at most $\frac{1}{\sqrt{2}}n$.
- **NO case**: any vertex cover of $G$ has size at least $(1 - \varepsilon)n$.

In particular, it is NP-hard to approximate the size of the minimum vertex cover of $G$ within factor $\sqrt{2} - \varepsilon$. This improves the result of Dinur and Safra [DS05].

By the results of Dinur, Mossel and Regev [DMR09], the following result for almost-4-coloring holds:

**Theorem 1.2.9.** For every $\delta > 0$, given an $n$-vertex graph $G$ it is NP-hard to distinguish between:

- **YES case**: $G$ has four disjoint independent sets of size $(\frac{1}{4} - \delta)n$ each.
- **NO case**: there is no independent set in $G$ of size $\delta n$.

By the result of Guruswami and Sinop [GS13] we have the following:

**Theorem 1.2.10.** There exists $c > 0$, such that for every $\delta > 0$, $k \in \mathbb{N}$, $k \geq 3$, given a $n$-vertex graph $G$ it is NP-hard to distinguish between:

- **YES case**: there is a proper $k$ coloring of $(1 - \delta)n$ of the vertices of $G$.
- **NO case**: any $k$-coloring of $G$ has at least $\frac{1}{k} - c \log(k) / k^2$ of the edges monochromatic.

Using the techniques of Khot and O’Donnell, [KO09], Theorem 1.2.5 implies:

**Theorem 1.2.11.** There exists a constant $c > 0$, such sufficiently small $\varepsilon > 0$, given a graph $G$ it is NP-hard to distinguish between:

- **YES case**: $G$ has a cut containing at least $\frac{1}{2} + \varepsilon$ fraction of the edges.
- **NO case**: any cut in $G$ has fractional size at most $\frac{1}{2} + c \log(1/\varepsilon)$.

**Unique-Games with Completeness close to 1.** It is well known that hardness of $\text{GapUG}[1 - \varepsilon, 1 - \varepsilon^\alpha]$ for a sufficient small $\alpha$ is equivalent to the Unique-Games Conjecture (the value needed for $\alpha$ is determined by quantitative version of the Parallel Repetition Theorem used; by Rao’s version [Rao11] any $\alpha < \frac{1}{2}$ is sufficient). However, there is a large discrepancy between this result and the best hardness result known due to Håstad et al. [HHM+17], who showed that $\text{GapUG}[1 - \varepsilon, 1 - C\varepsilon]$ is NP-hard for any $C < \frac{11}{8}$. By a trivial gap-shift, Theorem 1.2.5 implies the following modest improvement.

**Theorem 1.2.12.** For every $\varepsilon > 0$, $\text{GapUG}[1 - \varepsilon, 1 - 2\varepsilon + \varepsilon^2]$ is NP-hard.
Intermediate Complexity Conjecture. Barak [Bar17] proposed the Intermediate Complexity Conjecture, stating that there exist 2-Prover-1-Round Games (also called CSP’s) whose complexity is sub-exponential. That is, there are $0 < \alpha < \beta < 1$, $0 < s < c < 1$ such that given an instance of the CSP, the task of distinguishing between the case in which the value of the instance is at least $c$, and the case in which it is at most $s$, can be done in time $2^{n^\alpha}$ and not in time $2^{n^\beta}$. Our result, together with the sub-exponential time algorithm for Unique-Games [ABS15], imply that the Exponential Time Hypothesis [IP01] implies the Intermediate Complexity Conjecture:

**Theorem 1.2.13.** Assuming the Exponential Time Hypothesis, for every $\delta > 0$ there are $0 < \alpha < \beta < 1$ such GapUG[$\frac{1}{2}$, $\delta$] can be solved in time $2^{n^{\alpha}}$, but not in time $2^{n^{\beta}}$.

Lassere Integrality Gaps with Perfect Completeness. The starting point of our reduction to 2-to-2 games is the 3Lin problem. Very strong Integrality gaps for polynomially many rounds of the Lassere Hierarchy for the 3Lin problem are known [Gri01, Sch08], and these translate, by our reduction, into integrality gaps for 2-to-2 games with perfect completeness. In particular, in the hardness result for the Coloring problem, the SDP program has value indicating that the graph is 4 colorable, while the graph does not even have an independent set of size $\delta n$.

Unique Games Conjecture versus the Small Set Expansion Conjecture Raghavendra and Steurer [RS10] proposed the Small Set Expansion Conjecture and showed that it implies the Unique Games Conjecture. Roughly speaking, it states that GapSSE($\epsilon, 1 - \epsilon$), the problem of distinguishing whether a graph has a small set of expansion at most $\epsilon$ or whether every small set has expansion at least $1 - \epsilon$, is NP-hard.

The 2-to-2 Games Theorem arguably supports the suspicion that the Unique Games Conjecture may be correct while the Small Set Expansion Conjecture may be incorrect. An informal reasoning is as follows.

Raghavendra and Steurer give a reduction from GapSSE[$\epsilon$, $1 - \epsilon$] to GapUG[$1 - \epsilon'$, $\epsilon'$]. The same reduction also shows that GapSSE[$\beta$, $1 - \epsilon$] reduces to GapUG[$\frac{1}{2}$, $\epsilon'$] for some absolute constant $\beta$ (say $\beta = \frac{3}{4}$). If one were to show that the latter problem is NP-hard without concluding anything about the former, that may support the above suspicion. Indeed, this is precisely what happens in the proof of the 2-to-2 Games Theorem. One gets a reduction to GapUG[$\frac{1}{2}$, $\epsilon'$] without getting a reduction to Gap SSE; the graphs in the reduction always have small non-expanding sets.

2-to-1 Games. $d$-to-1 Games are the bipartite variant of $d$-to-$d$ Games. An instance of $d$-to-1 Games consists of a bipartite graph $(L \cup R, E)$, alphabets $\Sigma_L, \Sigma_R$ where $|\Sigma_L| = d |\Sigma_R|$ and $d$-to-1 constraints $\Phi = \{\phi_e\}_{e \in E}$. A constraint $\phi_e : \Sigma_L \to \Sigma_R$ is called $d$-to-1 if every $a \in \Sigma_R$ has exactly $d$ pre-images.

Hardness of 2-to-1 Games between $1 - \epsilon$ and $\epsilon$ can be achieved with relatively simple adjustments of the construction used in the proof of Theorem 1.2.6. This is not discussed further herein.

**Source material.** This chapter is based on the papers [KMS17, DKK+18a]. We chose to present the construction, as well as the soundness analysis slightly differently. The material presented in the chapter includes the reduction as well as its correctness analysis, relying on a result on the Grassmann Graph (Theorem 3.2.7) proved in the next chapter.

---

4Actually, the weaker hypothesis that $\text{NP} \not\subseteq \bigcap_{\epsilon > 0} \text{Dtime}(2^{n^{\epsilon}})$ suffices.
1.3 Pseudo-random Sets in the Grassmann Graph Have Near-Perfect Expansion (Chapter 4)

A key component in the proof of the 2-to-2 conjecture (with imperfect completeness) is the analysis of a 2-to-2 test on the Grassmann Graph (Theorem 3.2.7). For a vector space $V$ over $\mathbb{F}_2$ and a parameter $\ell \ll \dim(V)$, the vertices of the Grassmann Graph $G(V, \ell)$ are all $\ell$-dimensional subspaces of $V$,

$$U = \{ L \subseteq V \mid \dim(L) = \ell \},$$

and the edge set is

$$E = \{ (L, L') \mid \dim(L \cap L') = \ell - 1 \},$$

in words $L, L'$ are adjacent if their intersection is of dimension one less than theirs.

The graph $G(V, \ell)$ is naturally associated with a code: any linear function $f : V \to \mathbb{F}_2$ has a corresponding codeword $F$ over the alphabet $\mathbb{F}_2^n$, that assigns to each vertex of the Grassmann Graph a linear function: $F[L] = f|_L$.

The edges of the graph are interpreted as local consistency checks, that test if the values assigned to the vertices correspond to some global linear function $f$. More precisely, for a given word $A : U \to \mathbb{F}_2^n$ (not necessarily a codeword), the test picks an edge $(L, L')$ uniformly at random, and tests if $A[L]|_{L \cap L'} = A[L']|_{L \cap L'}$. Clearly, if the word $A$ is a legal codeword of the Grassmann Code, then this test passes with probability 1. The interesting question is: what can be said about words $A$ that pass the test with noticeable probability $\delta > 0$? It turns out that the most obvious speculations about such assignments $A$ is false (for more details, please see Section 3.2 in Chapter 3), but there are more complex speculations that are plausible. Importantly, the examples showing that the most obvious speculations fail suggest that the question of understanding the test is intimately related to the structure of sets in the graph whose expansion is bounded away from 1. This is the main question of interest of this chapter.

**Definition 1.3.1.** Let $G = (U, E)$ be a $d$-regular graph, and let $S \subseteq U$ be a subset of vertices. The edge expansion of $S$ is defined as

$$\Phi_G(S) = \frac{|E(S, U \setminus S)|}{d |S|},$$

where $E(S, U \setminus S)$ is the set of edges going from $S$ to $U \setminus S$. In words, the expansion of the set $S$ is fraction of edges that go outside $S$ among the edges that touch $S$.

Usually, a graph $G$ is called an $\gamma$-expander if $\Phi_G(S) \geq \gamma$ for all sets $S$ containing at most half of the vertices, where $\gamma$ is thought of as a positive constant, usually small. We will be interested in sets that have near-perfect expansion, i.e. sets for which $\Phi_G(S) \geq 1 - \delta$, where $\delta > 0$ is a small constant. We often drop the subscript $G$ when the graph $G$ is clear from context.

The question of which sets have near perfect expansion in a graph only makes sense for small sets, since large sets naturally contain a sizable fraction of the edges. We quantify the size of a set by its density, i.e. its fractional size with respect to the set of all vertices. The following fact is proved using Cauchy-Schwarz Inequality:

**Fact 1.3.2.** Let $S$ be a set of vertices of $G(V, \ell)$ of density $\delta$. Then $\Phi(S) \leq 1 - \delta + 2^{\ell-k}$.  

---

5 Indeed, let $G'$ be the Grassmann Graph with self-loops, where the probability to traverse from a vertex $L$ to itself is $2^{\ell-k}$. Then $\Phi_G(S) = (1 - 2^{\ell-k})\Phi_{G'}(S) + 2^{\ell-k}$. We have $1 - \Phi_{G'}(S) = \Pr_{u \in S, v \in G'[u]} [v \in S] = 1/4 \Pr_{u \in G', v \in G'[u]} [u, v \in S]$. Sampling an edge in $G'$ is equivalent to sampling an $\ell - 1$ dimensional subspace $B$ and then two extensions of it $L, L'$, and the probability in question is equal to $\mathbb{E}_{B} [\mathbb{E}_{L, L' \supseteq B} | 1.L \subseteq S] \leq \mathbb{E}_{B} [\mathbb{E}_{L \supseteq B} | 1.L \subseteq S] \geq \mathbb{E}_{B} [\mathbb{E}_{L \supseteq B} | 1.L \subseteq S]^2 = \delta^2$. Substituting we get that $\Phi_{G'}(S) \leq 1 - \delta$, and therefore the claim of the fact.
Thus, we have a certain limitation to how perfect of an expansion we can expect from any set of density $\delta$. Is this tight for all sets of density $\delta$? I.e., are there any sets of density at most $\delta$, for which $\Phi(S) \ll 1 - \delta$? The answer to this questions is yes, as evidenced by the following examples.

1. Take any $x \in V \setminus \{x\}$, and consider $S_x = \{L \subseteq V \mid x \in L\}$. Then clearly, as long as $\ell \ll \dim(V)$, the density of $S_x$ is $o(1)$. However, the expansion of $S_x$ is roughly half: for any $L \in S_x$, what is the probability that a random neighbour $L'$ of it lies inside $S_x$? This is precisely the probability that $x \in L'$; note that since $L, L'$ intersect on dimension $\ell - 1$, they share half of their vectors, and by symmetry the probability that $x$ is one of these shared vectors is $\approx \frac{1}{2}$. Thus, for any $L \in S_x$ roughly half of the edges incident to it end up in $L' \in S_x$, and the expansion of $S_x$ is $\approx \frac{1}{2}$.

2. Take any hyperplane $W \subseteq V$, and consider $S_W = \{L \subseteq V \mid L \subseteq W\}$. Again, as long as $\ell \ll \dim(V)$, we have that the density of $S_W$ is $o(1)$. For any $L \in S_W$, the probability that a random neighbour $L'$ is also in $S_W$ is $\approx \frac{1}{2}$: a random neighbour $L'$ can be sampled by taking an $\ell - 1$ dimensional subspace $A \subseteq L$, and then extending it to $L' = A \oplus \text{Span}(y)$ for a randomly chosen $y \in V \setminus L$. The probability $L' \in S_W$, is the probability that the chosen $y$ is in $W$, which is $\approx \frac{1}{2}$ since $W$ contains roughly half of the vectors of $V$.

Are there any other examples of sets with far-from-perfect expansion? Obviously, one could take one such example and perturb it a little without affecting the expansion very much – however these examples are essentially the same. One could also compose such examples: e.g. for non zero vector $x$, such example and perturb it a little without affecting the expansion very much – however these examples are essentially the same. One could also compose such examples: e.g. for non zero vector $x$, and hyperplane $W$ containing $x$, the set $\{L \mid x \in L \subseteq W\}$ has expansion nearly $\frac{3}{4}$. So, we must somehow quantify when a given set of vertices $S$ is “close” or “far” to such examples. This is captured by the notion of a pseudo-random sets.

**Definition 1.3.3.** A set of vertices $S$ is $(r, \varepsilon)$ pseudo-random ($r \in \mathbb{N}, \varepsilon > 0$) if for any subspaces $Q \subseteq W \subseteq V$ such that $\dim(Q) + \text{codim}(W) \leq r$, the density of $S$ in the induced subgraph on $\{L \subseteq V \mid Q \subseteq L \subseteq W\}$ is at most $\varepsilon$, i.e.,

$$\frac{|S \cap \{L \mid Q \subseteq L \subseteq W\}|}{|\{L \mid Q \subseteq L \subseteq W\}|} \leq \varepsilon.$$

The main result of Chapter 4 is a characterization stating that small sets whose expansion is far from perfect must be non pseudo-random (as are the examples that we described previously). Counter-positively, this results states:

**Theorem 1.3.4.** For every $\eta > 0$ there are $r \in \mathbb{N}, \varepsilon > 0$, such that any large enough $\ell$, large enough $k$, if $S \subseteq G(V, \ell)$ is $(r, \varepsilon)$ pseudo-random set ($\dim(V) = k$) we have $\Phi(S) \geq 1 - \eta$.

This theorem, which is the main result of this chapter, implies by a result of Barak et al. [BKS18] Theorem 3.2.7, thereby completing the proof of Theorem 1.2.6. For completeness, the chapter also includes the reduction due to Barak et al. [BKS18].

**Source material.** This chapter is mainly based on the papers [DKK+18b, KMS18]. The paper [DKK+18b] obtained partial results towards Theorem 1.3.4 – more specifically form $\eta > \frac{1}{4}$ and introduced some analytical machinery for the Grassmann Graph. To study the case $\eta \leq \frac{1}{4}$, an analogous question was analyzed in [KMMS18] for the Johnson Graph where an analog of Theorem 1.3.4 can be proven, giving some insight into the case of the Grassmann Graph. Finally, the paper [KMS18] significantly extended the analytical techniques for the Grassmann Graph, and actually worked with a different but related graph. This paper proved 1.3.4 (in its full generality).

Since the main focus of this chapter is the completion of the proof of the 2-to-2 conjecture, we do not present much of the material from [DKK+18b, KMMS18] and instead focus on presenting a proof for Theorem 1.3.4.
1.4 The Biased Long Code and Hardness of Vertex-Cover (Chapter 5)

In Chapter 5 we prove Theorem 1.2.7 and Corollary 1.2.8 that give new gaps of hardness for the Independent Set and Vertex Cover problem. The reduction as well as the analysis are along the lines of [DS05, Kho02], and is included here for sake of completeness.

1.5 Conclusions and Open Problems (Chapter 6)

In Chapter 6 we conclude the main results of this thesis and mention several open problems.
Chapter 2

Monotonicity Testing and Directed Isoperimetric Inequalities

2.1 Introduction

In this chapter, we study the problem of testing whether a given Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is monotone. We also study certain isoperimetric type theorems on the boolean hypercube that are closely related. Our main results are: (1) a directed and robust analogue of a theorem of Talagrand [Tal93], generalizing prior related theorems and (2) a monotonicity tester that is optimal in terms of query complexity (see Section 2.1.5 for subtle issues regarding its optimality).

2.1.1 Boolean Isoperimetric Type Theorems

Given a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, define the variance of the function as $\text{var}(f) = p(1 - p)$ where $p = \Pr_x[f(x) = 1]$. Let $S_f$ denote the set of sensitive edges, i.e. the set of pairs $(x, y)$ such that $x, y \in \{0, 1\}^n$ differ in exactly one coordinate, $f(x) = 1$ and $f(y) = 0$. Let $I_f = \frac{|S_f|}{2^n}$ denote the “total influence” of the function. A folklore theorem states:\footnote{A Fourier analytic proof: $\text{var}(f) = \sum_{S \subseteq \{1, \ldots, n\}, S \neq \phi} f(S)^2$ whereas $I_f = 2 \cdot \sum_{S \subseteq \{1, \ldots, n\}, S \neq \phi} f(S)^2 \cdot |S|$.}

**Theorem 2.1.1.**

$$I_f \geq \Omega(\text{var}(f)).$$

The parameter $I_f$ reflects the size of the edge boundary of the function $f$ (or more precisely of the subset $\{x \in \{0, 1\}^n | f(x) = 1\}$ of the hypercube). The size of the vertex boundary $\Gamma_f$ is defined as

$$\Gamma_f = \frac{1}{2^n} \cdot |\{x | f(x) = 1, \exists (x, y) \in S_f\}|.$$

Margulis [Mar74] shows that the size of the edge boundary and that of the vertex boundary cannot both be small. Specifically,

**Theorem 2.1.2.**

$$I_f \cdot \Gamma_f \geq \Omega(\text{var}(f)^2).$$

It is instructive to note that the inequality above is tight up to a constant factor, as shown by a dictatorship function as well as the majority function. Both functions have a constant variance. For the dictatorship function, both $I_f$ and $\Gamma_f$ are $\Theta(1)$. For the majority function, $I_f = \Theta(\sqrt{n})$ and $\Gamma_f = \Theta(\frac{1}{\sqrt{n}})$.
For \( x \in \{0, 1\}^n \), the sensitive edges incident on \( x \) are precisely the edges in \( S_f \) that are incident on \( x \). Let \( I_f(x) \) be equal to 0 if \( f(x) = 0 \) and equal to the number of sensitive edges incident on \( x \) if \( f(x) = 1 \). Talagrand [Tal93] shows that:

**Theorem 2.1.3.**

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

It is easily seen that Theorem 2.1.1 is implied by Theorem 2.1.2 which in turn is implied by Theorem 2.1.3. For the former implication, one observes that (here \( I(x) \) denotes indicator of an event)

\[
I_f = \mathbb{E}_x [I_f(x)] \geq \mathbb{E}_x \left[ I_{f(x)>0} \right] = \Gamma_f.
\]

For the latter, one observes using Cauchy-Schwarz that

\[
I_f \cdot \Gamma_f = \mathbb{E}_x [I_f(x)] \cdot \mathbb{E}_x \left[ I_{f(x)>0} \right] \geq \mathbb{E}_x \left[ \sqrt{I_f(x)} \right]^2 \geq \Omega(\var(f)^2).
\]

### 2.1.2 Directed Analogues of Boolean Isoperimetric Type Theorems

A function \( h : \{0, 1\}^n \mapsto \{0, 1\} \) is called monotone if for any two inputs \( x \) and \( y \) where \( y \) is obtained by changing a coordinate of \( x \) from 0 to 1 it holds that \( h(x) = 1 \Rightarrow h(y) = 1 \). Equivalently, writing \( x \leq y \) to mean that \( x_i \leq y_i \) for every coordinate \( i \in \{1, \ldots, n\} \), \( f \) is monotone if and only if

\[
\forall x, y \in \{-1, 1\}^n, \quad x \leq y \Rightarrow f(x) \leq f(y).
\]

For a function \( f : \{0, 1\}^n \mapsto \{0, 1\} \), let \( \varepsilon(f) \) denote the distance of \( f \) from the class of monotone functions, i.e. minimum fraction of its values that need to be changed to turn \( f \) into a monotone function.

The monotonicity testing problem asks for an algorithm that queries a given function \( f \) at as few places and distinguishes whether the function is monotone or is far from being monotone (more on this later).

The problem has been very well-studied since late 1990s, on the boolean hypercube as well as over more general posets, for functions that take real values instead of boolean values, and also in the context of related problems such as estimating distance to monotonicity, approximating total influence and shortest path routing on the hypercube [GGL+00, CS13a, CST14, BCGM12, FLN+02, ACCL07, RRS+12, CS13b, CS13c, LR01]. Still, designing an optimal tester for boolean functions on the boolean hypercube (the most basic and interesting case in our opinion) remained open.

Let \( S_f^- \) denote the set of negatively sensitive edges, i.e. the set of pairs \((x, y)\) such that \( y \) is obtained by changing a single coordinate of \( x \) from 0 to 1 while \( f(x) = 1 \), \( f(y) = 0 \). These are precisely the edges that violate the monotonicity property. Let \( I_f^- = \frac{|S_f^-|}{2^n} \) be the “total negative influence”. Motivated by an application to the monotonicity testing problem, Goldreich et al [GGL+00] show that:

**Theorem 2.1.4.**

\[
I_f^- \geq \Omega(\varepsilon(f)).
\]

A hypercube can be thought of as a directed graph by orienting all its edges “monotonically upwards”. In hindsight, Theorem 2.1.4 is viewed as a “directed” analogue of Theorem 2.1.1, where \( I_f \) is replaced by its analogue \( I_f^- \) and \( \var(f) \) is replaced by its analogue \( \varepsilon(f) \). As far as we know, Chakrabarty and Seshadhri [CS13a] are the first to suggest this analogy. Also motivated by an application to the monotonicity testing problem, they show the following directed analogue of Margulis’ Theorem 2.1.2:
Theorem 2.1.5.

\[ I_f \cdot \Gamma_f \geq \Omega(\varepsilon(f)^2). \]

We note that again \( I_f \) is replaced by its analogue \( I_f^{-} \) (which sounds intuitive) and \( \text{var}(f) \) is replaced by its analogue \( \varepsilon(f) \) (which is not so intuitive, and hence quite remarkable, in our opinion). Lastly, \( \Gamma_f \) is replaced by its analogue \( \Gamma_f^{-} \), size of the negative vertex boundary, defined as:

\[ \Gamma_f^{-} = \frac{1}{2^n} \cdot \left| \{ x \mid f(x) = 1, \exists (x, y) \in S_f^{-} \} \right|. \]

For \( x \in \{0, 1\}^n \), the negatively sensitive edges incident on \( x \) are precisely the edges in \( S_f^{-} \) that are incident on \( x \). Let \( I_f^{-}(x) \) be equal to 0 if \( f(x) = 0 \) and equal to the number of negatively sensitive edges incident on \( x \) if \( f(x) = 1 \). Carrying the analogy between the undirected and directed case further and still motivated by an application to the monotonicity testing problem, we show a directed analogue of Talagrand’s Theorem 2.1.3:

Theorem 2.1.6.

\[ \mathbb{E}_x \left[ \sqrt{I_f^{-}(x)} \right] \geq \Omega \left( \frac{\varepsilon(f)}{\log n/\varepsilon} \right). \]

Note that unlike previous theorems, our lower bound has dependence on the dimension \( n \), which might just be an artifact of our proof method and not inherent. Indeed, Waingarten [Wai18] observed that the theorem above holds without the logarithmic factor: i.e. the right hand side can be replaced with \( \Omega(\varepsilon) \).

Just like the undirected case, it is easily observed that Theorem 2.1.6 implies Theorem 2.1.5 (up to the poly-log factor), which in turn implies Theorem 2.1.4. We note that even though an informal analogy holds between the theorems in the undirected and directed settings, the proofs in the directed setting are completely different and much more involved (as an aside, we do show that the theorems in the directed setting imply the corresponding theorems in the undirected setting and hence are more general, see Section 2.9.2). One difficulty is that the parameter \( \varepsilon(f) \) is not too friendly to work with (as opposed to its analogue \( \text{var}(f) \)). In particular, there is no straightforward way to characterize or estimate \( \varepsilon(f) \). The Proofs of Theorems 2.1.4, 2.1.6 proceed in reverse: assuming an upper bound on the L.H.S. of the respective inequality, one gives a sequence of transformations that turns the given function \( f \) into a monotone function and hence upper bounding \( \varepsilon(f) \). The proof of Theorem 2.1.5 is combinatorial and uses the characterization of the distance from monotonicity by the size of maximal matching in the violation graph.

We also remark that our proof of Theorem 2.1.6 is very different from that of Theorems 2.1.4 and 2.1.5 and involves several new technical ingredients that might be useful towards further research. In particular, our proof does not use routing schemes on the hypercube as in [LR01, CS13a] and instead relies on a new “split operator” on functions. The split operator and its properties are presented in Section 2.3 and the main proof appears in Section 2.4. The proof involves applying the split operator on random restrictions of \( f \).

Towards an application to the monotonicity testing problem, Chakrabarty and Seshadhri [CS13a] actually need and prove a stronger form of Theorem 2.1.5. Let \( \Gamma_{f, \text{matching}}^{-} \) denote the size of the maximum matching among the edges in \( S_f^{-} \) (divided by a normalizing factor of \( 2^n \)), which is clearly at most \( \Gamma_f^{-} \) since the endpoints \( x \) of the matching with \( f(x) = 1 \) are also points on the negative vertex boundary.

Theorem 2.1.7 ([CS13a]).

\[ I_f^{-} \cdot \Gamma_{f, \text{matching}}^{-} \geq \Omega(\varepsilon(f)^2). \]

In this chapter, we are faced with a similar issue. We do not know how to use Theorem 2.1.6 directly towards an application to the monotonicity testing problem. Also, we do not know how to deduce Theorem 2.1.7 from Theorem 2.1.6. However it turns out that a “robust” version holds both for Theorem 2.1.3 (i.e.
the undirected case) and Theorem 2.1.6 (i.e. the directed case). The latter is now enough for our application to the monotonicity testing problem and if one wishes, to deduce Theorem 2.1.7 (up to the poly-log factor). Since the specific robust version wasn’t considered before, we first describe it in an undirected setting.

2.1.3 Robust version of Talagrand’s Theorem

The robust version concerns the scenario when the sensitive edges are colored with two colors, red or blue. Let \( \text{col} : S_f \mapsto \{\text{red, blue}\} \) be an arbitrary 2-coloring of the edges in \( S_f \). For \( x \in \{0, 1\}^n \), let \( I_{f, \text{red}}(x) \) be equal to 0 if \( f(x) = 0 \) and equal to the number of red sensitive edges incident on \( x \) if \( f(x) = 1 \). For \( y \in \{0, 1\}^n \), let \( I_{f, \text{blue}}(y) \) be equal to 0 if \( f(y) = 1 \) and equal to the number of blue sensitive edges incident on \( y \) if \( f(y) = 0 \). The robust version of Talagrand’s Theorem 2.1.3 is as follows:

**Theorem 2.1.8.** For a function \( f : \{0, 1\}^n \mapsto \{0, 1\} \) and an arbitrary coloring \( \text{col} : S_f \mapsto \{\text{red, blue}\} \),

\[
\mathbb{E}_x \left[ \sqrt{I_{f, \text{red}}(x)} \right] + \mathbb{E}_y \left[ \sqrt{I_{f, \text{blue}}(y)} \right] \geq \Omega(\text{var}(f)).
\]

We note that this theorem implies Theorem 2.1.3 by considering the coloring that colors all sensitive edges red. The theorem is proved by adapting Talagrand’s proof appropriately, see Section 2.2. Our presentation is a bit different (in addition to being a proof of the more general robust version) and more reader-friendly in our opinion. Also, the theorem is needed in the proof of the robust version of the directed analogue of Talagrand’s Theorem (i.e. of Theorem 2.1.6), stated next.

2.1.4 A Robust and Directed Analogue of Talagrand’s Theorem

We finally state the robust and directed analogue of Talagrand’s Theorem, which is what we really need towards an application to the monotonicity testing problem.

As before, let \( S_f^- \) denote the set of negatively sensitive edges. The robust version concerns the scenario when the negatively sensitive edges are colored with two colors, red and blue. Let \( \text{col} : S_f^- \mapsto \{\text{red, blue}\} \) be an arbitrary 2-coloring of the edges in \( S_f^- \). For \( x \in \{0, 1\}^n \), let \( I_{f, \text{red}}^-(x) \) be equal to 0 if \( f(x) = 0 \) and equal to the number of red negatively sensitive edges incident on \( x \) if \( f(x) = 1 \). For \( y \in \{0, 1\}^n \), let \( I_{f, \text{blue}}^-(y) \) be equal to 0 if \( f(y) = 1 \) and equal to the number of blue negatively sensitive edges incident to \( y \) if \( f(y) = 0 \). The robust and directed analogue of Talagrand’s Theorem is as follows:

**Theorem 2.1.9.** For a function \( f : \{0, 1\}^n \mapsto \{0, 1\} \) and an arbitrary coloring \( \text{col} : S_f^- \mapsto \{\text{red, blue}\} \),

\[
\mathbb{E}_x \left[ \sqrt{I_{f, \text{red}}^-(x)} \right] + \mathbb{E}_y \left[ \sqrt{I_{f, \text{blue}}^-(y)} \right] \geq \tilde{\Omega}(\varepsilon(f)).
\]

Again the precise bound is \( \Omega\left(\frac{\varepsilon(f)}{\log n + \log(1/\varepsilon(f))}\right) \). This theorem is proved by combining (part of) proof of Theorem 2.1.6 along with a careful manipulation of underlying edge-coloring and the undirected robust version, i.e. Theorem 2.1.8. The theorem implies Theorem 2.1.6 by considering a coloring that colors all negatively sensitive edges red. It also implies Theorem 2.1.7 (up to the poly-log factor), see Section 2.9.1.

2.1.5 Monotonicity Testing

As mentioned before, the monotonicity testing problem asks for a randomized algorithm that queries a given function \( f : \{0, 1\}^n \mapsto \{0, 1\} \) at a few places and distinguishes whether the function is monotone or is far from being monotone. Let us focus on the case when the tester is non-adaptive, has perfect completeness
and is a “pair tester” (all testers studied, including the one herein, have all three properties). Here nonadaptive means that the queries of the tester do not depend on the answers to the previous queries. Perfect completeness means that a monotone function must be accepted with probability 1. A “pair tester” picks a pair of inputs \((x, y)\) from a pre-determined distribution such that \(y\) is monotonically above \(x\) and rejects if a violation to monotonicity is detected, i.e. if \(f(x) = 1\) and \(f(y) = 0\). For a pair tester, a measure of its quality is its rejection probability \(\text{rej}(n, \varepsilon(f))\) expressed in terms of \(n\) and the distance of \(f\) from the class of monotone functions. If one desires, one can (non-adaptively) repeat a pair tester \(\frac{1}{\text{rej}(n, \varepsilon(f))}\) times and achieve a constant rejection probability. Thus, the number of queries is often expressed as \(\frac{1}{\text{rej}(n, \varepsilon(f))}\), with a constant rejection probability as the stated goal.

Goldreich et al [GGL +00] present a pair tester that picks a uniformly random edge \((x, y)\) of the hypercube (i.e. \(x\) and \(y\) differ in one coordinate). This is referred to as an “edge tester”. The rejection probability is exactly \(\frac{f}{n}\) and hence \(\Omega\left(\frac{\varepsilon(f)}{n}\right)\) by their Theorem 2.1.4. Chakrabarty and Seshadhri [CS13a] present a pair tester that picks a number \(\tau \in \{1, 2, \ldots, \sqrt{n}\}\) with a certain distribution and then a pair \((x, y)\) is picked, roughly uniformly, so that \(y\) is monotonically above \(x\) by a distance \(\tau\). This is referred to as a “path tester” and its rejection probability is \(\tilde{\Omega}\left(\frac{\varepsilon(f)^{3/2}}{n^{7/6}}\right)\). As far as dependence on \(n\) is concerned, this is the first improvement over the work of Goldreich et al [GGL +00], further improved to \(\tilde{\Omega}\left(\frac{\varepsilon(f)^{2}}{n^{7/6}}\right)\) by Chen et al [CST14]. The analysis of the tester relies on their Theorem 2.1.5.

Equipped with our Theorem 2.1.9, we present and analyze a path tester\(^7\) whose rejection probability is \(\tilde{\Omega}\left(\frac{\varepsilon(f)^{2}}{n^{5/6}}\right)\):

**Theorem 2.1.10.** Given a function \(f : \{0, 1\}^n \mapsto \{0, 1\}\), there is a path tester that is non-adaptive, has perfect completeness and rejection probability \(\tilde{\Omega}\left(\frac{\varepsilon(f)^{2}}{n^{5/6}}\right)\).

In the following sections, we elaborate a bit on how Theorem 2.1.9 leads to a monotonicity tester, and then comment on the optimality of our tester.

**Monotonicity Testing from Good Subgraphs**

Given a function \(f : \{0, 1\}^n \mapsto \{0, 1\}\), let \(G_f(V, W, E)\) denote the bipartite graph of negatively sensitive edges, i.e. \(V, W \subseteq \{0, 1\}^n, \forall x \in V f(x) = 1, \forall y \in W f(y) = 0\), \(E\) is precisely the set of negatively sensitive edges \(S_f\), and every vertex in \(V \cup W\) has at least one negatively sensitive edge incident to it.

Roughly speaking, Chakrabarty and Seshadhri [CS13a] use their Theorem 2.1.5 to deduce that the graph \(G_f(V, W, E)\) has a large matching and relate the rejection probability of their test to the size of this matching. We, on the other hand, use our Theorem 2.1.9 to deduce that the graph \(G_f(V, W, E)\) has a “\((K, d)\)-good subgraph” with appropriate parameters \(K\) and \(d\) (a matching corresponds to the case \(d = 1\) and then \(K\) is the size of the matching). We analyze our tester with reference to this good subgraph. Here, a bipartite graph \(G'(V', W', E')\) is called \((K, d)\)-good if \(|W'| = K\), every vertex in \(W'\) has degree \(d\) and every vertex in \(V'\) has degree at most \(2d\) (or the symmetric case with the roles of the two sides of the bipartite graph reversed). Leaving out some important details and caveats, the analysis of our tester is informally stated as:

**Theorem 2.1.11.** *(Informal)* If for a function \(f : \{0, 1\}^n \mapsto \{0, 1\}\), the graph \(G_f(V, W, E)\) has a \((\sigma \cdot 2^n, d)\)-good subgraph, then there is a pair tester with rejection probability \(\tilde{\Omega}\left(\frac{\varepsilon(f)^{2}}{n^{5/6}}\right)\).

---

\(^7\)Our path tester chooses \(\tau\) uniformly from \(\{1, 2, 4, 8, \ldots, 2^\left\lfloor \frac{\log n}{2} \right\rfloor\}\), attempting to guess the “correct” value for \(\tau\). A similar guess is made in [CST14] and their distribution of \(\tau\) is, morally speaking, the same as ours. In [CS13a], \(\tau\) is chosen uniformly from \(\{1, 2, 3, \ldots, n^{1/8}\}\) with probability \(\frac{1}{n}\) and \(\tau = 1\) with probability \(\frac{1}{n}\). This apparent difference, however, is only because the authors did not try to guess \(\tau\), which is later fixed in [CST14].
We use Theorem 2.1.9 to deduce that $G_{f}^{-}(V, W, E)$ has a $(\sigma \cdot 2^n, d)$-good subgraph with $\sigma \sqrt{d} \geq \tilde{\Omega}(\varepsilon(f))$ (see Section 2.6.1 for the necessary combinatorial argument). Combined with the informal statement of our tester above, we get a tester with rejection probability $\tilde{\Omega}(\frac{\varepsilon(f)^2}{n})$ as claimed. Additional new ingredients used are bounds on the total influence of the function and on the fraction of “non-persistent” inputs (see Sections 2.6.2 and 2.6.3 respectively). We would like to emphasize that the analysis of our tester is qualitatively different and a bit simpler than that of Chakrabarty and Seshadhri [CS13a]. We do not elaborate this point further, but as a demonstration, we note (omitting the proof) that just using the large matching as in [CS13a] as a good subgraph, we already get a tester with rejection probability $\tilde{\Omega}(\frac{2^{3/3}}{n^{6/6}})$, improving the bound in both [CS13a, CST14].

**Upper Bounds on the Total Influence**

It is well known that if a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is monotone, then the total influence $I_f$ of $f$ is at most $O(\sqrt{n})$. Suppose now that a function $f$ is $\varepsilon$-close to monotone; in this case, one can only give the trivial $O(\sqrt{n} + \varepsilon n)$ upper bound on the total influence of $f$, which can be seen to be tight by taking a monotone function with total influence $\Theta(\sqrt{n})$ (such as the majority function), and change it randomly in $\varepsilon$ fraction of the inputs. For such function however, testing monotonicity can be done using small number of queries, by simply running the edge tester (since this function clearly has $I_f^{-} = \Omega(\varepsilon n)$, and the edge tester rejects with probability $\frac{I_f^{-}}{n}$).

An important ingredient in our analysis is the observation that if the edge tester doesn’t work well enough, then the function has small total influence. More precisely, we show that if $I_f^{-} = O(\sqrt{n})$, then $I_f = O(\sqrt{n})$, and in particular $I_f^{+} = O(\sqrt{n})$.

**Lower Bounds for Monotonicity Testing**

We now give an overview of lower bounds on the number of queries required by a monotonicity testing algorithm and compare our tester against these (and new) lower bounds. Towards a uniform comparison of known bounds, for a parameter $\varepsilon$, let us require that a tester rejects any function that is $\varepsilon$-far from being monotone with a constant probability. Seemingly, the dependence of the number of queries on the two parameters $n$ and $\varepsilon$ can be traded against each other, so the situation is a bit subtle.

Let us first consider the case of pair testers that are non-adaptive and have perfect completeness (the most interesting case in our opinion, especially since all known testers are of this kind). The tester of Goldreich et al [GGL+00] achieves $O(n/\varepsilon)$ queries and Briet et al [BCGM12] show that if a pair tester has $F(n)/\varepsilon$ query complexity, then the dependence on $n$ must be $F(n) \geq \Omega(n)$. We show that if a pair tester makes $O(n^{\alpha}/\varepsilon^{\beta})$ queries, then $\alpha + \frac{\beta}{2} \geq \frac{3}{2}$. This follows from:

**Theorem 2.1.12.** For $\varepsilon = \Theta(1/\sqrt{n})$, a pair tester that is non-adaptive, has perfect completeness and rejects a function that is $\varepsilon$-far from being monotone with constant probability must make $\Omega(n^{3/2})$ queries.

We note that for any $\alpha, \beta \geq 0$ such that $\alpha + \frac{\beta}{2} \geq \frac{3}{2}$, we have $\frac{n^{\alpha}}{\varepsilon^{\beta}} \geq \min \left\{ \frac{n}{\varepsilon}, \sqrt{\frac{n}{\varepsilon^{2}}} \right\}$. Hence, for any setting of $\alpha, \beta$, either the $O(n/\varepsilon)$-tester of Goldreich et al [GGL+00] or our $O(\sqrt{n}/\varepsilon^{2})$-tester performs as well as a potential $O(n^{\alpha}/\varepsilon^{\beta})$-tester. Thus, our tester in conjunction with Goldreich et al’s tester is optimal. Also, if only the dependence on $n$ is concerned (which is more interesting in our opinion than the dependence on $\varepsilon$), our tester is optimal even if compared against testers that possibly have imperfect completeness and not necessarily pair testers (see below).

Now we turn to more general testers, where there are still gaps between the upper and lower bounds. We already stated all the upper bounds before. We do not know a scenario where it helps to be adaptive, have
imperfect completeness, or not be a pair tester. From the lower bound side, if a tester is non-adaptive and has perfect completeness (but is not necessarily a pair tester), a lower bound of $\Omega(\sqrt{n})$ is shown by Fischer et al [FLN+02] for a constant $\varepsilon$. For non-adaptive testers that possibly have imperfect completeness, a lower bound of $\tilde{\Omega}(n^{1/5})$ is shown by Chen et al [CST14] for a constant $\varepsilon$ and further improved to $\Omega(n^{1/2-o(1)})$ by Chen et al [CDST15].

2.1.6 Followup works

Following our work, a $\tilde{\Omega}(n^{1/4})$ lower bound for adaptive testers was shown by Belovs and Blais [BB15a]. This was improved to a $\tilde{\Omega}(n^{1/3})$ by Chen, Waingarten and Xie [CWX17a], leaving a small gap for improvements via adaptive testers.

Theorem 2.1.9 (or rather a consequence of it, Lemma 2.7.1) was used by other works. Belovs and Blais [BB15b] showed that it is possible to test monotonicity using $O(n^{1/4}e^{-1/2})$ quantum queries, and Chen, Waingarten and Xie [CWX17b] showed an adaptive unateness tester that uses $O(n^{3/4}e^{-2})$ queries.

Finally Waingarten [Wai18] observed that Theorems 2.1.6, 2.1.9 hold without the logarithmic factor (namely, the right hand side in the inequalities in both theorems can be replaced with $\Omega(\varepsilon)$). Their proof is a surprisingly simple argument that accounts for the slack in inequality (2.6).

2.2 Proof of Theorem 2.1.8

In this section, we present a proof of Theorem 2.1.8. The proof is an easy adaptation of Talagrand’s proof [Tal93] of Theorem 2.1.3. Our presentation is a bit different and more reader-friendly in our opinion.

We recall that for a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, $S_f$ denotes the set of sensitive edges. Let col : $S_f \rightarrow \{\text{red}, \text{blue}\}$ be an arbitrary 2-coloring of the edges in $S_f$. For $x \in \{0, 1\}^n$, $I_{f, \text{red}}(x)$ is equal to 0 if $f(x) = 0$ and equal to the number of red sensitive edges incident on $x$ if $f(x) = 1$. For $y \in \{0, 1\}^n$, $I_{f, \text{blue}}(y)$ is equal to 0 if $f(y) = 1$ and equal to the number of blue sensitive edges incident on $y$ if $f(y) = 0$.\footnote{In other word, blue a sensitive edges contribute to the sensitivity of the 0-valued endpoint, and red sensitive edges contribute to the sensitivity of the 1-valued endpoint.} We intend to show that

$$
\mathbb{E}_x \left[ \sqrt{I_{f, \text{red}}(x)} \right] + \mathbb{E}_y \left[ \sqrt{I_{f, \text{blue}}(y)} \right] \geq \Omega(\text{var}(f)).
$$

(2.1)

The proof is by induction on $n$. We show that the L.H.S. of the inequality above is lower bounded by $\frac{1}{5} \cdot \text{var}(f)$. The claim is correct when $n = 1$. For any $n \geq 2$, let $f : \{0, 1\}^n \rightarrow \{0, 1\}$ be the given function and $f_0 = f(0, x_2, \ldots, x_n)$ and $f_1 = f(1, x_2, \ldots, x_n)$ be the two sub-functions on $n - 1$ coordinates. Let $\mathcal{H}_0$ and $\mathcal{H}_1$ denote the “bottom” and the “top” hypercubes on which the functions $f_0$ and $f_1$ are defined so that $\{0, 1\}^n = \mathcal{H}_0 \cup \mathcal{H}_1$. Let $\mathcal{I}$ be the set of sensitive edges along the first coordinate so that

$$
S_f = S_{f_0} \cup S_{f_1} \cup \mathcal{I}.
$$

Also, a given coloring of edges in $S_f$ induces a coloring of edges in $S_{f_0}$, $S_{f_1}$ and $\mathcal{I}$. Let $s = \frac{\lvert \mathcal{I} \rvert}{2^{n-1}}$ be the influence of the first coordinate. By Lemma 2.2.1,

$$
\text{var}(f) \leq \frac{1}{2} \cdot \text{var}(f_0) + \frac{1}{2} \cdot \text{var}(f_1) + \frac{1}{4} \cdot s^2.
$$

The main idea is to express the L.H.S. of Equation (2.1), denoted $T$, as a sum of four terms $A, B, C, D$ denoting similar quantities from functions $f_0$ and $f_1$ and in addition, a term $\Phi$ that corresponds to the “incremental
contribution” from the edges in $I$ (along the first coordinate). Towards this end, let

$$A = \mathbb{E}_x \left[ \sqrt{I_{f_0, \text{red}}(x)} \right], \quad B = \mathbb{E}_y \left[ \sqrt{I_{f_0, \text{blue}}(y)} \right],$$

$$C = \mathbb{E}_z \left[ \sqrt{I_{f_1, \text{red}}(z)} \right], \quad D = \mathbb{E}_w \left[ \sqrt{I_{f_1, \text{blue}}(w)} \right].$$

In the following $x, y$ will always denote vertices in the bottom hypercube $H_0$ and $z, w$ will always denote vertices in the top hypercube $H_1$. Also, $f(x) = f(z) = 1$ and we are concerned about the number of red edges incident on them. Similarly, $f(y) = f(w) = 0$ and we are concerned about the number of blue edges incident on them. We may assume, inductively, that

$$A + B \geq \frac{1}{8} \cdot \text{var}(f_0),$$

$$C + D \geq \frac{1}{8} \cdot \text{var}(f_1).$$

It is clear that $T$, i.e. the LHS. of Equation (2.1), equals

$$T = \frac{1}{2} (A + B) + \frac{1}{2} (C + D) + \Phi,$$

where $\Phi$ denotes the incremental contribution from edges in $I$ as follows. The edges in $I$ have the effect that for vertices $x, y \in H_0$ and $z, w \in H_1$, the relevant degrees $I_{f_0, \text{red}}(x), I_{f_0, \text{blue}}(y), I_{f_1, \text{red}}(z), I_{f_1, \text{blue}}(w)$ may increase by one. Each (colored) edge in $I$ leads to an increase in the degree of exactly one vertex $x, y, z$ or $w$. Partition the edges in $I$ as $I = I_A \cup I_B \cup I_C \cup I_D$ depending on whether an edge leads to an increase in the degree of $x, y, z$ or $w$ “type” of vertex. Defining $s_A = \frac{|I_A|}{2^n}$ and similarly $s_B, s_C, s_D$, we have $s = s_A + s_B + s_C + s_D$.

Let us focus on the edges in $I_A$. For $d = 1, 2, \ldots, n$, consider the vertices $x \in H_0$ such that the edges in $I_A$ increase their degree from $d - 1$ to $d$ and let the number of such vertices be $s_d \cdot 2^{n-1}$. Specifically, for such a vertex $I_{f_0, \text{red}}(x) = d - 1$ and $I_{f, \text{red}}(x) = d$ (when $x$ is viewed as a vertex from $\{0, 1\}^n = H_0 \cup H_1$). The incremental contribution of $x$ is:

$$\sqrt{I_{f, \text{red}}(x)} - \sqrt{I_{f_0, \text{red}}(x)} = \sqrt{d} - \sqrt{d - 1} \geq \frac{1}{4\sqrt{d}}.$$ 

We note that $s_A = \sum_{d=1}^n s_d$. Denoting by $\Phi_A$ the incremental contribution averaged over all $x \in H_0$, we get

$$\Phi_A \geq \sum_{d=1}^n s_d \cdot \frac{1}{4\sqrt{d}}.$$

We also note that (by considering the contribution to $T$ from only the $x$ “type” vertices)

$$T \geq \frac{1}{2} \cdot \sum_{d=1}^n s_d \sqrt{d}.$$

By Cauchy-Schwarz, we have

$$\Phi_A \cdot T \geq \frac{1}{2} \left( \sum_{d=1}^n s_d \cdot \frac{1}{4\sqrt{d}} \right) \left( \sum_{d=1}^n s_d \sqrt{d} \right) \geq \frac{1}{8} \cdot \left( \sum_{d=1}^n s_d \right)^2 = \frac{1}{8} \cdot s_A^2.$$
If \( T \geq \frac{1}{8} \) we are done, since we only intend to prove a lower bound of \( T \geq \frac{1}{8} \cdot \text{var}(f) \). Thus we may assume that \( \Phi_A \geq s_A^2 \). Similarly we may assume that \( \Phi_B \geq s_B^2 \), \( \Phi_C \geq s_C^2 \), \( \Phi_D \geq s_D^2 \) and

\[
\Phi = \frac{1}{2} \cdot (\Phi_A + \Phi_B + \Phi_C + \Phi_D) \geq \frac{1}{2} \cdot (s_A^2 + s_B^2 + s_C^2 + s_D^2) \geq \frac{1}{8} \cdot (s_A + s_B + s_C + s_D)^2 = \frac{1}{8} s^2.
\]

Finally, we get inductively that

\[
T = \frac{1}{2} (A + B) + \frac{1}{2} (C + D) + \Phi
\geq \frac{1}{2} \left( \frac{1}{8} \cdot \text{var}(f_0) \right) + \frac{1}{2} \left( \frac{1}{2} \cdot \text{var}(f_1) \right) + \frac{1}{8} \cdot s^2
\geq \frac{1}{8} \cdot \left( \frac{1}{2} \cdot \text{var}(f_0) + \frac{1}{2} \cdot \text{var}(f_1) + \frac{1}{4} \cdot s^2 \right)
\geq \frac{1}{8} \cdot \text{var}(f).
\]

This completes the inductive proof.

**Lemma 2.2.1.** For a function \( f : \{0, 1\}^n \mapsto \{0, 1\} \), let \( s \) denote the influence of the first coordinate and \( f_0 \) and \( f_1 \) denote the sub-functions on the remaining \( n - 1 \) coordinates. Then

\[
\text{var}(f) \leq \frac{1}{2} \cdot \text{var}(f_0) + \frac{1}{2} \cdot \text{var}(f_1) + \frac{1}{4} \cdot s^2.
\]

**Proof.** Let \( p_0 = \mathbb{E}_x[f_0(x)] \) and \( p_1 = \mathbb{E}_x[f_1(x)] \) so that \( \mathbb{E}_x[f(z)] = \frac{p_0 + p_1}{2} \). We have, by definition, \( \text{var}(f_0) = p_0(1 - p_0) \) and \( \text{var}(f_1) = p_1(1 - p_1) \). Thus we have the identity:

\[
\text{var}(f) = \frac{p_0 + p_1}{2} (1 - \frac{p_0 + p_1}{2}) = \frac{1}{2} \cdot \text{var}(f_0) + \frac{1}{2} \cdot \text{var}(f_1) + \frac{1}{4} (p_0 - p_1)^2.
\]

Observing that

\[
|p_0 - p_1| = |\mathbb{E}_x[f_0(x)] - \mathbb{E}_x[f_1(x)]| \leq \mathbb{E}_x[|f_0(x) - f_1(x)|] = s.
\]

finishes the proof. \( \square \)

### 2.3 The Switch and the Split Operators

Goldreich et al [GGL+00] define a “switch operator” \( S_i \) for a coordinate \( i \in \{1, \ldots, n\} \) that constructs a function \( S_i(f) : \{0, 1\}^n \mapsto \{0, 1\} \) from a given function \( f : \{0, 1\}^n \mapsto \{0, 1\} \). We define a new operator that we call a “split operator” \( \nabla_i \) for a coordinate \( i \in \{1, \ldots, n\} \) that constructs a function \( \nabla_i(f) : \{0, 1\}^{n+1} \mapsto \{0, 1\} \) from a given function \( f : \{0, 1\}^n \mapsto \{0, 1\} \). Note that \( \nabla_i(f) \) is a function of \( n + 1 \) coordinates. Both the operators \( S_i \) and \( \nabla_i \) are “applied” on coordinate \( i \) and can be sequentially applied on coordinates 1 through \( n \) in any desired order. The operators are non-commutative in the sense that the resulting function, in general, depends on the order in which the operators are applied on multiple coordinates. In this section, we define both the operators and prove several important properties of the latter.

#### 2.3.1 The Switch Operator

**Definition 2.3.1.** For a function \( f : \{-1, 1\}^n \mapsto \{-1, 1\} \) and a coordinate \( i \in \{1, \ldots, n\} \), the function \( S_i(f) : \{-1, 1\}^n \mapsto \{-1, 1\} \) is defined as follows. For each \( \alpha \in \{-1, 1\}^{i-1} \), \( \beta \in \{-1, 1\}^{n-i} \), and a coordinate variable \( x_i \in \{0, 1\} \),

\[
S_i(f)(\alpha, x_i, \beta) \overset{\text{def}}{=} \begin{cases} 
\min \{ f(\alpha, 0, \beta), f(\alpha, 1, \beta) \} & \text{if } x_i = 0 \\
\max \{ f(\alpha, 0, \beta), f(\alpha, 1, \beta) \} & \text{if } x_i = 1.
\end{cases}
\]
In the following, we will abuse notation for the sake of conciseness and convenience: with the understanding that for \( x_{-i} \in \{-1, 1\}^{n-1} \), written as \( x_{-i} = \alpha \circ \beta, \alpha \in \{-1, 1\}^{i-1}, \beta \in \{-1, 1\}^{n-i} \), and that the coordinate \( x_i \) is inserted in the \( i^{th} \) position as \( \alpha \circ x_i \circ \beta \), we will write the above definition of the switch operator as:

\[
S_i(f)(x_i, x_{-i}) \overset{def}{=} \begin{cases}
\min\{f(0, x_{-i}), f(1, x_{-i})\} & \text{if } x_i = 0 \\
\max\{f(0, x_{-i}), f(1, x_{-i})\} & \text{if } x_i = 1.
\end{cases}
\]

The switch operator considers the edges of the hypercube along the \( i^{th} \) dimension and for every edge that violates the monotonicity of \( f \) in that dimension, switches the values of the function at the two endpoints of that edge. The “corrected” function \( S_i(f) \) is now monotone along the \( i^{th} \) dimension. A remarkable property of the switch operator, as shown by Goldreich et al \([GGL+00]\), is that if another switch is now applied on coordinate \( j \neq i \), the resulting function (i.e. \( S_j(S_i(f)) \)) stays monotone along the \( i^{th} \) dimension. In particular, if the switch operator is applied on coordinates 1 through \( n \), one after another in some order, the final function is monotone. Another remarkable property is that the Hamming distance between a pair of functions can only decrease after applying the switch operator to both. We state and prove these properties below.

**Lemma 2.3.2.** If \( f : \{0, 1\}^n \mapsto \{-1, 1\} \) is monotone in the \( j^{th} \) coordinate, then so is \( S_i(f) \) for a coordinate \( i \in \{1, \ldots, n\}, i \neq j \).

**Proof.** It is enough to fix a setting \( x_{-i-j} \in \{-1, 1\}^{n-2} \) of coordinates except \( i \) and \( j \) and consider the behavior of four values

\[
a = f(x_i = 0, x_j = 0, x_{-i-j}), \quad b = f(x_i = 1, x_j = 0, x_{-i-j}),
\]

\[
c = f(x_i = 0, x_j = 1, x_{-i-j}), \quad d = f(x_i = 1, x_j = 1, x_{-i-j}).
\]

Monotonicity of \( f \) in the \( j^{th} \) coordinate implies that \( a \leq c, b \leq d \). After the switch on the \( i^{th} \) coordinate, the four values change to

\[
S_i(f)(0, 0, x_{-i-j}) = \min\{a, b\}, \quad S_i(f)(1, 0, x_{-i-j}) = \max\{a, b\},
\]

\[
S_i(f)(0, 1, x_{-i-j}) = \min\{c, d\}, \quad S_i(f)(1, 1, x_{-i-j}) = \max\{c, d\}.
\]

Now the monotonicity of \( S_i(f) \) in the \( j^{th} \) coordinate amounts to saying that

\[\min\{a, b\} \leq \min\{c, d\}, \quad \max\{a, b\} \leq \max\{c, d\},\]

which follows since \( \min\{\cdot, \cdot\}, \max\{\cdot, \cdot\} \) are monotone functions and \( (a, b) \leq (c, d) \) by hypothesis. \( \square \)

**Definition 2.3.3.** For functions \( f_1, f_2 : \{0, 1\}^n \mapsto \{-1, 1\} \), their Hamming distance is

\[
\Delta(f_1, f_2) \overset{def}{=} \mathbb{E}_x \left[ 1_{f_1(x) \neq f_2(x)} \right].
\]

**Lemma 2.3.4.** For functions \( f, h : \{0, 1\}^n \mapsto \{-1, 1\} \) and a coordinate \( i \in \{1, \ldots, n\} \),

\[
\Delta(S_i(f), S_i(h)) \leq \Delta(f, h).
\]

**Proof.** For any fixed setting of \( x_{-i} \in \{-1, 1\}^{n-1} \), we show that the inequality holds as far as contribution from that setting of \( x_{-i} \) is concerned. Indeed, denoting

\[
a = f(0, x_{-i}), \quad b = f(1, x_{-i}),
\]

\[
b = f(1, x_{-i}),
\]
\[ c = h(0, x_{-i}), \quad d = h(1, x_{-i}), \]

we see that the contribution to the RHS is

\[ 1_{a \neq c} + 1_{b \neq d}, \]

whereas the contribution to the LHS is

\[ 1_{\min\{a, b\} \neq \min\{c, d\}} + 1_{\max\{a, b\} \neq \max\{c, d\}}. \]

The inequality can be checked by case analysis. For example, if \( a = b \) then LHS = RHS. The same holds if \( c = d \) or if \( (a, b) = (c, d) \). The only remaining case is when \( (a, b) = (0, 1) \land (c, d) = (1, 0) \) (and the other way round) in which case the Hamming distance actually decreases. \( \square \)

We recall that for a function \( f \), the parameter \( \varepsilon(f) \) denotes the distance of \( f \) from the class of monotone functions, i.e. the minimum fraction of values of \( f \) that need to be changed to turn \( f \) into a monotone function. The parameter \( \varepsilon(f) \) seems difficult to characterize in a “constructive” manner. However it turns out that it can be well-approximated in a constructive manner. Goldreich et al [GGL+00] show, thanks to Lemma 2.3.2, that if the switch operator is applied to a function \( f \), on coordinates 1 through \( n \), say in that order, then the resulting function is monotone. We observe that this transformation is efficient in the sense that the fraction of values of \( f \) changed is at most 2 \( \varepsilon(f) \). This observation also appears in a paper of Fattal and Ron [FR10, Lemma 4.3], but several researchers (including us) seemed unaware of this (thanks to Andrej Bogdanov for pointing out).

**Lemma 2.3.5.** For any function \( f : \{-1, 1\}^n \rightarrow \{-1, 1\} \),

\[ \varepsilon(f) \leq \Delta(f, S_n(S_{n-1}(\ldots S_2(S_1(f)) \ldots))) \leq 2 \varepsilon(f). \]

**Proof.** By a repeated application of Lemma 2.3.2, the function \( S_n(S_{n-1}(\ldots S_2(S_1(f)) \ldots)) \) is monotone and hence the left inequality holds by definition of \( \varepsilon(f) \). Towards the right inequality, let \( h \) be a monotone function such that \( \Delta(f, h) = \varepsilon(f) \). By a repeated application of Lemma 2.3.4,

\[ \Delta(S_n(S_{n-1}(\ldots S_2(S_1(f)) \ldots)), S_n(S_{n-1}(\ldots S_2(S_1(h)) \ldots))) \leq \Delta(f, h). \]

We note however that since \( h \) is already monotone, applying switch operators keeps it unaffected. Thus, the above inequality is same as

\[ \Delta(S_n(S_{n-1}(\ldots S_2(S_1(f)) \ldots)), h) \leq \Delta(f, h) = \varepsilon(f). \]

Now by triangle inequality,

\[ \Delta(f, S_n(S_{n-1}(\ldots S_2(S_1(f)) \ldots))) \leq \Delta(f, h) + \Delta(S_n(S_{n-1}(\ldots S_2(S_1(f)) \ldots)), h) \leq 2 \varepsilon(f), \]

and the right inequality holds. \( \square \)

It will be useful for us to consider the scenario when the switch operator is applied to coordinates 1 through \( n \) in a random order. Let \( \rho \in S_n \) denote a permutation of \( \{1, 2, \ldots, n\} \) and \( \gamma(f) \) denote the expected fraction of values of \( f \) changed by selecting \( \rho \in S_n \) at random and then applying switches according to the order \( \rho \). Since in the proof of Lemma 2.3.5, we didn’t use the fact that the switches were in a specific order, it follows that \( \gamma(f) \) is sandwiched between \( \varepsilon(f) \) and 2 \( \varepsilon(f) \).

**Definition 2.3.6.** For a permutation \( \tau \in S_n \), \( \tau \circ f \) denotes the function \( S_{\tau(n)}(S_{\tau(n-1)}(\ldots S_{\tau(2)}(S_{\tau(1)}(f)) \ldots)) \).

\[ \gamma(f) \overset{\text{def}}{=} \mathbb{E}_{\tau \in S_n} [\Delta(f, \tau \circ f)]. \]

We have

\[ \varepsilon(f) \leq \gamma(f) \leq 2 \varepsilon(f). \]
2.3.2 The Split Operator

Now we define the split operator $\nabla_i$ applied on a coordinate $i \in \{1, \ldots, n\}$. For a function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$, the function $\nabla_i(f)$ is a function of $n + 1$ coordinates. It is best to think that the $i^{th}$ coordinate is now split into two coordinates indexed as $(i, +)$ and $(i, -)$.

**Definition 2.3.7.** For a function $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ and a coordinate $i \in \{1, \ldots, n\}$, the function $\nabla_i(f)$ is defined as follows. For each $x_{-i} \in \{-1, 1\}^{n-1}$

\[
\nabla_i(f)(x_{i+}, x_{i-}, x_{-i}) \overset{\text{def}}{=} \begin{cases} 
    f(0, x_{-i}) & \text{if } x_{i+} = 0, x_{i-} = 0 \\
    \min \{f(0, x_{-i}), f(1, x_{-i})\} & \text{if } x_{i+} = 0, x_{i-} = 1 \\
    \max \{f(0, x_{-i}), f(1, x_{-i})\} & \text{if } x_{i+} = 1, x_{i-} = 0 \\
    f(1, x_{-i}) & \text{if } x_{i+} = 1, x_{i-} = 1.
\end{cases}
\]

If the function $f$ is written as $f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)$, then the function $\nabla_i(f)$ is written as

\[
(\nabla_i(f))(x_1, \ldots, x_{i-1}, x_{i+}, x_{i-}, x_{i+1}, \ldots, x_n).
\]

In particular, the coordinates of $\nabla_i(f)$ have the same indices as that of $f$, except that the coordinate $i$ is split into two coordinates indexed as $(i, +)$ and $(i, -)$. We start with a preliminary observation about the split operator.
Lemma 2.3.8. Let \( f : \{-1, 1\}^n \to \{-1, 1\} \), and \( i, j \in \{1, \ldots, n\} \) be coordinates such that \( i \neq j \). Then:

1. \( \nabla_i(f) \) is monotone in coordinate \((i, +)\).
2. \( \nabla_i(f) \) is anti-monotone in coordinate \((i, -)\).
3. If \( f \) is monotone in coordinate \( j \), then so is \( \nabla_i(f) \).
4. If \( f \) is anti-monotone in coordinate \( j \), then so is \( \nabla_i(f) \).

Proof. We verify that \( \nabla_i(f) \) is monotone in coordinate \((i, +)\) by verifying its monotonicity for every fixing of \( x_{-i} \in \{0, 1\}^{n-1} \) and \( x_{-i} \in \{0, 1\} \). Fix some \( x_{-i} \in \{0, 1\}^{n-1} \). When \( x_{i,-} = 0 \), by definition,

\[
\nabla_i(f)(x_{i,+} = 0, x_{i,-} = 0, x_{-i}) = f(0, x_{-i}) \\
\nabla_i(f)(x_{i,+} = 1, x_{i,-} = 0, x_{-i}) = \max \{ f(0, x_{-i}), f(1, x_{-i}) \},
\]

and the first value is less or equal the second. When \( x_{i,-} = 1 \), by definition,

\[
\nabla_i(f)(x_{i,+} = 0, x_{i,-} = 1, x_{-i}) = \min \{ f(0, x_{-i}), f(1, x_{-i}) \} \\
\nabla_i(f)(x_{i,+} = 1, x_{i,-} = 1, x_{-i}) = f(1, x_{-i}),
\]

and again, the first value is less or equal the second. This confirms that \( \nabla_i(f) \) is monotone in coordinate \((i, +)\). Similarly we can confirm that \( \nabla_i(f) \) is anti-monotone in coordinate \((i, -)\).

Now suppose that \( f \) is monotone in coordinate \( j \). We will show that \( \nabla_i(f) \) is also monotone in coordinate \( j \). We write \( \nabla_i(f) \) as \( \nabla_i(f)(x_{i,+}, x_{i,-}, x_{j}, x_{-i-j}) \) where \( x_{-i-j} \) denotes all coordinates except \( i \) and \( j \). We verify monotonicity of \( \nabla_i(f) \) in coordinate \( j \) by going over all four fixings of \( x_{i,+} \) and \( x_{i,-} \). When \( x_{i,+} = x_{i,-} = 0 \), we have

\[
\nabla_i(f)(x_{i,+} = 0, x_{i,-} = 0, x_{j}, x_{-i-j}) = f(x_i = 0, x_j, x_{-i-j}),
\]

and monotonicity of \( \nabla_i(f) \) in coordinate \( j \) follows from that of \( f \). Similarly we can verify the case \( x_{i,+} = x_{i,-} = 1 \). Now consider the case \( x_{i,+} = 1 \) and \( x_{i,-} = 0 \). In this case,

\[
\nabla_i(f)(x_{i,+} = 1, x_{i,-} = 0, x_{j} = 0, x_{-i-j}) = \max \{ f(x_i = 0, x_j = 0, x_{-i-j}), f(x_i = 1, x_j = 0, x_{-i-j}) \}, \\
\nabla_i(f)(x_{i,+} = 1, x_{i,-} = 0, x_{j} = 1, x_{-i-j}) = \max \{ f(x_i = 0, x_j = 1, x_{-i-j}), f(x_i = 1, x_j = 1, x_{-i-j}) \},
\]

and using monotonicity of \( f \) in coordinate \( j \), the first value is less or equal the second (as written, there appears to be a \( 2 \times 2 \) array of values such that each column is increasing downwards; this implies that the maximum of the two values in the upper row is less or equal the maximum of the two values in the lower row). Lastly, the case when \( x_{i,+} = 0 \) and \( x_{i,-} = 1 \) is handled similarly, confirming the monotonicity of \( \nabla_i(f) \) in coordinate \( j \).

Finally, a similar argument shows that if \( f \) is anti-monotone in coordinate \( j \), then so is \( \nabla_i(f) \), completing the proof of the lemma.

Though we do not need the following fact, we do note that applying the split operator (just like the switch operator, see Lemma 2.3.4) can only decrease the distance to monotonicity.

Lemma 2.3.9.

\[ \epsilon(\nabla_i(f)) \leq \epsilon(f). \]
Proof. Let \( h \) be a monotone function nearest to \( f \), i.e. \( \varepsilon(f) = \Delta(f, h) \). Since \( h \) is monotone, so is \( \nabla_i(h) \) and it is enough to show that
\[
\Delta(\nabla_i(f), \nabla_i(h)) \leq \Delta(f, h).
\]
We observe that the inequality above holds for every fixed setting of \( x_{-i} \in \{ -1, 1 \}^n \). Indeed, letting
\[
a_0 = f(x_i = 0, x_{-i}), \quad a_1 = f(x_i = 1, x_{-i}),
\]
\[
b_0 = h(x_i = 0, x_{-i}), \quad b_1 = h(x_i = 1, x_{-i}),
\]
the contribution to \( \Delta(f, h) \) is \( \Delta((a_0, a_1), (b_0, b_1)) \), whereas the contribution to \( \Delta(\nabla_i(f), \nabla_i(h)) \) is
\[
\Delta((a_0, \min\{a_0, a_1\}, \max\{a_0, a_1\}, a_1), (b_0, \min\{b_0, b_1\}, \max\{b_0, b_1\}, b_1)).
\]
It can be checked by easy case analysis that the former always dominates the latter. \( \square \)

### 2.3.3 Pure Functions

We will need to consider functions \( g_{f,S,\rho} \) that are obtained from function \( f : \{ -1, 1 \}^n \mapsto \{ -1, 1 \} \) by applying the split operator on coordinates in some index set \( S \subseteq \{ 1, \ldots, n \} \) in an order specified by a permutation \( \rho \) on the set \( S \). In particular if \( S = \{ 1, \ldots, n \} \) and \( \rho \) is arbitrary, then applying the split operator \( n \) times successively in the order \( \rho \) yields a function
\[
g(x_{1,+}, x_{1,-}, x_{2,+}, x_{2,-}, \ldots, x_{n,+}, x_{n,-}),
\]
which, by re-arranging the coordinates can be written as
\[
g(x_{1,+}, x_{2,+}, \ldots, x_{n,+}, x_{1,-}, x_{2,-}, \ldots, x_{n,-}).
\]
We may write the function as \( g(x, y) \) where \( x, y \in \{ 0, 1 \}^n \), \( x \) denotes the ‘+’ coordinates and \( y \) denotes the ‘−’ coordinates. By Lemma 2.3.8, \( g(x, y) \) is monotone in the \( x \)-coordinates and anti-monotone in the \( y \)-coordinates. Functions with this property will be important for us and we call them “pure”.

**Definition 2.3.10.** A function \( g : \{ 0, 1 \}^n \times \{ 0, 1 \}^n \mapsto \{ -1, 1 \} \), written as \( g(x, y) \), is called pure if it is monotone in every \( x \)-coordinate and anti-monotone in every \( y \)-coordinate.

As we said, pure functions will arise as intermediate functions during our proof and the main point is that pure functions are also “simple” in the sense that several of the parameters and theorems are easily characterized and proved respectively for pure functions. We start with the observation that \( \varepsilon(g) \) has an easy characterization for a pure function \( g(x, y) \) (although this is not the case for a general function). For a fixed \( x \in \{ -1, 1 \}^n \), let \( g(x, \cdot) \) denote the restriction of \( g \) when the first argument is fixed to \( x \).

**Lemma 2.3.11.** For a pure function \( g(x, y) \),
\[
\varepsilon(g) = \Theta \left( \mathbb{E}_x [\text{var}(g(x, \cdot))] \right).
\]

*Proof.* Define \( p_x = \mathbb{E}_y [g(x, y)] \), and define a function \( h(x, y) \) as
\[
h(x, y) = \begin{cases} 1 & \text{if } p_x > \frac{1}{2}, \\ 0 & \text{if } p_x \leq \frac{1}{2}. \end{cases}
\]
We note that \( h(x, y) \) depends only on \( x \). Since \( g \) is monotone in \( x \), whenever \( x \leq x' \), it holds that \( g(x, y) \leq g(x', y) \) and then taking expectation over \( y \), one gets \( p_x \leq p_{x'} \). Thus \( h \) is monotone and

\[
\varepsilon(g) \leq \Delta(h, g) = \mathbb{E}_x [\Delta(h(x, \cdot), g(x, \cdot))] = \mathbb{E}_x [\min(p_x, 1 - p_x)] \\
\leq \mathbb{E}_x [2p_x(1 - p_x)] = 2 \cdot \mathbb{E}_x [\text{var}(g(x, \cdot))].
\]

On the other hand if \( h' \) is any monotone function, then since \( g(x, \cdot) \) is anti-monotone, by Lemma 2.3.12,

\[
\Delta(h', g) = \mathbb{E}_x [\Delta(h'(x, \cdot), g(x, \cdot))] \geq \mathbb{E}_x [\text{var}(g(x, \cdot))].
\]

\[\square\]

**Lemma 2.3.12.** If \( \psi, \psi' : \{-1, 1\}^n \to \{-1, 1\} \) are such that \( \psi \) is monotone and \( \psi' \) is anti-monotone, then

\[
\Delta(\psi, \psi') \geq \max \{\text{var}(\psi), \text{var}(\psi')\}.
\]

**Proof.** Let \( p = \mathbb{E}[\psi] \) and \( p' = \mathbb{E}[\psi'] \). We note that \( (\psi, 1 - \psi') \) is a pair of monotone functions. The lemma follows using FKG inequality [FKG71],

\[
\Delta(\psi, \psi') = \mathbb{P}[\psi = 1, \psi' = 0] + \mathbb{P}[\psi = 0, \psi' = 1] \\
= \mathbb{P}[\psi = 1, (1 - \psi') = 1] + \mathbb{P}[\psi = 0, (1 - \psi') = 0] \\
\geq p(1 - p') + (1 - p)p' \\
\geq p(1 - p) + p'(1 - p') = \text{var}(\psi) + \text{var}(\psi').
\]

\[\square\]

We now show that Theorem 2.1.6 holds for a pure function \( g(x, y) \), which has a simple enough structure that it follows immediately from the undirected version of the theorem, i.e. Talagrand’s Theorem 2.1.3. We note that for a pure function, we do not lose any poly-log factor.

**Lemma 2.3.13.** For a pure function \( g(x, y) \),

\[
\mathbb{E}_{x,y} \left[ \sqrt{I_g^{-1}(x, y)} \right] \geq \Omega(\varepsilon(g)).
\]

**Proof.** Since \( g(x, y) \) is monotone in \( x \), all its negatively sensitive edges are incident on the \( y \)-coordinates. Moreover, since \( g(x, y) \) is anti-monotone in \( y \), all sensitive edges incident on the \( y \)-coordinates are actually negatively sensitive. Thus,

\[
\mathbb{E}_{x,y} \left[ \sqrt{I_g^{-1}(x, y)} \right] = \mathbb{E}_x \left[ \mathbb{E}_y \left[ \sqrt{I_g(x, \cdot)(y)} \right] \right].
\]

Using Theorem 2.1.3 on function \( g(x, \cdot) \) along with Lemma 2.3.11, we have

\[
\mathbb{E}_x \left[ \mathbb{E}_y \left[ \sqrt{I_g(x, \cdot)(y)} \right] \right] \geq \mathbb{E}_x \left[ \Omega(\text{var}(g(x, \cdot))) \right] \geq \Omega(\varepsilon(g)).
\]

\[\square\]

Similarly, Theorem 2.1.9 holds for a pure function \( g(x, y) \), which is noted as the lemma below. This follows in a similar manner from the undirected version of the theorem i.e. Theorem 2.1.8. We skip the proof. We note again that for a pure function, we do not lose any poly-log factor.
Lemma 2.3.14. For a pure function \( g(x, y) \) and for an arbitrary coloring \( \text{col} : \mathcal{S}_g^- \mapsto \{ \text{red}, \text{blue} \} \),

\[
\mathbb{E}_{x,y} \left[ \sqrt{I^-_{g, \text{red}}(x, y)} \right] + \mathbb{E}_{x,y} \left[ \sqrt{I^-_{g, \text{blue}}(x, y)} \right] \geq \Omega(\varepsilon(g)).
\]

Finally, we will need the notion of a strongly monotone function, which is defined only for functions of the type

\[
h(x_{1,+}, x_{1,-}, x_{2,+}, x_{2,-}, \ldots, x_{n,+}, x_{n,-}),
\]

i.e. for functions for which the coordinates are paired and in each pair, one coordinate is designated as ‘+’ and the other as ‘−’.

Definition 2.3.15. A function \( h : \{0, 1\}^{2n} \mapsto \{0, 1\} \), written as,

\[
h(x_{1,+}, x_{1,-}, x_{2,+}, x_{2,-}, \ldots, x_{n,+}, x_{n,-}),
\]

is strongly monotone if it is monotone and moreover, for any \( i \in \{1, \ldots, n\} \), changing the pair of coordinates \( ((i,+), (i,-)) \) from \( 01 \) to \( 10 \) cannot change the function from \( 1 \) to \( 0 \).

For a function \( g(x_{1,+}, x_{1,-}, x_{2,+}, x_{2,-}, \ldots, x_{n,+}, x_{n,-}) \), let \( \delta(g) \) denote the minimum distance of \( g \) from any strongly monotone function. Clearly \( \delta(g) \geq \varepsilon(g) \), but it turns out that if \( g = g(x, y) \) is a pure function, then \( \delta(g) \) is same as \( \varepsilon(g) \) up to a constant factor. Here, it is understood that we re-arrange the coordinates of \( g \) as

\[
g(x_{1,+}, x_{2,+}, \ldots, x_{n,+}, x_{1,-}, x_{2,-}, \ldots, x_{n,-}),
\]

and write \( g = g(x, y) \) with \( x \) and \( y \) denoting the ‘+’ inputs and ‘−’ inputs respectively (and vice versa; from the representation \( g(x, y) \) one can go back to the representation in terms of pairs of coordinates).

Lemma 2.3.16. For a pure function \( g(x, y) \),

\[
\varepsilon(g) \geq \Omega(\delta(g)).
\]

Proof. The lemma can be equivalently stated as \( \delta(g) \leq O(\varepsilon(g)) \). By Lemma 2.3.11, and also examining its proof, we see that for a pure function \( g(x, y) \),

\[
\varepsilon(g) = \Theta\left( \mathbb{E}_x[\text{var}(g(x, \cdot))] \right),
\]

i.e. the most efficient way to turn \( g \) into a monotone function (up to a constant factor) is to change \( g(x, \cdot) \) to identically 0 or identically 1 depending on whether \( g(x, \cdot) \) is more likely to be 0 or 1. Note that, once this is done, the new function \( h(x, y) \) does not depend on the \( y \) coordinates at all and hence is a strongly monotone function.

2.3.4 Splitting only Decreases Talagrand Objective

In this section, we show that the “Talagrand objective”, i.e. the LHS. of the inequality in Theorem 2.1.6 (and later the robust Theorem 2.1.9) can only decrease when a split operator is applied to \( f \).

Lemma 2.3.17. For a coordinate \( i \in \{1, \ldots, n\} \),

\[
\mathbb{E}_x \left[ \sqrt{I^-_f(x)} \right] \geq \mathbb{E}_{x'} \left[ \sqrt{I^-_{f_i}(x')} \right].
\]
Let \( j \) values of \( f \) nates except \( i \) is sensitive edges on the input otherwise. The values of \( I^1 \) coordinate to 

**Proof.** First we examine how a typical contribution to the LHS. looks like. Fix some setting \( x_{-i} \) of coordinates except \( i \) and we look at the value of \( I_f^-(\cdot) \) at the two inputs \( (x_i = 0, x_{-i}) \) and \( (x_i = 1, x_{-i}) \). Consider values of \( f \) at these two inputs:

\[
    a = f(x_i = 0, x_{-i}), \quad b = f(x_i = 1, x_{-i}).
\]

Let \( j \in \{1, \ldots, n\}, j \neq i \) be any coordinate such that the \( j^{th} \) coordinate of \( x_{-i} \) equals 0. Changing this coordinate to 1, let the values of \( f \) at the two new inputs be:

\[
    a' = f(x_i = 0, x_j = 1, x_{-i-j}), \quad b' = f(x_i = 1, x_j = 1, x_{-i-j}).
\]

Consider the notation:

\[
    (a, b) \rightarrow [(0, 0), \#\alpha], \quad [(1, 0), \#\beta], \quad [(0, 1), \#\gamma], \quad [(1, 1), \#\delta],
\]

meaning that the number of coordinates \( j \) for which \( (a', b') = (0, 0) \) is exactly \( \alpha \), the number of coordinates \( j \) for which \( (a', b') = (1, 0) \) is exactly \( \beta \), and so on (see Figure 2).

Recall that the value of \( I_f^-(\cdot) \) at an input equals zero if \( f = 0 \) at the input and the number of negatively sensitive edges on the input otherwise. The values of \( I_f^-(\cdot) \) at the two inputs \( (x_i = 0, x_{-i}) \) and \( (x_i = 1, x_{-i}) \) depend on the values of \( f \) at these inputs, i.e. \( a \) and \( b \), and the numbers \( \alpha, \beta, \gamma, \delta \) in the following manner (the extra +1 at one place takes into account a negatively sensitive edge contributed by coordinate \( i \) itself):

\[
    I_f^-(x_i = 0, x_{-i}), I_f^-(x_i = 1, x_{-i}) = \begin{cases} 
        0, & \text{if } a = 0, b = 0 \\
        \alpha + \gamma + 1, & \text{if } a = 1, b = 0 \\
        0, & \text{if } a = 0, b = 1 \\
        \alpha + \beta, & \text{if } a = 1, b = 1 \\
        \alpha + \gamma, & \text{if } a = 1, b = 1 \\
    \end{cases}
\]

Now we consider the effect of applying the split operator on the coordinate \( i \). Under the action of the operator, the two inputs \( (x_i = 0, x_{-i}) \) and \( (x_i = 1, x_{-i}) \) give rise to four inputs, namely

\[
    (x_{i,+} = 0, x_{i,-} = 0, x_{-i}), (x_{i,+} = 0, x_{i,-} = 1, x_{-i}), (x_{i,+} = 1, x_{i,-} = 0, x_{-i}), (x_{i,+} = 1, x_{i,-} = 1, x_{-i}),
\]

with the values of \( \nabla_i(f) \) at these inputs respectively as

\[
    a, \min\{a, b\}, \max\{a, b\}, b.
\]

As before, we investigate the scenario when, for some \( j \in \{1, \ldots, n\}, j \neq i \), the \( j^{th} \) coordinate of these four inputs is changed from 0 to 1. It is not difficult to see that the values of \( \nabla_i(f) \) at these four new inputs are:

\[
    a', \min\{a', b'\}, \max\{a', b'\}, b',
\]
if the values of $f$ at the two new inputs are $(a', b')$ as discussed above. Using a similar notation as before:

$$(a, \min\{a, b\}, \max\{a, b\}, b) \mapsto \left[(0, 0, 0, 0), \#\alpha\right], \left[(1, 0, 1, 0), \#\beta\right], \left[(0, 0, 1, 1), \#\gamma\right], \left[(1, 1, 1, 1), \#\delta\right],$$

meaning that the number of coordinates $j$ for which $(a', \min\{a', b'\}, \max\{a', b'\}, b') = (0, 0, 0, 0)$ is exactly $\alpha$ and so on (where $\alpha, \beta, \gamma, \delta$ are precisely the numbers defined before). Now we are ready to complete the proof considering the four cases depending on values $a$ and $b$. In each case, we compare the contribution of $I^+_{f^*}(\cdot)$ at the two relevant inputs against the contribution of $I^+_{\nabla^i(f)}(\cdot)$ at the four relevant inputs to the inequality in the statement of the lemma.

When $a = 0, b = 0$, values of $I^+_{\nabla^i(f)}(\cdot)$ are:

$$0, 0, 0, 0,$$

and the inequality in the statement of the lemma amounts to saying that:

$$\frac{\sqrt{0} + \sqrt{0}}{2} \geq \frac{\sqrt{0} + \sqrt{0} + \sqrt{0} + \sqrt{0}}{4}.$$

When $a = 1, b = 0$, values of $I^+_{\nabla^i(f)}(\cdot)$ are:

$$\alpha + \gamma + 1, 0, \alpha + 1, 0,$$

and the inequality in the statement of the lemma amounts to saying that:

$$\frac{\sqrt{\alpha + \gamma + 1} + \sqrt{0}}{2} \geq \frac{\sqrt{\alpha + \gamma + 1} + \sqrt{0} + \sqrt{\alpha + 1} + \sqrt{0}}{4}.$$

When $a = 0, b = 1$, values of $I^+_{\nabla^i(f)}(\cdot)$ are:

$$0, 0, \alpha, \alpha + \beta,$$

and the inequality in the statement of the lemma amounts to saying that:

$$\frac{\sqrt{0} + \sqrt{\alpha + \beta}}{2} \geq \frac{\sqrt{0} + \sqrt{\alpha + \beta} + \sqrt{\alpha} + \sqrt{\alpha + \beta}}{4}.$$

When $a = 1, b = 1$, values of $I^+_{\nabla^i(f)}(\cdot)$ are:

$$\alpha + \gamma, \alpha + \beta + \gamma, \alpha, \alpha + \beta,$$

and the inequality in the statement of the lemma amounts to saying that:

$$\frac{\sqrt{\alpha + \gamma} + \sqrt{\alpha + \beta}}{2} \geq \frac{\sqrt{\alpha + \gamma} + \sqrt{\alpha + \beta} + \gamma + \sqrt{\alpha} + \sqrt{\alpha + \beta}}{4}.$$

All of the assertions above hold by the convexity of $z \mapsto \sqrt{z}$. □

The lemma below shows that the LHS. of inequality in the robust Theorem 2.1.9 can only decrease when a split operator is applied to $f$. This is a bit subtle as one shows that for any coloring of the negatively sensitive edges of $f$, there is a coloring of the negatively sensitive edges of $\nabla^i(f)$ such that the objective can only decrease.
Lemma 2.3.18. For any coloring \( \text{col} : \mathcal{S}^- \rightarrow \{\text{red, blue}\} \) and coordinate \( i \in \{1, \ldots, n\} \), there exists a coloring \( \text{col} : \mathcal{S}^-_{\nabla_i(f)} \rightarrow \{\text{red, blue}\} \) such that
\[
\mathbb{E}_x \left[ \sqrt{I^-_{\nabla_i(f), \text{red}}(x)} \right] + \mathbb{E}_x \left[ \sqrt{I^-_{\nabla_i(f), \text{blue}}(x)} \right] \geq \mathbb{E}_{x'} \left[ \sqrt{I^-_{\nabla_i(f), \text{red}}(x')} \right] + \mathbb{E}_{x'} \left[ \sqrt{I^-_{\nabla_i(f), \text{blue}}(x')} \right].
\]

Proof. After defining the appropriate coloring \( \text{col} : \mathcal{S}^-_{\nabla_i(f)} \rightarrow \{\text{red, blue}\} \), we will in fact show two separate inequalities:
\[
\mathbb{E}_x \left[ \sqrt{I^-_{\nabla_i(f), \text{red}}(x)} \right] \geq \mathbb{E}_{x'} \left[ \sqrt{I^-_{\nabla_i(f), \text{red}}(x')} \right],
\]
\[
\mathbb{E}_x \left[ \sqrt{I^-_{\nabla_i(f), \text{blue}}(x)} \right] \geq \mathbb{E}_{x'} \left[ \sqrt{I^-_{\nabla_i(f), \text{blue}}(x')} \right].
\]
The coloring \( \text{col} : \mathcal{S}^-_{\nabla_i(f)} \rightarrow \{\text{red, blue}\} \) is derived naturally from the coloring \( \text{col} : \mathcal{S}^- \rightarrow \{\text{red, blue}\} \) in the following manner. Fix a setting of \( x_{-i} \in \{-1, 1\}^{n-1} \) for the rest of the proof. We recall that there are two relevant inputs for \( f \) with values
\[
a = f(x_i = 0, x_{-i}), \quad b = f(x_i = 1, x_{-i}),
\]
and four relevant inputs for \( \nabla_i(f) \) with values
\[
a = (\nabla_i(f))(x_{i,+} = 0, x_{i,-} = 0, x_{-i}), \quad \min\{a, b\} = (\nabla_i(f))(x_{i,+} = 0, x_{i,-} = 1, x_{-i}),
\]
\[
\max\{a, b\} = (\nabla_i(f))(x_{i,+} = 1, x_{i,-} = 0, x_{-i}), \quad b = (\nabla_i(f))(x_{i,+} = 1, x_{i,-} = 1, x_{-i}).
\]

Towards the desired coloring, we first consider the potential, negatively sensitive edge between inputs \((x_i = 0, x_{-i})\) and \((x_i = 1, x_{-i})\). This edge, denoted \( e \), is negatively sensitive if and only if \( a = 1, b = 0 \) and in that case, it induces two negatively sensitive edges of \( \nabla_i(f) \), between two inputs in the same row, for each of the two rows above, and these two edges are colored with the same color as \( e \).

Next, we consider potential, negatively sensitive edges along dimension \( j \in \{1, \ldots, n\}, j \neq i \). We assume, w.l.o.g. that the \( j^{\text{th}} \) coordinate of \( x_{-i} \) equals 0. When this coordinate is changed to 1, let us denote the values of \( f \) at the two new inputs as \((a', b')\) and the values of \( \nabla_i(f) \) at the four new inputs as
\[
a', \quad \min\{a', b'\}, \quad \max\{a', b'\}, \quad b'.
\]

Let \( e_0 \) denote the edge whose endpoints have \( f \)-values \((a, a')\) and \( e_1 \) denote the edge whose endpoints have \( f \)-values \((b, b')\). Similarly, Let \( e_{00}, e_{\min}, e_{\max}, e_{11} \) denote the edges whose endpoints have \( (\nabla_i(f)) \)-values \((a, a')\), \((\min\{a, b\}, \min\{a', b'\})\), \((\max\{a, b\}, \max\{a', b'\})\), \((b, b')\) respectively. We now assign colors to \( e_{00}, e_{\min}, e_{\max}, e_{11} \) depending on the colors of \( e_0 \) and \( e_1 \). We note that colors are assigned to only negatively sensitive edges.

- If \( e_0 \) and \( e_1 \) are uncolored, so are \( e_{00}, e_{\min}, e_{\max}, e_{11} \).

- If exactly one of the two edges \( e_0 \) and \( e_1 \) is colored, say by color \( c \in \{\text{red, blue}\} \), then all negatively sensitive edges among \( e_{00}, e_{\min}, e_{\max}, e_{11} \) are colored with the same color \( c \) (depending on the case, there are one or two such negatively sensitive edges).

- If both the edges \( e_0 \) and \( e_1 \) are colored (which happens if and only if \( a = b = 1, a' = b' = 0 \)), then:
  - If \( e_0 \) and \( e_1 \) are colored with the same color, say \( c \in \{\text{red, blue}\} \), then \( e_{00}, e_{\min}, e_{\max}, e_{11} \) all get the same color \( c \).
If \(e_0\) and \(e_1\) are colored with different colors, then \(e_{00}\) gets the color of \(e_0\), \(e_{11}\) gets the color of \(e_1\), \(e_{\text{min}}\) is colored red and \(e_{\text{max}}\) is colored blue.

We now prove that (the other “blue inequality” being symmetric):

\[
\mathbb{E}_x \left[ \sqrt{I_{f,\text{red}}(x)} \right] \geq \mathbb{E}_{x'} \left[ \sqrt{I_{\nabla_i(f),\text{red}}(x')} \right].
\]  \hspace{1cm} (2.2)

We note that the above inequality only concerns the number of red, negatively sensitive edges that are incident on an input with \(f\) or \(\nabla_i(f)\)-value equal to 1. For the fixed setting of \(x_{-i}\) we examine the contribution to LHS. of the inputs \((x_i \in \{0, 1\}, x_{-i})\) and to RHS. of the inputs \((x_{i,+} \in \{0, 1\}, x_{i,-} \in \{0, 1\}, x_{-i})\), depending on four cases according to values of \(a\) and \(b\). Also, in the following, coordinate \(j\) always refers to a coordinate \(j \in \{1, \ldots, n\}, j \neq i\) such that the \(j^{th}\) coordinate of \(x_{-i}\) equals 0.

**Case** \(a = b = 0\):

In this case, since the relevant inputs have \(f\) and \(\nabla_i(f)\) values equal to 0, there is no contribution to either side of inequality (2.2).

**Case** \(a = b = 1\):

We partition the set of coordinates \(j\) as in proof of Lemma 2.3.17, however we need a finer partition. Let \(a_1, a_2, a_3\) be the number of coordinates \(j\) such that \((a, b) \mapsto (0, 0)\), and the coloring of \((e_0, e_1)\) is (red,red), (blue,red), or (red,blue) respectively. Let \(\beta\) be the number of coordinates \(j\) such that \((a, b) \mapsto (1, 0)\) and the edge \(e_1\) is red. Let \(\gamma\) be the number of coordinates \(j\) such that \((a, b) \mapsto (0, 1)\) and the edge \(e_0\) is red. Then,

\[
I_{f,\text{red}}(x_i = 0, x_{-i}) = a_1 + a_3 + \gamma, \quad I_{f,\text{red}}(x_i = 1, x_{-i}) = a_1 + a_2 + \beta
\]

\[
I_{\nabla_i(f),\text{red}}(x_{i,+}, x_{i,-}, x) = \begin{cases} 
\alpha + a_3 + \gamma & \text{if } x_{i,+} = 0, x_{i,-} = 0 \\
\alpha + a_2 + a_3 + \beta + \gamma & \text{if } x_{i,+} = 0, x_{i,-} = 1 \\
\alpha & \text{if } x_{i,+} = 1, x_{i,-} = 0 \\
\alpha + a_2 + \beta & \text{if } x_{i,+} = 1, x_{i,-} = 1 
\end{cases}
\]

So the inequality (2.2) amounts to saying

\[
\frac{\sqrt{\alpha_1 + a_3 + \gamma} + \sqrt{\alpha_1 + a_2 + \beta}}{2} \geq \frac{\sqrt{\alpha_1 + a_3 + \gamma} + \sqrt{\alpha_1 + a_2 + \beta} + \gamma + \sqrt{\alpha_1 + \sqrt{\alpha_1 + \alpha_2 + \beta}}}{4}.
\]

**Case** \(a = 0, b = 1\):

In this case, only \((x_i = 1, x_{-i})\) contributes to the LHS. of inequality (2.2) and only \((x_{i,+} = 1, x_{i,-} \in \{0, 1\}, x_{-i})\) contributes to the RHS. Again, we look at a partition of the coordinates \(j\). Let \(\alpha\) be the number of coordinates \(j\) such that \((a, b) \mapsto (0, 0)\) and the edge \(e_1\) is red. Let \(\beta\) be the number of coordinates \(j\) such that \((a, b) \mapsto (1, 0)\) and the edge \(e_0\) is red. Then,

\[
I_{f,\text{red}}(x_i = 1, x_{-i}) = \alpha + \beta.
\]

\[
I_{\nabla_i(f),\text{red}}(x_{i,+} = 1, x_{i,-}, x_{-i}) = \begin{cases} 
\alpha & \text{if } x_{i,-} = 0 \\
\alpha + \beta & \text{if } x_{i,-} = 1 
\end{cases}
\]

So the inequality (2.2) amounts to saying

\[
\frac{\sqrt{0 + \alpha + \beta}}{2} \geq \frac{\sqrt{0 + \alpha + \beta}}{4}.
\]
Case $a = 1, b = 0$:

In this case, only $(x_i = 0, x_{-i})$ contributes to the LHS of inequality (2.2) and only $(x_{i,+} \in \{0, 1\}, x_{i,-} = 0, x_{-i})$ contributes to the RHS. Again, we look at a partition of the coordinates $j$. Let $\alpha$ be the number of coordinates $j$ such that $(a, b) \mapsto (0, 0)$ and the edge $e_0$ is red. Let $\beta$ be the number of coordinates $j$ such that $(a, b) \mapsto (0, 1)$ and the edge $e_0$ is red. We note that in this case, we also have negatively sensitive edges along dimension $i$ itself. Let $\chi = 1$ if the negatively sensitive edge with endpoints $(x_i = 0, x_{-i})$ and $(x_i = 1, x_{-i})$ is red and $\chi = 0$ otherwise. Then,

\begin{align*}
I_{f, \text{red}}(x_i = 0, x_{-i}) &= \alpha + \beta + \chi. \\
I_{\nabla_i(f), \text{red}}(x_{i,+}, x_{i,-} = 0, x_{-i}) &= \begin{cases} 
\alpha + \beta + \chi & \text{if } x_{i,+} = 0 \\
\alpha + \chi & \text{if } x_{i,+} = 1 
\end{cases}
\end{align*}

So the inequality (2.2) amounts to saying

\[
\frac{\sqrt{\alpha + \beta + \chi + \sqrt{0}}}{2} \geq \frac{\sqrt{\alpha + \beta + \chi + \sqrt{0}} + \sqrt{\alpha + \chi + \sqrt{0}}}{4}.
\]

\[
\Box
\]

### 2.4 Proof of Theorem 2.1.6

In this section, we present a proof of Theorem 2.1.6, which we consider the most interesting part of this chapter. The first subsection describes an informal overview of the proof and the formal proof appears in the next subsection. The overview omits several of the key ingredients and is meant for intuition only, and the reader should expect the formal proof to be quite different in terms of notation, additional ideas etc.

#### 2.4.1 An Overview

We recall that for a function $f : \{0, 1\}^n \mapsto \{0, 1\}$ and input $x$, $I_f(x)$ is equal to 0 if $f(x) = 0$ and equal to the number of negatively sensitive edges incident on $x$ if $f(x) = 1$. The minimum distance of $f$ from any monotone function is denoted as $\varepsilon(f)$. We intend to show that (ignoring the difference between $\Omega$ and $\tilde{\Omega}$ notations):

\[
\mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \Omega(\varepsilon(f)). \tag{2.3}
\]

We start by describing a preliminary attempt towards a proof, then point out why it doesn’t quite work, and then describe how to extend this preliminary attempt to a correct proof. We attempt to “reduce” the inequality (2.3) concerning function $f$ to the same inequality concerning the function $g : \{0, 1\}^{2n} \mapsto \{0, 1\}$ that is obtained by applying the split operator to $f$, successively on coordinates 1 through $n$, say in that order. Partitioning the set of coordinates of $g$ into two blocks of size $n$ each, we use the notation $g = g(x, y)$. As in Section 2.3.3, $g(x, y)$ is a pure function in the sense that $g$ is monotone in $x$-coordinates and anti-monotone in $y$-coordinates. Let $\varepsilon(g)$ be the minimum distance of $g$ from any monotone function. Towards proving inequality (2.3), we observe that:

- when one replaces the function $f$ by the pure function $g$, the LHS of the inequality (2.3) can only decrease (this is by Lemma 2.3.17; splits can only decrease the Talagrand objective).
- the inequality (2.3) holds for the pure function $g(x, y)$ (this is by Lemma 2.3.13).
Thus, we can conclude
\[ \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \mathbb{E}_{x,y} \left[ \sqrt{I_g(x,y)} \right] \geq \Omega(\varepsilon(g)). \] (2.4)

Now, the inequality (2.3) would be proved if it were (always) the case that \( \varepsilon(g) \geq \Omega(\varepsilon(f)) \) - this turns out to be incorrect\(^9\). Still, by a careful examination of the split operator and relating it to the switch operator, we are able to show a lower bound (up to a constant factor)
\[ \varepsilon(g) \geq \Delta(f, \pi_n \circ f) - \mathbb{E}_{\pi_{n/2}}[\Delta(f, \pi_{n/2} \circ f)]. \] (2.5)

We elaborate more on this lower bound. Here \( \pi_n \circ f \) denotes the function obtained from \( f \) by applying the switch operator on all \( n \) coordinates (and \( \pi_n \) denotes the full set \( \{1, \ldots, n\} \)). Also, \( \pi_{n/2} \) denotes a (random) subset of \( \frac{n}{2} \) coordinates and \( \pi_{n/2} \circ f \) denotes the function obtained from \( f \) by applying the switch operator on precisely the coordinates in \( \pi_{n/2} \). We certainly know that \( \pi_n \circ f \) is a monotone function and we think of applying the switch operator on coordinates, one by one, as “progressing” towards the “target function” \( \pi_n \circ f \). By definition, applying the switch operator on all \( n \) coordinates, attains the target. However, it is possible that applying the switch operator on only (random) \( \frac{n}{2} \) coordinates gets us very close to the target. If so, \( \pi_{n/2} \circ f \approx \pi_n \circ f \) for almost every choice of \( \pi_{n/2} \) and one does not get a good lower bound in inequality (2.5). Nevertheless, combining inequalities (2.4), (2.5) and thinking of \( \pi_n \) itself as a random set of size \( n \) (though there is only one set of size \( n \) and no randomness is involved), we get
\[ \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \mathbb{E}_{\pi_n}[\Delta(f, \pi_n \circ f)] - \mathbb{E}_{\pi_{n/2}}[\Delta(f, \pi_{n/2} \circ f)]. \]

This inequality now suggests that we ought to get another inequality,
\[ \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \mathbb{E}_{\pi_{n/2}}[\Delta(f, \pi_{n/2} \circ f)] - \mathbb{E}_{\pi_{n/4}}[\Delta(f, \pi_{n/4} \circ f)], \]

that reflects the difference between applying the switch operator on random \( \frac{n}{2} \) coordinates versus applying it on random \( \frac{n}{4} \) coordinates. Indeed, we do obtain such an inequality. Instead of working with the function \( g = g_n \) that is obtained by applying the split operator to \( f \) on all \( n \) coordinates, we work with the function \( g_{n/2} \) that is obtained by applying the split operator to \( f \) on (random) \( n/2 \) coordinates. More generally, we are able to obtain similar inequalities for \( i = 0, 1, 2, \ldots, \lfloor \log n \rfloor \) as
\[ \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \mathbb{E}_{\pi_{2^i}}[\Delta(f, \pi_{2^i} \circ f)] - \mathbb{E}_{\pi_{2^{i+1}}}[\Delta(f, \pi_{2^{i+1}} \circ f)]. \]

Summing up these inequalities in a telescoping manner and ignoring the negative term for the last inequality numbered \( i = \lfloor \log n \rfloor \) (corresponding to not applying switch operator at all and the term amounts to \( \Delta(f, f) = 0 \)), we get
\[ \lfloor \log n \rfloor \cdot \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \Delta(f, \pi_n \circ f) \geq \varepsilon(f), \]

since \( \pi_n \circ f \) is a monotone function. This proves inequality (2.3), but with a loss of log-factor.

### 2.4.2 The Formal Proof

We now prove Theorem 2.1.6 formally. Let \( f : \{-1, 1\}^n \mapsto \{ -1, 1 \} \) be the given function. Fix a non-empty set \( S \subseteq \{1, \ldots, n\} \) and partition the set of coordinates \( \{1, \ldots, n\} \) as \( S \cup \overline{S} \). We write \( f(x) = f(w, z) \) where

---

\(^9\)A counter example is given by \( f(x) = 1 - \text{majority}(x) \). It has distance 1/2 from monotone, yet \( g \) is close to monotone. Roughly speaking, the reason is that on most sub-cubes \( g \) looks like \( f \) with half of the switches applied.
the input \( x \) is partitioned into \( w, z \) denoting the inputs on coordinates in sets \( S \) and \( \overline{S} \) respectively. For every fixed \( z \), we will consider the function \( f(\cdot, z) \), apply switch and split operators on it, obtain certain inequalities and finally take expectation over \( z \). We start by observing that

\[
\mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \mathbb{E}_z \left[ \mathbb{E}_w \left[ \sqrt{I_{f(\cdot, z)}(w)} \right] \right].
\]

(2.6)

We elaborate more on this inequality. The function \( f(\cdot, z) \) is considered to be a function of only the coordinates in \( S \). The inequality follows by observing that for any input \( x = (w, z) \),

\[
I_f(x) \geq I_{f(\cdot, z)}(w).
\]

Indeed, if \( f(x) = f(w, z) = 0 \), both sides equal zero. Otherwise the LHS. equals the number of negatively sensitive edges of \( f \) incident on \( x = (w, z) \), whereas the RHS. equals the number of negatively sensitive edges of \( f(\cdot, z) \) incident on \( w \), which is same as the number of negatively sensitive edges of \( f \) incident on \( x = (w, z) \), but only considering the edges along coordinates in \( S \).

Now we consider the function \( g_{S, \rho} \) that is obtained from \( f \) by applying the split operator on coordinates in \( S \) in the order given by the permutation \( \rho \) of the set \( S \). Since \( f = f(w, z) \) and the split operator does not “touch” the coordinates in \( \overline{S} \), we may write \( g_{S, \rho} = g_{S, \rho}(u, v, z) \) where \( (u, v) \) denote the ‘+’ and ‘−’ coordinates obtained after splitting the coordinates of \( w \). For every fixed \( z \), the function \( g_{S, \rho}(\cdot, \cdot, z) \) is pure in the sense that, regarded only as function of \( u \) and \( v \), it is monotone in the \( u \)-coordinates and anti-monotone in the \( v \)-coordinates. Moreover, for every fixed \( z \), we can consider \( g_{S, \rho}(\cdot, \cdot, z) \) as the function obtained from \( f(\cdot, z) \) by applying the split operator on coordinates in \( S \) in the order \( \rho \). By Lemma 2.3.17, splitting can only decrease the Talagrand objective, and hence for every fixed \( z, S, \rho \),

\[
\mathbb{E}_w \left[ \sqrt{I^{-}_{g_{S, \rho}(\cdot, \cdot, z)}(w)} \right] \geq \mathbb{E}_{u,v} \left[ \sqrt{I^{-}_{g_{S, \rho}(\cdot, \cdot, z)}(u, v)} \right].
\]

Since \( g_{S, \rho}(\cdot, \cdot, z) \) is a pure function, by Lemma 2.3.13, Definition 2.3.15 of strong monotonicity, and Lemma 2.3.16 regarding distance to strong monotonicity, we have

\[
\mathbb{E}_{u,v} \left[ \sqrt{I^{-}_{g_{S, \rho}(\cdot, \cdot, z)}(u, v)} \right] \geq \Omega(\delta(g_{S, \rho}(\cdot, \cdot, z))).
\]

Combining the inequalities above, we have, for some absolute constant \( C \),

\[
C \cdot \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \mathbb{E}_z [\delta(g_{S, \rho}(\cdot, \cdot, z))].
\]

(2.7)

**Relating Splits to Switches**

Now we take a closer look at the function \( g_{S, \rho}(\cdot, \cdot, z) \). This function is defined on input \( (u, v) \) where \( |u| = |v| = |S| \). Re-arranging the input coordinates into \((+, -)\) pairs, and denoting by \( \Sigma = \{00, 01, 10, 11\} \) as the four possible values that a pair of coordinates may take, we view the function as:

\[
g_{S, \rho}(\cdot, \cdot, z) : \Sigma^{|S|} \mapsto \{0, 1\}.
\]

We also write \( g_{S, \rho}(\sigma, z) \) where \( \sigma \in \Sigma^{|S|} \) and denote coordinates of \( \sigma \) as \( \sigma_i \) for \( i \in S \). To avoid confusion, we emphasize that when we write \( g_{S, \rho}(\sigma, z) \), the coordinates of \( \sigma \) are understood to be re-ordered according to the permutation \( \rho \). For example, if \( S = \{1, 2, 3, 4, 5, 6, 7, 8\} \) and \( \rho = (5, 4, 8, 1, 7, 2, 3, 6) \), then \( g_{S, \rho}(\sigma, z) \) is interpreted as:

\[
g_{S, \rho}(\sigma_5, \sigma_4, \sigma_8, \sigma_1, \sigma_7, \sigma_2, \sigma_3, \sigma_6, z).
\]
Let \( \pi : \Sigma = \{00, 01, 10, 11\} \mapsto \{Y, N\} \) be defined as \( \pi(00) = \pi(11) = N \) and \( \pi(01) = \pi(10) = Y \). We extend \( \pi \) to \( \pi : \Sigma^i \mapsto \{Y, N\}^i \) for integer \( i \) by its application on each coordinate. Let \( \varphi : \Sigma = \{00, 01, 10, 11\} \mapsto \{0, 1\} \) be defined as \( \varphi(00) = \varphi(01) = 0 \) and \( \varphi(10) = \varphi(11) = 1 \). Thus \( \varphi \) simply selects the first coordinate of a pair of coordinates. We extend \( \varphi : \Sigma^i \mapsto \{0, 1\}^i \) for integer \( i \) by its application on each coordinate.

The lemma below shows that \( g_{S, \rho}(\cdot, z) \) is actually composed of copies of \( f(\cdot, z) \) with suitable switch operators applied. Before stating the lemma, we need to explain further notation. For a permutation \( \rho \) of set \( S \) and a vector \( \pi \in \{Y, N\}^{|S|} \), we denote by \( \rho \ast \pi \), the permutation \( \rho \) with the elements whose \( \pi \)-coordinate is ‘\( N \)’ “dropped”. More explicitly, we think of the permutation \( \rho \) as an ordered list \( (t_1, \ldots, t_{|S|}) \) of elements of \( S \subseteq \{1, \ldots, n\} \) and then \( \rho \ast \pi \) is this ordered list with element \( t_i \) dropped if \( \pi_i = N \). Thus \( \rho \ast \pi \) is also a ordered list and then \( (\rho \ast \pi) \circ f(\cdot, z) \) denotes the function obtained from \( f(\cdot, z) \) by applying the switch operator on coordinates in the order \( 5, 1, 2, 3 \). As an illustration, suppose

\[
S = \{1, 2, 3, 4, 5, 6, 7, 8\}, \\
\rho = (5, 4, 8, 1, 7, 2, 3, 6), \\
\pi = (Y, N, N, Y, N, Y, Y, N),
\]

so that \( \rho \ast \pi = (5, 1, 2, 3) \) and \( (\rho \ast \pi) \circ f(\cdot, z) \) is the function obtained from \( f(\cdot, z) \) by applying the switch operator on coordinates in the order \( 5, 1, 2, 3 \). Now we are ready to state the lemma.

**Lemma 2.4.1.** For \( \sigma \in \Sigma^{|S|} \),

\[
g_{S, \rho}(\sigma, z) = ((\rho \ast \pi(\sigma)) \circ f)(\varphi(\sigma), z).
\]

**Proof.** Since the input \( z \) is “auxiliary” and just “floats around”, we can drop it from the notation. Equivalently, we can assume that \( S = \{1, \ldots, n\} \) is the full set. Also, we can assume w.l.o.g. that the permutation \( \rho \) is the identity permutation, i.e. the ordered list \((1, 2, \ldots, n)\). Thus the function \( g_{S, \rho} \) is the function obtained from \( f \) by applying the split operator on coordinates 1 through \( n \), in that order. We write \( g = g_{S, \rho} \) and drop \( S \) and \( \rho \) from the notation. Further, for \( \pi \in \{Y, N\}^n \), by denoting \( \pi \circ f \) as the function obtained from \( f \) by considering the coordinates 1 through \( n \) in that order, and applying the switch operator on \( j \)-th coordinate if and only if \( \pi_j = Y \), the lemma amounts to saying

\[
\forall \sigma \in \Sigma^n, \quad g(\sigma) = (\pi(\sigma) \circ f)(\varphi(\sigma)). \tag{2.8}
\]

Using the short-form \( \nabla_{[1, \ldots, i]}(f) \) to denote the “prefix”

\[
\nabla_i(\nabla_{i-1}(\ldots \nabla_2(\nabla_1(f)) \ldots)),
\]

we note that \( g = \nabla_{[1, \ldots, n]}(f) \). Also, \( \nabla_{[1, \ldots, i]}(f) \) is a function

\[
\nabla_{[1, \ldots, i]}(f) : \Sigma^i \times \{0, 1\}^{n-i} \mapsto \{0, 1\}.
\]

We prove by induction on \( i \) that

\[
\forall \sigma \in \Sigma^i, \ x \in \{0, 1\}^{n-i}, \quad \nabla_{[1, \ldots, i]}(f)(\sigma, x) = (\pi(\sigma) \circ f)(\varphi(\sigma), x), \tag{2.9}
\]

and the lemma follows from the case \( i = n \). We note that in the inductive statement above, \( \sigma \in \Sigma^i, \pi(\sigma) \in \{Y, N\}^i \) and \( \pi(\sigma) \circ f \) denotes the function obtained from \( f \) by considering the coordinates 1 through \( i \) in that order, and applying the switch operator on \( j \)-th coordinate if and only if \( \pi_j = Y \).
For $i = 0$, there is nothing to prove as the statement is $f(x_1, \ldots, x_n) = f(x_1, \ldots, x_n)$. Assume the statement (2.9) for some $0 \leq i < n - 1$ and for convenience, write $x = (x_{i+1}, y)$ where $y = (y_{i+2}, \ldots, y_n)$ is a vector of formal boolean variables. Thus the inductive hypothesis is that:

$$\nabla_{[1, \ldots, i]}(f)(\sigma, x_{i+1}, y) = (\pi(\sigma) \circ f)(\varphi(\sigma), x_{i+1}, y).$$

By applying the split operator $\nabla_{i+1}$ to the LHS.

$$\nabla_{[1, \ldots, i+1]}(f)(\sigma, \sigma_{i+1}, y) = \begin{cases} \nabla_{[1, \ldots, i]}(f)(\sigma, 0, y) & \sigma_{i+1} = 00 \\ \min \{ \nabla_{[1, \ldots, i]}(f)(\sigma, 0, y), \nabla_{[1, \ldots, i]}(f)(\sigma, 1, y) \} & \sigma_{i+1} = 01 \\ \max \{ \nabla_{[1, \ldots, i]}(f)(\sigma, 0, y), \nabla_{[1, \ldots, i]}(f)(\sigma, 1, y) \} & \sigma_{i+1} = 10 \\ \nabla_{[1, \ldots, i]}(f)(\sigma, 1, y) & \sigma_{i+1} = 11. \end{cases}$$

By inductive hypothesis, the RHS. can be replaced as

$$\nabla_{[1, \ldots, i+1]}(f)(\sigma, \sigma_{i+1}, y) = \begin{cases} (\pi(\sigma) \circ f)(\varphi(\sigma), 0, y) & \sigma_{i+1} = 00 \\ \min \{ (\pi(\sigma) \circ f)(\varphi(\sigma), 0, y), (\pi(\sigma) \circ f)(\varphi(\sigma), 1, y) \} & \sigma_{i+1} = 01 \\ \max \{ (\pi(\sigma) \circ f)(\varphi(\sigma), 0, y), (\pi(\sigma) \circ f)(\varphi(\sigma), 1, y) \} & \sigma_{i+1} = 10 \\ (\pi(\sigma) \circ f)(\varphi(\sigma), 1, y) & \sigma_{i+1} = 11. \end{cases}$$

The RHS. can be further replaced, by the definition of the switch operator, as

$$\nabla_{[1, \ldots, i+1]}(f)(\sigma, \sigma_{i+1}, y) = \begin{cases} (\pi(\sigma) \circ f)(\varphi(\sigma), 0, y) & \sigma_{i+1} = 00 \\ S_{i+1}(\pi(\sigma) \circ f)(\varphi(\sigma), 0, y) & \sigma_{i+1} = 01 \\ S_{i+1}(\pi(\sigma) \circ f)(\varphi(\sigma), 1, y) & \sigma_{i+1} = 10 \\ (\pi(\sigma) \circ f)(\varphi(\sigma), 1, y) & \sigma_{i+1} = 11. \end{cases}$$

The RHS. can be written succinctly as

$$((\pi(\sigma), N) \circ f)(\varphi(\sigma), \varphi(\sigma_{i+1}), y) \quad \text{for } \sigma_{i+1} = 00, 11,$$

$$((\pi(\sigma), Y) \circ f)(\varphi(\sigma), \varphi(\sigma_{i+1}), y) \quad \text{for } \sigma_{i+1} = 01, 10,$$

which can be written even more succinctly as

$$((\pi(\sigma), \pi(\sigma_{i+1})) \circ f)(\varphi(\sigma), \varphi(\sigma_{i+1}), y),$$

completing the inductive step. \hfill \Box

Using Lemma 2.4.1, it is now easy to lower bound the distance of $g_{S, \rho}(\cdot, z)$ to strong monotonicity by the average distance of $(\rho \ast \pi) \circ f(\cdot, z)$ to monotonicity, averaged over a random choice of $\pi \in \{Y, N\}^{|S|}$.

**Lemma 2.4.2.**

$$\delta(g_{S, \rho}(\cdot, z)) \geq \mathbb{E}_{\pi \in \{Y, N\}^{|S|}} \left[ \varepsilon((\rho \ast \pi) \circ f(\cdot, z)) \right].$$

The idea is given $g$ and a strongly-monotone function $h$ close to it, associate different sub-cubes of $g$ with switched versions of $f$ by Lemma 2.4.1, and argue that the induced $n$-bit function $h$ on that sub-cube is monotone.
Proof. As before, we may drop $z$, assume that $S = \{1, \ldots, n\}$, that $\rho$ is the identity permutation, so that $g_{S, \rho} = g = \nabla_{[1, \ldots, n]} (f) : \Sigma^n \mapsto \{-1, 1\}$ and prove that
\[
\delta(g) \geq \mathbb{E}_{\pi \in \{Y, N\}^n} [\varepsilon(\pi \circ f)].
\]

Let $h : \Sigma^n \mapsto \{-1, 1\}$ be a strongly monotone function that is $\delta(g)$-close to $g$. Identify each $\pi \in \{Y, N\}^n$ with the sub-cube $V = V_\pi \subseteq \Sigma^n$ defined as
\[
V = \prod_{\pi[i] = Y} \{01, 10\} \times \prod_{\pi[i] = N} \{00, 11\}.
\]

We look at the restriction of functions $g, h, \varphi$ to the sub-cube $V$. The map $\varphi|_V : V \mapsto \{-1, 1\}^n$ is a bijection and Lemma 2.4.1 (or rather the specialized statement (2.8)) implies that
\[
g|_V (\sigma) = (\pi \circ f)(\varphi|_V (\sigma)).
\]

Since $h : \Sigma^n \mapsto \{-1, 1\}^n$ is strongly monotone, i.e. monotone under the ordering $00 \leq 01 \leq 10 \leq 11$ on $\Sigma$, it follows that $h|_V (\varphi|_V^{-1}(\cdot))$ viewed as a function on $\{0, 1\}^n$ is monotone. Thus
\[
\varepsilon(\pi \circ f) \leq \Delta(\pi \circ f, h|_V (\varphi|_V^{-1}(\cdot))) = \Delta((\pi \circ f)(\varphi(\cdot)), h|_V (\cdot)) = \Delta(g|_V, h|_V).
\]

Now taking expectations over the choice of $\pi$, we get as desired
\[
\mathbb{E}_{\pi \in \{Y, N\}^n} [\varepsilon(\pi \circ f)] \leq \mathbb{E}_{\pi \in \{Y, N\}^n} [\Delta(g|_{V_\pi}, h|_{V_\pi})] = \delta(g).
\]

\[\square\]

The Telescoping Argument

We are now ready to complete the proof of Theorem 2.1.6. Combining inequality (2.7) and Lemma 2.4.2, we get
\[
C \cdot \mathbb{E}_x \left[ \sqrt{I_f (x)} \right] \geq \mathbb{E}_z \left[ \mathbb{E}_{\pi} [\varepsilon ((\rho \ast \pi) \circ f (\cdot, z))] \right].
\]

Note that on the RHS, it is understood that there is a underlying set $S$ of coordinates, $f (\cdot, z)$ is regarded as a function of coordinates in $S$, $\rho$ is a permutation of $S$ and $\pi \in \{Y, N\}^{|S|}$. Using Definition 2.3.6, we can replace $\varepsilon (\cdot)$ by $\gamma (\cdot)$ and write
\[
2C \cdot \mathbb{E}_x \left[ \sqrt{I_f (x)} \right] \geq \mathbb{E}_z \left[ \mathbb{E}_{\pi} [\gamma ((\rho \ast \pi) \circ f (\cdot, z))] \right]. \tag{2.10}
\]

By definition of $\gamma (\cdot)$, denoting by $\tau$ a random permutation on $S$, we have
\[
\gamma ((\rho \ast \pi) \circ f (\cdot, z)) = \mathbb{E}_\tau [\Delta ((\rho \ast \pi) \circ f (\cdot, z), \tau \circ (\rho \ast \pi) \circ f (\cdot, z))],
\]

which by triangle inequality is at least
\[
\mathbb{E}_\tau [\Delta (f (\cdot, z), \tau \circ (\rho \ast \pi) \circ f (\cdot, z))] - \Delta (f (\cdot, z), (\rho \ast \pi) \circ f (\cdot, z)).
\]

Now we look at the function $\tau \circ (\rho \ast \pi) \circ f (\cdot, z)$ closely. Here, after applying the switch operator according to $\rho \ast \pi$, one applies the switch operator again according to $\tau$. The coordinates on which the switch operator was applied in the first phase, i.e. according to $\rho \ast \pi$, are already "monotonized" and these coordinates do not
mutation of

Clearly, if $\rho \neq \pi$, then the negatively sensitive edges of $S$ are affected in the same manner. We only point out the minor differences. Let $\rho \star \pi = (5, 1, 2, 3)$. Then we have $\lambda = \lambda(\tau, \rho, \pi) = (5, 1, 2, 3, 4)$.

Taking the expectation of inequality (2.10) over a random choice of $\rho$, we thus see that

$$2 \cdot C \cdot \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \Psi_f(p) - \Psi_f \left( \frac{P}{2} \right),$$

where $\Psi_f(p)$ denotes the expected change in $f$ by considering the coordinates $1$ through $n$ in a random order and applying the switch operator on each coordinate with probability $p$ independently. The reason is that in the first expectation, each coordinate is not-restricted (in $S$) with probability $p$; thus $\lambda \circ f(\cdot, z)$ is the function $f$ where each switch is performed with probability $p$. Similarly, in the second expectation each switch is performed with probability $p \cdot \frac{1}{2}$.

Considering this lower bound for $p = 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{2^s}$ where $s = \lceil 5 \log n + 5 \log(1/\varepsilon(f)) \rceil$, and using telescoping sum, we get that

$$(s + 1) \cdot 2 \cdot C \cdot \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \Psi_f(p = 1) - \Psi_f \left( \frac{1}{2^{s+1}} \right) \geq \varepsilon(f) - \frac{1}{2} \cdot \varepsilon(f) \geq \frac{1}{2} \cdot \varepsilon(f),$$

and hence

$$\mathbb{E}_x \left[ \sqrt{I_f(x)} \right] \geq \Omega \left( \frac{\varepsilon(f)}{\log n + \log(1/\varepsilon(f))} \right),$$

proving Theorem 2.1.6. We noted that $\Psi_f(p = 1) = \gamma(f) \geq \varepsilon(f)$ and for the choice of $s$, with probability $1 - \frac{1}{2} \cdot \varepsilon(f)$, no switch is applied at all (not even on a single coordinate), so $\Psi_f \left( \frac{1}{2^{s+1}} \right) \leq \frac{1}{2} \cdot \varepsilon(f)$.

## 2.5 Proof of Theorem 2.1.9

Now that we have proved Theorem 2.1.6, we observe that the proof of Theorem 2.1.9 follows in essentially the same manner. We only point out the minor differences. Let $\text{col} : S \rightarrow \{\text{red}, \text{blue}\}$ be any coloring of the negatively sensitive edges of $f$. We intend to lower bound

$$\mathbb{E}_x \left[ \sqrt{I_{f, \text{red}}(x)} \right] + \mathbb{E}_x \left[ \sqrt{I_{f, \text{blue}}(x)} \right].$$

Similar to the beginning of Section 2.4.2, thinking of the input $x$ as $x = (w, z)$ w.r.t. an underlying partition of the coordinates as $S \cup S^c$, we get the lower bound

$$\mathbb{E}_z \left[ \mathbb{E}_w \left[ \sqrt{I_{f(\cdot, z), \text{red}}(w)} \right] + \mathbb{E}_w \left[ \sqrt{I_{f(\cdot, z), \text{blue}}(w)} \right] \right].$$
The reason is the same as before, that we are now ignoring edges along the coordinates in $S$. We now consider the function $g_{S,\rho}$ as before. Using Lemma 2.3.18, there is a coloring of negatively sensitive edges of $g_{S,\rho}(\cdot, z)$ so that we get a lower bound

$$
\mathbb{E}_z \left[ \mathbb{E}_\sigma \left[ \sqrt{I_{g_{S,\rho}(\cdot, z), \text{red}}(\sigma)} \right] + \mathbb{E}_\sigma \left[ \sqrt{I_{g_{S,\rho}(\cdot, z), \text{blue}}(\sigma)} \right] \right].
$$

By Lemma 2.3.14 and 2.3.16, we get a lower bound, up to a constant factor,

$$
\mathbb{E}_z \left[ \delta(g_{S,\rho}(\cdot, \cdot, z)) \right].
$$

Now Sections 2.4.2 and 2.4.2 only concern a lower bound on this last quantity (and taking expectation over the choice of $S, \rho$ and telescoping) and we are done. We note that even though the proof in Section 2.4.2 is written with reference to the quantity $\mathbb{E}_x \left[ \sqrt{I_f(x)} \right]$, one only needs to replace it with the analogue $\mathbb{E}_x \left[ \sqrt{I_{f, \text{red}}(x)} \right] + \mathbb{E}_x \left[ \sqrt{I_{f, \text{blue}}(x)} \right]$ everywhere.

### 2.6 Some Useful Results

#### 2.6.1 A Combinatorial Fact

**Definition 2.6.1.** A bipartite graph $G(V, W, E)$ is called right-$(K, d)$-good if $|W| \geq K$, degree of every vertex in $W$ is in the range $[d, 2d]$ and degree of every vertex in $V$ is at most $2d$.

**Definition 2.6.2.** A bipartite graph $G(V, W, E)$ is called left-$(K, d)$-good if $|V| \geq K$, degree of every vertex in $V$ is in the range $[d, 2d]$ and degree of every vertex in $W$ is at most $2d$.

**Definition 2.6.3.** A bipartite graph $G(V, W, E)$ is called $(K, d)$-good if it is either right-$(K, d)$-good or left-$(K, d)$-good.

**Definition 2.6.4.** A bipartite graph $G(V, W, E)$ is called $L$-robust if for any 2-coloring of its edges $\text{col} : E \mapsto \{\text{red}, \text{blue}\}$, we have

$$
\sum_{v \in V} \sqrt{D_{\text{red}}(v)} + \sum_{w \in W} \sqrt{D_{\text{blue}}(w)} \geq L.
$$

Here $D_{\text{red}}(v)$ denotes the number of red edges incident on $v$ and $D_{\text{blue}}(w)$ denotes the number of blue edges incident on $w$.

**Lemma 2.6.5.** If a bipartite graph $G(V, W, E)$ is $L$-robust and all its vertices have degree less than $2^s$, then it has a subgraph that is $(K, d)$-good with $K \sqrt{d} \geq \frac{L}{8s}$.

**Proof.** We search for the desired good subgraph by looking at a decreasing sequence of subgraphs of $G(V, W, E)$. Let $G_0(V_0, W_0, E_0) = G(V, W, E)$ be the starting graph. For $j = 0, 1, 2, \ldots$, we assume that $G_j(V_j, W_j, E_j)$ is $(L - j \cdot \frac{L}{2s})$-robust and all its vertices have degree less than $2^{s-j}$. Then we show that either $G_j(V_j, W_j, E_j)$ has a subgraph that is $(K, d)$-good with $K \sqrt{d} \geq \frac{L}{8s}$ (in which case we are done and we can stop) or else has a subgraph $G_{j+1}(V_{j+1}, W_{j+1}, E_{j+1})$ that is $(L - (j+1) \cdot \frac{L}{2s})$-robust and has all degrees less than $2^{s-j-1}$ and we resume the next iteration. This iterative process must find a good subgraph since otherwise for $j = s$, the graph $G_s$ will have no edges and still be $L/2$-robust.

We prove the iterative claim. In the graph $G_j(V_j, W_j, E_j)$, let $A \subseteq V_j, B \subseteq W_j$ be the sets of vertices whose degree is in the range $[2^{s-j-1}, 2^{s-j})$. Let $V_{j+1} = V_j \setminus A, W_{j+1} = W_j \setminus B$, and $G_{j+1}(V_{j+1}, W_{j+1}, E_{j+1})$ be the induced subgraph of $G_j(V_j, W_j, E_j)$ on vertex set $(V_{j+1}, W_{j+1})$. Note that the degrees of vertices in $G_{j+1}$ are less than $2^{s-j-1}$.
Let $H(A, W_j, E_H)$ be the induced subgraph of $G_j(V_j, W_j, E_j)$ on the vertex set $(A, W_j)$. With $d = 2^{s-j-1}$, note that in the graph $H(A, W_j, E_H)$, every vertex in $A$ has degree in the range $[d, 2d]$ and every vertex in $W_j$ has degree at most $2d$. Thus $H$ is left-$(|A|, d)$-good and if $|A| \sqrt{d} \geq \frac{L}{8}$, we are done. Hence we may assume that $|A| \sqrt{d} \leq \frac{L}{8}$. Similarly, we may assume that $|B| \sqrt{d} \leq \frac{L}{8}$. Now we prove that $G_j(V_{j+1}, W_{j+1}, E_{j+1})$ is $(L - (j + 1) \cdot \frac{L}{2s})$-robust.

Consider any \{red, blue\}-coloring of edges of $G_{j+1}(V_{j+1}, W_{j+1}, E_{j+1})$. Extend this to a coloring of edges of $G_j(V_j, W_j, E_j)$ by coloring all edges between $A$ and $W_{j+1}$ as red, all edges between $B$ and $V_{j+1}$ as blue, and coloring edges between $A$ and $B$ arbitrarily. Using the fact that $G_j$ is $(L - j \cdot \frac{L}{2s})$-robust, we get (degrees below are degrees in $G_j$):

$$L - j \cdot \frac{L}{2s} \leq \sum_{v \in V_j} \sqrt{D_{\text{red}}(v)} + \sum_{w \in W_j} \sqrt{D_{\text{blue}}(w)}$$

$$= \sum_{v \in V_{j+1}} \sqrt{D_{\text{red}}(v)} + \sum_{v \in A} \sqrt{D_{\text{red}}(v)} + \sum_{w \in W_{j+1}} \sqrt{D_{\text{blue}}(w)} + \sum_{w \in B} \sqrt{D_{\text{blue}}(w)}$$

Using the upper bound on $|A|$ and $|B|$ as above, we get

$$\sum_{v \in V_{j+1}} \sqrt{D_{\text{red}}(v)} + \sum_{w \in W_{j+1}} \sqrt{D_{\text{blue}}(w)} \geq L - (j + 1) \cdot \frac{L}{2s}. $$

Noting that for vertices in $V_{j+1}$ (resp. $W_{j+1}$) their red-degree (resp. blue-degree) is same in the graphs $G_j$ and $G_{j+1}$, the claim follows. \hfill \Box

2.6.2 Bound on Total Influence

Recall that for a function $f : \{0, 1\}^n \mapsto \{0, 1\}$, $S_f$ denotes the set of sensitive edges and $S_f^- \subseteq S_f$ denotes the set of negatively sensitive edges. The total influence is $I_f = \frac{|S_f|}{2^n}$ and the total negative influence is $I_f^- = \frac{|S_f^-|}{2^n}$. We show that if $I_f^- \ll I_f$ then $I_f$ is at most $O(\sqrt{n})$. This upper bound on the total influence is very useful in the analysis of our monotonicity tester. Formally:

**Theorem 2.6.1.**

$$I_f^- \leq \frac{1}{3} \cdot I_f \implies I_f \leq 6\sqrt{n}. $$

**Proof.** For coordinate $i \in \{1, \ldots, n\}$, let $(x_{-i}, b)$ denote an input that equals $b$ in the $i$th coordinate and $x_{-i} \in \{0, 1\}^{n-1}$ in the remaining coordinates. Let $1_{\cdot}$ denote the indicator of an event. One observes that the Fourier coefficient $\hat{f}(\{i\})$ is, by definition:

$$2 \cdot \hat{f}(\{i\}) = \mathbb{E}[f(x_{-i}, 0) - f(x_{-i}, 1)] = \mathbb{E} \left[ 2 \cdot \mathbb{1}_{f(x_{-i}, 0) \neq f(x_{-i}, 1)} \cdot \mathbb{1}_{f(x_{-i}, 0) > f(x_{-i}, 1)} - \mathbb{1}_{f(x_{-i}, 0) = f(x_{-i}, 1)} \right]. $$

Rearranging and summing over $i \in \{1, \ldots, n\}$ gives

$$\sum_{i=1}^{n} \mathbb{E} \left[ \mathbb{1}_{f(x_{-i}, 0) \neq f(x_{-i}, 1)} \right] = 2 \cdot \sum_{i=1}^{n} \mathbb{E} \left[ \mathbb{1}_{f(x_{-i}, 0) > f(x_{-i}, 1)} \right] - 2 \cdot \sum_{i=1}^{n} \hat{f}(\{i\}).$$
Observing that the first and the second sums are precisely $I_f$ and $I_f^-$ respectively and taking absolute values,

$$I_f \leq 2 \cdot I_f^- + 2 \cdot \sum_{i=1}^{n} |\hat{f}(\{i\})|.$$ 

Now if $I_f^- \leq \frac{1}{3} I_f$, it follows that $I_f$ is at most $6 \cdot \sum_{i=1}^{n} |\hat{f}(\{i\})|$, which by Cauchy-Schwarz is at most

$$6\sqrt{n} \sum_{i=1}^{n} \hat{f}(\{i\})^2 \leq 6\sqrt{n}.$$ 

\[ \square \]

### 2.6.3 Bound on Fraction of Non-persistent Inputs

Fix an integer parameter $\tau \in \left[1, \sqrt{\frac{n}{\log n}}\right]$ and a function $f : \{0, 1\}^n \mapsto \{0, 1\}$. For an input $x \in \{0, 1\}^n$, a random input $y$ that is monotonically above $x$ at distance $\tau$ is picked by changing $\tau$ coordinates of $x$ from 0 to 1, picked uniformly at random from the set of 0-coordinates of $x$. In the following, we assume that all inputs under consideration have Hamming weight in the range $\frac{1}{2}n \pm O(\sqrt{n \log n})$, since the fraction of remaining “atypical” inputs is at most, say $\frac{1}{10n}$, and ignoring these atypical inputs has no effect on our arguments.

**Definition 2.6.6.** An input $x \in \{0, 1\}^n$ is called $\tau$-persistent if over the choice of a random input $y$ that is monotonically above $x$ at distance $\tau$,

$$\Pr_{x,y}[f(x) \neq f(y)] \leq \frac{1}{10}.$$ 

The following lemma bounds the fraction of non-persistent inputs in terms of the total influence of the function. Here is an intuitive argument. For a non-persistent input, changing $\tau$ of its coordinates changes the value of the function with a constant probability. Thus, roughly speaking, changing one of its coordinates changes the value of the function with probability $\Omega(1/\tau)$ and hence, the total influence is at least the fraction of non-persistent inputs multiplied by $\Omega(n/\tau)$. However, a formal proof offers some subtleties.

**Lemma 2.6.7.** Let $\tau \in \left[1, \sqrt{\frac{n}{\log n}}\right]$. The fraction of $\tau$-non-persistent inputs is at most $O\left(\frac{n}{\tau} \cdot \tau\right)$.

**Proof.** Let $\alpha$ be the fraction of non-persistent inputs. Consider the random process that picks input $x$ uniformly at random and then picks, at random, input $y$ that is monotonically above $x$ by distance $\tau$. We consider the probability that $f(x) \neq f(y)$. Clearly,

$$\Pr_{x,y}[f(x) \neq f(y)] = \frac{1}{2^n} \cdot \sum_{x \in \{0,1\}^n} \Pr_y[f(x) \neq f(y)] \geq \alpha \cdot \frac{1}{10}, \quad (2.11)$$

since for a non-persistent input $x$, the probability $\Pr_y[f(x) \neq f(y)]$ is at least $\frac{1}{10}$. Let $x = x^0, x^1, \ldots, x^\tau = y$ denote the sequence of inputs, starting with $x$ and changing one coordinate at a time (from ‘0’ to ‘1’) till one reaches $y$. The order in which the coordinates are changed is itself random. Clearly,

$$\Pr_{x,y}[f(x) \neq f(y)] \leq \sum_{\ell=0}^{\tau-1} \Pr[f(x^\ell) \neq f(x^{\ell+1})]. \quad (2.12)$$
We note that while moving from \( x^\ell \) to \( x^{\ell+1} \) the coordinate that is changed is itself random among the ‘0’-coordinates of \( x^\ell \). The distribution of the input \( x^\ell \) is not necessarily uniform (but close to it, as we will see). We observe that the lemma follows immediately if we were to assume that the distribution of \( x^\ell \) is uniform. Indeed, supposing so, the distribution of the pair \((x^\ell, x^{\ell+1})\) is same as the distribution of the pair \((u, v)\) where \( u \) is uniformly random and \( v \) is obtained by changing a random ‘0’ coordinate of \( u \) to ‘1’. Thus, we would have, combining Equations (2.11), (2.12),

\[
\alpha/10 \leq \sum_{\ell=0}^{\tau-1} Pr_{u,v}[f(u) \neq f(v)] = \tau \cdot \frac{I_f}{n},
\]

completing the proof of the lemma. We now show how to handle the issue that the distribution of \( x^\ell \) is not necessarily uniform. We show that for any fixed input \( z \in \{0,1\}^n \) with Hamming weight in the range \( \frac{1}{2}n \pm O(\sqrt{n \log n}) \),

\[
Pr[z = x^\ell] \leq C \cdot Pr[z = u],
\]

for some constant \( C \), i.e. the probability of sampling \( z \) from the distribution corresponding to that of \( x^\ell \) is at most a constant times the probability of sampling \( z \) from the uniform distribution. Indeed, suppose the Hamming weight of \( z \) is \( \frac{1}{2}n + k \) for some \( k \in [-O(\sqrt{n \log n}), O(\sqrt{n \log n})] \). We note that if \( z = x^\ell \), then \( x \) must have Hamming weight \( \frac{1}{2}n + k - \ell \) and since \( x \) is uniformly random,

\[
Pr[z = x^\ell] = \left( \frac{1}{2^n} \cdot \left( \frac{n}{\frac{1}{2}n + k - \ell} \right) \right) \cdot \frac{1}{\binom{n}{\frac{1}{2}n + k}}.
\]

Here the first factor reflects the probability that \( x \) has Hamming weight \( \frac{1}{2}n + k - \ell \) and the second factor reflects the probability that \( z \) happens to be one specific input among all inputs with Hamming weight \( \frac{1}{2}n + k \). By Lemma 2.6.8, \( Pr[z = x^\ell] \) is at most a constant times \( \frac{1}{2^n} = Pr[z = u] \) as claimed. Combining Equations (2.11), (2.12) as before,

\[
\frac{\alpha}{10} \leq \sum_{\ell=0}^{\tau-1} Pr[f(x^\ell) \neq f(x^{\ell+1})] \leq \sum_{\ell=0}^{\tau-1} C \cdot Pr[f(u) \neq f(v)] = \tau \cdot C \cdot \frac{I_f}{n},
\]

completing the proof of the lemma.

**Lemma 2.6.8.** Consider integers \( \ell \in \left[0, \sqrt{\frac{n}{\log n}}\right] \) and \( k \in [-O(\sqrt{n \log n}), O(\sqrt{n \log n})] \). Then layers \( \frac{1}{2}n + k \) and \( \frac{1}{2}n + k - \ell \) of the \( n \)-dimensional hypercube have the same number of vertices up to a constant factor. That is:

\[
\left( \frac{\binom{n}{\frac{1}{2}n + k - \ell}}{\binom{n}{\frac{1}{2}n + k}} \right) = \Theta(1).
\]

**Proof.**

\[
\left( \frac{\binom{n}{\frac{1}{2}n + k - \ell}}{\binom{n}{\frac{1}{2}n + k}} \right) = \frac{(\frac{1}{2}n + k)! (\frac{1}{2}n - k)!}{(\frac{1}{2}n + k - \ell)! (\frac{1}{2}n - k + \ell)!} = \frac{(\frac{1}{2}n + k)(\frac{1}{2}n + k - 1)...(\frac{1}{2}n + k - \ell + 1)}{(\frac{1}{2}n - k + \ell)(\frac{1}{2}n - k + \ell - 1)...(\frac{1}{2}n - k + 1)} = \prod_{j=0}^{\ell-1} \left( 1 + \frac{2k - \ell}{\frac{1}{2}n - k + \ell - j} \right) = \left( 1 \pm O\left( \sqrt{\frac{\log n}{n}} \right) \right)^\ell = \Theta(1),
\]

since \(|2k - \ell|\) and \(|k - \ell + j|\) are \( O(\sqrt{n \log n}) \) and \( \ell \) is at most \( \sqrt{\frac{n}{\log n}} \).
2.7 Algorithm for Monotonicity Testing

In this section, we present our monotonicity testing algorithm. Let \( f : \{0, 1\}^n \mapsto \{0, 1\} \) be the given function that is \( \varepsilon(f) \)-far from being monotone. The tester is a simple path tester that picks a distance parameter \( \tau \in \{1, \ldots, \lceil \sqrt{n} \rceil\} \) from a certain distribution and tries to detect a violation to monotonicity on a pair \((x, y)\) of inputs such that \( y \) is monotonically above \( x \) at distance \( \tau \). Using Theorem 2.1.9, we deduce existence of a good subgraph of violated edges and then the tester is analyzed with reference to this good subgraph (note that the good subgraph is used for purposes of the analysis only).

As we will see, we only need to consider the case when \( \varepsilon(f) \geq \frac{1}{\sqrt{n}} \), hence we make this assumption henceforth and in particular, \( \log(1/\varepsilon(f)) \) is \( O(\log n) \). Secondly, we can ignore “atypical” inputs, i.e. those whose Hamming weight is outside the range \( \frac{n}{2} \pm 10\sqrt{n \log n} \). The fraction of these inputs is at most, say \( \frac{1}{n^3} \), and ignoring them does not affect the analysis. One can get around this issue in a formal manner also, as follows. One may pretend that instead of a query access to the given function \( f \), we have a query access to the function \( \bar{f} \) that is same as \( f \) except that on inputs of Hamming weight less than \( \frac{n}{2} - 10\sqrt{n \log n} \), \( \bar{f} = 0 \) and on inputs of Hamming weight greater than \( \frac{n}{2} + 10\sqrt{n \log n} \), \( \bar{f} = 1 \). It is easily seen that the atypical inputs do not participate in any violating pair \((x, y)\) for \( \bar{f} \). One may carry out the analysis for \( \bar{f} \) giving a lower bound on the rejection probability. Since the rejection probability amounts to detecting a violating pair and any violating pair for \( \bar{f} \) is also a violating pair for \( f \), the same lower bound on rejection probability applies to \( f \).

2.7.1 Existence of a Good Subgraph

We recall that \( G_f^- (V, W, E) \) denotes the bipartite graph of negatively sensitive edges, i.e. \( V, W \subseteq \{0, 1\}^n \), \( \forall x \in V \ f(x) = 1, \forall y \in W \ f(y) = 0, E \) is precisely the set of negatively sensitive edges \( S_f^- \), and every vertex in \( V \cup W \) has at least one negatively sensitive edge incident on it.

Theorem 2.1.9 amounts to saying that the graph \( G_f^- (V, W, E) \) is \( \Omega(2^n \cdot \frac{\varepsilon(f)}{\log n}) \)-robust. We note that the degrees of vertices in \( G_f^- \) are at most \( n \), hence less than \( 2^n \) with \( s = \lceil \log n \rceil + 1 \). Applying Lemma 2.6.5, we get that \( G_f^- (V, W, E) \) has a subgraph \( G_{\text{good}}(A, B, E_{AB}) \) that is \((K, d)\)-good where \( K \sqrt{d} \geq \Omega(\frac{2^n \cdot \varepsilon(f)}{\log^2 n}) \).

We assume w.l.o.g. that \( G_{\text{good}} \) is a \((K, d)\)-good graph with \(|B| = K \) (see Remark 2.7.2). By definition of a good graph, every vertex in \( A \) has degree at most \( 2d \) and every vertex in \( B \) has degree in the range \([d, 2d]\). By deleting some edges if necessary, we may assume that every vertex in \( B \) has degree exactly \( d \). By deleting some vertices from \( B \) if necessary, we may assume that \( K \sqrt{d} = \Theta(\frac{2^n \cdot \varepsilon(f)}{\log^2 n}) \). Finally, we write \( K = \sigma \cdot 2^n \) for convenience. We summarize the conclusion below.

**Lemma 2.7.1.** \( G_f^- (V, W, E) \) contains a subgraph \( G_{\text{good}}(A, B, E_{AB}) \) such that \(|B| = \sigma \cdot 2^n \), vertices in \( B \) have degree exactly \( d \), vertices in \( A \) have degree at most \( 2d \) and \( \sigma \sqrt{d} = \Theta(\frac{\varepsilon(f)}{\log^2 n}) \).

2.7.2 The Tester

Let \( p \) be the largest integer such that \( 2^p \leq \sqrt{\frac{n}{\log n}} \). Our tester works as follows:

1. Pick an integer \( k \in \{0, 1, 2, \ldots, p\} \) uniformly at random and let \( \tau = 2^k \).
2. Pick an input \( x \in \{0, 1\}^n \) uniformly at random.
3. Let \( S \subseteq \{1, \ldots, n\} \) be the set of 0-coordinates of \( x \). Pick a random subset \( T \subseteq S, |T| = \tau \) and obtain \( z \) by changing coordinates of \( x \) in \( T \) to 1.
4. Reject if and only if $f(x) = 1$ and $f(z) = 0$.

**Remark 2.7.2.** (i) In Step (3), as remarked before, we may assume that the Hamming weight of $x$ is in the range $\frac{n}{2} \pm 10\sqrt{n} \log n$, so $x$ does have enough 0-coordinates. (ii) The description of the tester and the analysis are written assuming that the good subgraph, as in Lemma 2.7.1, is a right-good subgraph. If it were a left-good subgraph, the tester (and the analysis) would work in the anti-symmetric manner, by letting the input $y$ to be monotonically below $x$ by a distance $\tau$ and then rejecting if $f(x) = 0$, $f(y) = 1$. Formally, the tester would pick one of the two options with probability $\frac{1}{2}$ each and proceed.

### 2.7.3 The Analysis of the Tester

We intend to show that the tester rejects with probability $\tilde{\Omega}(\frac{\epsilon(f)^2}{\sqrt{n}})$. We first make a series of preliminary observations and then proceed to the main analysis. We note first that the tester, when $\epsilon(f) \geq \frac{1}{\sqrt{n}}$, actually runs the edge tester and this happens with probability $\frac{1}{\sqrt{n}}$ over the choice of $x$. Thus our tester already includes, implicitly, the edge tester.

**The Lower Bound** $\epsilon(f) \geq \frac{1}{\sqrt{n}}$:

The edge tester has rejection probability $\Omega(\frac{\epsilon(f)}{\sqrt{n}})$ as shown by Goldreich et al [GGL+00] and when $\epsilon(f) \leq \frac{1}{\sqrt{n}}$, it also qualifies as a tester with rejection probability $\Omega(\frac{\epsilon(f)^2}{\sqrt{n}})$. Thus we assume henceforth that $\epsilon(f) \geq \frac{1}{\sqrt{n}}$.

**The Upper Bound** $I_f \leq 6\sqrt{n}$:

If the total negative influence $I_f^-$ is large, i.e. if $I_f^- \geq \sqrt{n}$, then the edge tester rejects with probability $\frac{I_f^-}{\sqrt{n}}$ which is $\Omega(\frac{1}{\sqrt{n}})$ and we are done. Thus we assume henceforth that $I_f^- \leq \sqrt{n}$. By Theorem 2.6.1, we then have an upper bound on the total influence, i.e. $I_f \leq \max\{3 \cdot I_f^-, 6\sqrt{n}\} = 6\sqrt{n}$.

**The Lower Bound** $\sigma \geq \frac{\log n}{\sqrt{n}}$:

We know that $\sigma \sqrt{d} = \Theta(\frac{\epsilon(f)}{\log n})$. If $\sigma \leq \frac{\log n}{\sqrt{n}}$, we get that

$$\sigma \cdot d = \frac{(\sigma \sqrt{d})^2}{\sigma} \geq \Omega\left(\frac{\epsilon(f)^2 \cdot \sqrt{n}}{\log^2 n}\right).$$

The graph $G_{\text{good}}(A, B, E_{AB})$ has exactly $|B| \cdot d = \sigma \cdot 2^n \cdot d$ edges and all these are violating edges. Thus $I_f^- \geq \sigma \cdot d \geq \tilde{\Omega}(\epsilon(f)^2 \cdot \sqrt{n})$. As before, the edge tester rejects with probability $\frac{I_f^-}{\sqrt{n}}$ which is $\tilde{\Omega}(\frac{\epsilon(f)^2}{\sqrt{n}})$ and we are done. Thus we assume henceforth that $\sigma \geq \frac{\log n}{\sqrt{n}}$.

**Choosing $\tau$ that “Works” and Persistence of $B$-vertices:**

Since $\tau$ takes values that are powers of 2, we can fix a value of $\tau$ such that

$$\tau \leq \sigma \sqrt{\frac{n}{\log n}} \leq 2 \tau.$$
Using the bounds \( \frac{\log n}{\sqrt{n}} \leq \sigma \leq 1 \), it holds that \( \frac{1}{2} \cdot \sqrt{\log n} \leq \tau \leq \sqrt{\frac{n}{\log n}} \). We will show that the tester “works” when this specific value of \( \tau \) is chosen by the tester, which happens with probability \( \Theta(\frac{1}{\log n}) \).

We call a vertex \( y \in B \) “(\( \tau - 1 \))-persistent” if changing \( \tau - 1 \) of its coordinates at random from 0 to 1 changes the value of the function from 0 to 1 with probability at most \( \frac{1}{10} \). By Lemma 2.6.7, the fraction of vertices \( y \in \{0, 1\}^n \) that are not \( (\tau - 1) \)-persistent is bounded by \( O(\frac{f^2}{n} \cdot \tau) \), which using upper bounds \( I_f \leq 6\sqrt{n} \) and \( \tau \leq \sigma \sqrt{\frac{n}{\log n}} \), is upper bounded by \( \frac{\sigma}{100} \). Since \( B \) constitutes \( \sigma \) fraction of vertices in \( \{0, 1\}^n \), at least a \( \frac{\sigma}{100} \) fraction of the vertices in \( B \) are \( (\tau - 1) \)-persistent. We retain only the persistent vertices in \( B \) and assume henceforth that all vertices in \( B \) are \( (\tau - 1) \)-persistent, redefining the parameter \( \sigma \) to reflect the new, slightly reduced, size of \( B \).

**Main Analysis**

We are now ready to present the main argument in our analysis. Let \( G_{\text{good}}(A, B, E_{AB}) \) be the good graph of violated edges. Let \( \tau \) be the specifically chosen parameter as above. The tester picks an input \( x \in \{0, 1\}^n \) uniformly at random and then picks input \( z \) at random that is monotonically above \( x \) by a distance \( \tau \). We consider the probability of the following event \( \mathcal{R} \). We note that in case of event \( \mathcal{R} \), the tester does detect a violation of monotonicity and rejects. Thus, the probability of event \( \mathcal{R} \) is a lower bound on the rejection probability of the tester. When computing the probability of the event \( \mathcal{R} \) however, one needs to be careful to avoid “double-counting” as explained towards the end of the analysis.

**The event \( \mathcal{R} \)**

- \( x \in A \) and hence \( f(x) = 1 \).
- There is a unique \( y \) such that \( (x, y) \in E_{AB} \) (i.e. \( x \leq y \) and \( y \in B \)) and moreover \( y \leq z \).
- \( f(y) = f(z) = 0 \).

Fix some \( x \in A \). Let \( s = \deg_A(x) \leq 2d \) be the degree of \( x \) in the graph \( G_{\text{good}}(A, B, E_{AB}) \) and let \( \{y_1, \ldots, y_s\} \) be the set of neighbors of \( x \) in this graph. We note that \( d \leq n \) and

\[
s \cdot \tau \leq 2d \cdot \sigma \sqrt{\frac{n}{\log n}} = O(\sigma \sqrt{d}) \cdot \sqrt{\frac{dn}{\log n}} \leq O\left(\frac{1}{\log^2 n}\right) \cdot \frac{n}{\sqrt{\log n}}.
\]

Consider the choice of a random \( z \) that is monotonically above \( x \) by a distance \( \tau \). This amounts to changing, at random, \( \tau \) of the coordinates of \( x \) from 0 to 1. The probability that \( y_i \leq z \) for some \( i \in \{1, \ldots, s\} \) is \( \Omega\left(\frac{s \cdot \tau}{n}\right) \). This reflects the probability that while changing \( \tau \) of the coordinates from \( \approx \frac{d}{2} \) 0-coordinates of \( x \), one of the \( s \) relevant coordinates corresponding to its neighbors \( y_1, \ldots, y_s \) gets changed. It is important that \( s \cdot \tau \ll n \) for this argument to work. On the other hand, the probability that \( y_i \leq z \) for two or more indices \( i \in \{1, \ldots, s\} \) is at most \( O(s^2 \cdot \frac{\tau^2}{n^2}) \), which is negligible compared to the probability that there is at least one such index. In other words, whenever \( y_i \leq z \) for some \( i \in \{1, \ldots, s\} \), such an index \( i \) is likely unique. Moreover, for any fixed \( i \in \{1, \ldots, s\} \), conditioning the choice of \( z \) so that \( y_i \leq z \), the choice of \( z \) amounts to changing, at random, \( \tau - 1 \) of the coordinates of \( y_i \) from 0 to 1. Since \( y_i \in B \) is \( (\tau - 1) \)-persistent, with probability \( \frac{9}{10} \) over the choice of \( z \) (conditional on \( y_i \leq z \)), it holds that \( f(z) = f(y_i) = 0 \).

The discussion above shows that for a fixed \( x \in A \), the probability of event \( \mathcal{R} \) over the choice of \( z \) is \( \Omega\left(\frac{\deg_A(x) \cdot \tau}{n}\right) \). Hence the overall probability of event \( \mathcal{R} \) can be lower bounded as:

\[
\Pr[\mathcal{R}] = \frac{1}{2^n} \sum_{x \in A} \Omega\left(\frac{\deg_A(x) \cdot \tau}{n}\right) = \Omega\left(\frac{\tau}{n}\right) \cdot \frac{1}{2^n} \sum_{x \in A} \deg_A(x) = \Omega\left(\frac{\tau}{n}\right) \cdot \frac{|B| \cdot d}{2^n} = \Omega\left(\frac{\tau}{n}\right) \cdot \sigma \cdot d,
\]
which by substituting \( \tau = \Theta\left(\sqrt{\frac{n}{\log n}}\right) \) and \( \sigma^2 d = (\sigma \sqrt{d})^2 \geq \Omega\left((\frac{\varepsilon(f)}{\log n})^2\right) \), gives a lower bound of \( \Omega\left(\frac{\varepsilon(f)^2}{\sqrt{n}}\right) \) on \( \Pr[R] \). We now note that whenever the event \( R \) occurs, the tester detects that \((x, z)\) is a violating pair. Moreover, the pair \((x, z)\) uniquely determines the edge \((x, y) \in E_{AB} \) with \( x \leq y \leq z \) and the violating pair \((x, z)\) can be “credited” to the edge \((x, y) \in E_{AB} \). The sets of violating pairs credited to different edges in \( E_{AB} \) are disjoint and there is no “double-counting”. This shows that \( \Pr[R] \) is also a lower bound on the rejection probability of the tester.

### 2.8 A Lower Bound for Monotonicity Testing

In this section, we prove Theorem 2.1.12. For a subset \( S \subseteq \{1, 2, \ldots, 4n\}, |S| = 2n \) and an index \( j \in \{1, 2, \ldots, 4n\}, j \notin S \), define a function \( f_{S,j} : \{0, 1\}^{4n} \rightarrow \{0, 1\} \) as follows: for input \( x \in \{0, 1\}^{4n} \), denoting by \( x_S \), its restriction to coordinates in \( S \),

- If \( x_S \) has Hamming weight less than \( n \), \( f(x) = 0 \).
- If \( x_S \) has Hamming weight larger than \( n \), \( f(x) = 1 \).
- If \( x_S \) has Hamming weight exactly \( n \), \( f(x) = 1 - x_j \).

In short, \( f_{S,j}(x) \) is the majority function on \( x_S \) except in the “middle layer” where it is the anti-dictatorship of coordinate \( j \). It is easily seen that \( \varepsilon(f_{S,j}) = \Theta\left(\frac{1}{\sqrt{n}}\right) \) (to turn \( f_{S,j} \) into a monotone function, the best strategy, up to a constant factor, is to make it 0 in the middle layer). Consider the family \( F \) of functions \( f_{S,j} \) over all choices of \( S, |S| = 2n \) and \( j \notin S \).

We show that for any “reasonable” pair of inputs \((x, y)\) such that \( y \) is monotonically above \( x \), the probability that \((x, y)\) is a violating pair for the function \( f_{S,j} \) is \( O(1/n^{3/2}) \) over a randomly chosen function \( f_{S,j} \in F \). It then follows that any pair tester that queries \( o(n^{3/2}) \) pairs, has only \( o(1) \) rejection probability on some function in \( F \), completing the proof. A pair \((x, y)\) is “reasonable” if Hamming weights of \( x, y \) are in the range \( 2n \pm O(\sqrt{n \log n}) \). The fraction of inputs that participate in an unreasonable pair is polynomially small in \( n \) and hence we may ignore unreasonable pairs without affecting our argument.

Fix any reasonable pair \((x, y)\) where \( y \) is monotonically above \( x \) by a distance \( \tau \). Note that the pair is violating pair for the function \( f_{S,j} \) if and only if \( x_S = y_S \) with Hamming weight exactly \( n \), \( x_j = 0 \) and \( y_j = 1 \). Denoting by \( D_{x,y} \) the set of coordinates where \( x \) and \( y \) differ, \( |D_{x,y}| = \tau \) and this is a violating pair if and only if

\[
\text{\( x_S \) has Hamming weight \( n \) and \( S \cap D_{x,y} = \emptyset \) and \( j \in D_{x,y}. \)}
\]

The probability over the choice of \( S \) of the second event is \( 2^{-O(\tau)} \). Conditioned on the second event, the distribution of the Hamming weight of \( x_S \) is hypergeometric, where the probability it equals \( k \) is \( |x| \) denotes the Hamming weight of \( x \):

\[
p(k) \overset{\text{def}}{=} \binom{|x|}{k} \binom{4n-|x| - \tau}{2n-k} \frac{1}{\binom{4n}{2n}}.
\]

A direct computation (in the spirit of Lemma 2.6.8) shows that \( \frac{p(n+k)}{p(n)} = \Theta(1) \) for \( 0 \leq k \leq \sqrt{n \log n} \). Since \( p(n + k) \) sums to at most 1 in that range of \( k \), we have that each one of the terms is \( O\left(\sqrt{\frac{\log n}{n}}\right) \), and in particular \( p(n) = O\left(\sqrt{\frac{\log n}{n}}\right) \).

Conditioned on the previous two, the probability of the third event is \( \frac{\tau}{4n-|S|} = O\left(\frac{\tau}{n}\right) \).
Overall, the probability that the above event happens for a random choice of $S, j$ is at most $O(\sqrt{\log n / n})$. This probability is maximized when $\tau = 1$ and is at most $O(\sqrt{\log n / n^{3/2}})$.

## 2.9 Additional Observations

### 2.9.1 Theorem 2.1.9 $\Rightarrow$ Theorem 2.1.7

We show how to derive Theorem 2.1.7 from Theorem 2.1.9 (up to the poly-log factor). For a function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$, let $G_f^-(V, W, E)$ denote the bipartite graph of negatively sensitive edges as before (i.e. $E = S_f^-$). We note that $I_f^- = |S_f^-| / 2^n$ and $\Gamma_{f,\text{matching}}^{-}$ is the maximum size of a matching in $G_f^-$ divided by $2^n$. We intend to show that

$$I_f^- \cdot \Gamma_{f,\text{matching}}^{-} \geq \Omega(f(\varepsilon)^2).$$

Towards this end, for subsets $A \subseteq V, B \subseteq W$, let $(A, B)$ be a minimum vertex cover in $G_f^-$ so that its size $|A| + |B|$ is also the maximum size of a matching. Color the edges of $G_f^-$ with two colors so that all edges incident on $A$ are colored red and all remaining edges (which must be incident on $B$) blue. Applying Theorem 2.1.9, we get that

$$\sum_{x \in A} \sqrt{I_{f,\text{red}}(x)} + \sum_{y \in B} \sqrt{I_{f,\text{blue}}(x)} \geq 2^n \cdot \Omega(\varepsilon(f)).$$

Using Cauchy-Schwarz, the LHS is upper bounded by

$$\sqrt{|A| + |B|} \cdot \sqrt{\sum_{x \in A} I_{f,\text{red}}(x) + \sum_{y \in B} I_{f,\text{blue}}(x)} = \sqrt{\Gamma_{f,\text{matching}}^- \cdot 2^n \cdot |S_f^-|}.$$  

Combining these two observations and squaring both sides of the inequality gives the desired result.

### 2.9.2 Undirected Theorems from Directed Theorems

In this section, we show that Theorems 2.1.1, 2.1.2, 2.1.3, 2.1.8 follow from the corresponding directed versions of these theorems, namely, Theorems 2.1.4, 2.1.5, 2.1.6, 2.1.9 respectively (possibly up to poly-log factor). Thus the directed versions are indeed generalizations of the undirected versions (as ought to be the case).

**Definition 2.9.1.** For input $x \in \{-1, 1\}^n$, let $x^* \in \{-1, 1\}^n$ denote the input with all the bits of $x$ flipped. For a function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$, let $f^* : \{-1, 1\}^n \mapsto \{-1, 1\}$ denote the function defined so that $\forall x \in \{-1, 1\}^n, f(x) = f^*(x^*)$.

**Lemma 2.9.2.** For a function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$,

$$\varepsilon(f) + \varepsilon(f^*) \geq \Omega(\text{var}(f)).$$

**Proof.** Let $h$ be a monotone function that is nearest to $f$ so that $\varepsilon(f) = \Delta(f, h)$. Let $p = \mathbb{E}[f]$ and $q = \mathbb{E}[h]$ so that $|p - q| \leq \varepsilon(f)$. Thus

$$\text{var}(h) = q(1-q) \geq p(1-p) - 2 |p-q| \geq \text{var}(f) - 2 \varepsilon(f).$$


If \( \varepsilon(f) \geq \frac{1}{2} \cdot \text{var}(f) \) we are done. Otherwise, we get \( \text{var}(h) \geq \frac{1}{2} \cdot \text{var}(f) \). Now let \( g \) be a monotone function that is nearest to \( f^* \) so that \( \varepsilon(f*) = \Delta(f^*, g) = \Delta(f, g^*) \). Since \( h \) is monotone and \( g^* \) is anti-monotone, using Lemma 2.3.12,

\[
\varepsilon(f) + \varepsilon(f^*) = \Delta(f, h) + \Delta(f, g^*) \geq \Delta(h, g^*) \geq \text{var}(h) \geq \frac{1}{2} \cdot \text{var}(f),
\]

and we are done.

We make some preliminary observations and then all the implications will follow immediately. Let \( G(V, W, E) \) denote the graph of sensitive edges of \( f \), i.e. \( f(V) \equiv 1 \), \( f(W) \equiv 0 \), \( E \) is precisely the set of sensitive edges of \( f \) and every vertex in \( V \cup W \) has at least one sensitive edge of \( f \) incident on it. By definition:

\[
I_f = \frac{|E|}{2^n}, \quad \Gamma_f = \frac{|V|}{2^n}, \quad \mathbb{E}_x \left[ \sqrt{I_f(x)} \right] = \frac{1}{2^n} \cdot \sum_{x \in V} \sqrt{\deg_G(x)}.
\]

\( G_1(V_1, W_1, E_1) \) be the graph of negatively sensitive edges of \( f \). This is a subgraph of \( G \) induced by precisely the negatively sensitive edges, i.e. \( (x, y) \in E_1 \) if and only if \( (x, y) \in E \) and \( x \leq y \). Also, \( V_1 \subseteq V \), \( W_1 \subseteq W \) are precisely the subsets of vertices that have at least one negatively sensitive edge on them. By definition:

\[
I_f^- = \frac{|E_1|}{2^n}, \quad \Gamma_f^- = \frac{|V_1|}{2^n}, \quad \mathbb{E}_x \left[ \sqrt{I_f^-(x)} \right] = \frac{1}{2^n} \cdot \sum_{x \in V_1} \sqrt{\deg_{G_1}(x)}.
\]

Let \( E_2 = E \setminus E_1 \) and let \( G_2(V_2, W_2, E_2) \) be the subgraph of \( G \) induced by edges in \( E_2 \). Thus \( V_2, W_2 \) are subsets of vertices that have at least one edge in \( E_2 \) incident on them. We may write

\[
G(V, W, E) = G_1(V_1, W_1, E_1) \cup G_2(V_2, W_2, E_2),
\]

noting that \( E_1 \cap E_2 = \emptyset \), but \( V_1, V_2 \) and similarly \( W_1, W_2 \) need not be disjoint.

Consider the mapping \( \varphi : x \mapsto x^* \). This is clearly an isomorphism of the hypercube and let us denote the isomorphic copy of \( G_2(V_2, W_2, E_2) \) under this isomorphism as \( G_2^*(V_2^*, W_2^*, E_2^*) \). The main observation is that \( G_2^*(V_2^*, W_2^*, E_2^*) \) is precisely the graph of negatively sensitive edges of the function \( f^* \). Indeed, an edge \( e = (x, y) \in E_2 \) satisfies \( f(x) = 1 \), \( f(y) = 0 \), \( x \geq y \). So its isomorphic copy \( e^* = (x^*, y^*) \in E_2^* \) satisfies \( f^*(x^*) = 1 \), \( f^*(y^*) = 0 \), \( x^* \leq y^* \) and hence is a negatively sensitive edge of \( f^* \). The same argument works backwards.

The isoperimetric parameters for \( f^* \) concerning the graph \( G_2^*(V_2^*, W_2^*, E_2^*) \) can now be expressed in terms of its isomorphic copy \( G_2(V_2, W_2, E_2) \) as:

\[
I_{f^*} = \frac{|E_2^*|}{2^n}, \quad \Gamma_{f^*} = \frac{|V_2^*|}{2^n}, \quad \mathbb{E}_x \left[ \sqrt{I_{f^*}(x)} \right] = \frac{1}{2^n} \cdot \sum_{x \in V_2} \sqrt{\deg_{G_2}(x)}.
\]

Theorem 2.1.1 now follows from Theorem 2.1.4 and Lemma 2.9.2 as:

\[
I_f = \frac{|E|}{2^n} = \frac{|E_1|}{2^n} + \frac{|E_2|}{2^n} = I_f^- + I_f^+ \geq \Omega(\varepsilon(f)) + \Omega(\varepsilon(f^*)) \geq \Omega(\text{var}(f)).
\]

Theorem 2.1.2 follows from Theorem 2.1.5 as:

\[
I_f \cdot \Gamma_f = (I_f^- + I_f^+) \cdot \frac{|V|}{2^n} \geq (I_f^- + I_f^+) \cdot \frac{1}{2} \cdot \left( \frac{|V_1|}{2^n} + \frac{|V_2|}{2^n} \right) = (I_f^- + I_f^+) \cdot \frac{1}{2} \cdot \left( \Gamma_f^- + \Gamma_f^+ \right),
\]
which is lower bounded, up to the factor \( \frac{1}{2} \), by

\[
I_f \cdot \Gamma_f + I_{f^*} \cdot \Gamma_{f^*} \geq \Omega(\varepsilon(f)^2) + \Omega(\varepsilon(f^*)^2) \geq \Omega((\varepsilon(f) + \varepsilon(f^*))^2) \geq \Omega(\text{var}(f)^2).
\]

Theorem 2.1.3 follows from Theorem 2.1.6 (up to a poly-log factor) as:

\[
\mathbb{E}_x \left[ \sqrt{I_f(x)} \right] = \frac{1}{2^n} \cdot \sum_{x \in V} \sqrt{\deg_G(x)} \geq \frac{1}{2} \cdot \frac{1}{2^n} \cdot \left( \sum_{x \in V_1} \sqrt{\deg_{G_1}(x)} + \sum_{x \in V_2} \sqrt{\deg_{G_2}(x)} \right),
\]

which is same, up to the factor \( \frac{1}{2} \), as

\[
\mathbb{E}_x \left[ \sqrt{I_f(x)} \right] + \mathbb{E}_x \left[ \sqrt{I_{f^*}(x)} \right] \geq \tilde{\Omega}(\varepsilon(f)) + \tilde{\Omega}(\varepsilon(f^*)) \geq \tilde{\Omega}(\text{var}(f)).
\]

Similarly, Theorem 2.1.8 follows from Theorem 2.1.9 (up to a poly-log factor), where a coloring of edges of \( E \) induces a coloring of edges of \( E_1 \) and \( E_2 \). We omit the straightforward proof.
Chapter 3

NP-hardness of 2-to-2 Games

3.1 Introduction

This chapter, together with Chapter 4, are devoted to the proof of the 2-to-2 games conjecture (with imperfect completeness), restated below.

**Theorem 1.2.6 (Restated).** For every constant \( \delta > 0 \) there exists \( k \in \mathbb{N} \) such that given a 2-to-2-Game instance \( G = (V, E, \Phi, \Sigma) \) with alphabet size \( k \), it is NP-hard to distinguish between:

- YES case: there is a coloring satisfying \( 1 - \delta \) fraction of the constraints of \( G \).
- NO case: no coloring satisfies more than \( \delta \) fraction of the constraints of \( G \).

In fact, stronger conditions hold: (a) the YES cases: there is a set \( X \) of \( 1 - \delta \) fraction of the vertices such that all constraints inside \( X \) are satisfied, and (b) the NO case: any set containing \( \delta \) fraction of the vertices contains \( \Omega(\delta^2) \) fraction of the edges of the graph. These conditions are necessary towards certain applications, e.g. to the Vertex Cover and the Independent Set problems.

For further implications, see Chapter 1. The discussion herein focuses on the presentation of techniques and ideas that go into the reduction.

Following the proof of the PCP Theorem, a general framework for proving hardness of approximation results has been developed [AS98, ALM+98, Raz98, BGS98, Hås01]. Using this framework, a PCP construction for a problem \( \mathcal{P} \) is a composition of two separate modules: “Inner PCP” and “Outer PCP”.

- The Inner PCP is a combinatorial gadget – usually code theoretic, such that the corresponding code has a “closeness test” on it only using tests of the form \( \mathcal{P} \). This test should have the following two properties: (a) any legal codeword passes the test with high probability, and (b) any word that passes the test with noticeable probability must be somewhat close to a legal codeword – this usually comes in the form of list-decoding.

- The Outer PCP is usually the 2-Prover 1-Round Problem (see Definition 1.2.1), which is known to be NP-hard (using the PCP Theorem and the Parallel Repetition Theorem).

The composition of the two modules amounts to combining several copies of the Inner PCP via the Outer PCP.

Our reduction for the proof of Theorem 1.2.6 follows this framework, and the first point that should be considered is the choice of the Inner PCP. Since we wish to prove hardness for 2-to-2 Games, we must choose the Inner PCP accordingly and the one that we choose is based on the Grassmann Graph (along with the Grassmann Code and the Grassmann Test). The objects that are being encoded locally are linear functions.
from \(\mathbb{F}_2^k\) to \(\mathbb{F}_2\). Therefore, to make the composition work, our Outer PCP must have linear structure, and a natural starting point of our reduction is the 3Lin problem [Hås01].

**Remark 3.1.1.** To prove hardness results assuming the Unique-Games Conjecture, one only has to design an appropriate Inner PCP for a problem, the Outer PCP is the Unique-Games problem and the composition comes in “for free”. This is in sharp contrast to proving \(NP\)-hardness results, in which case the choice of the Outer PCP and the composition between them presents significant challenges. For instance, towards the end of proving the Unique-Games Conjecture the well known 2Lin Inner PCP of [KKMO07] uses 1-to-1 tests and satisfies the condition above for the Inner PCP, however it is not known how to design an Outer PCP that composes well with it.

### 3.1.1 Informal Overview of the Reduction

#### The Inner PCP

The Inner PCP consists of an encoding scheme, a probabilistic test to check a supposed encoding, and a “list-decoding” procedure when a (supposed) encoding passes the test with “good” probability. In the present context, the Grassmann graph is the central combinatorial object at the Inner PCP level. The Grassmann graph, \(G(V, \ell)\), has a vertex for every \(\ell\)-dimensional subspace \(L\) of an \(k\)-dimensional (global) space \(V\) and there is an edge between two vertices \(L, L'\) if and only if \(\dim(L \cap L') = \ell - 1\). The Grassmann graph leads, in a natural manner, to an encoding of an \(\mathbb{F}_2\)-linear functions \(f : V \to \{0, 1\}\), along with a probabilistic test to check a supposed encoding. The function \(f\) is encoded by writing down \(f|_L\), the restriction of \(f\) to the subspace \(L\), for every \(\ell\)-dimensional subspace \(L\). The restriction \(f|_L\) is a linear function on \(L\) and since there are precisely \(2^\ell\) linear functions on \(L\), the (supposed) restriction \(f|_L\) can be written as a symbol over the alphabet \(\{2^\ell\} = \{1, \ldots, 2^\ell\}\).

A natural 2-query test to check a supposed encoding \(F[\cdot]\) is as follows. Here \(F[\cdot]\) is a table that assigns, to every \(\ell\)-dimensional subspace \(L\), a linear function \(F[L]\) on it. The intention is that \(F[L] = f|_L\) for all \(L\) and for some global linear function \(f : V \to \{0, 1\}\). The test attempts to check whether this is indeed the case. The test chooses a random pair of \(\ell\)-dimensional subspaces \(L, L'\) such that \(\dim(L \cap L') = \ell - 1\), reads \(F[L], F[L']\), and accepts if and only if \(F[L]|_{L \cap L'} = F[L']|_{L \cap L'}\), i.e. if and only if \(F[L], F[L']\) are consistent on the intersection \(L \cap L'\). Observe that

- If \(F[\cdot]\) is, as intended, derived from a global linear function, then the test accepts with probability 1.
- The test is 2-to-2 in the sense that for every linear function \(h\) on \(L \cap L'\), it has precisely two extensions \(f_1, f_2\) to \(L\) and precisely two extensions \(g_1, g_2\) to \(L'\) and then all four pairs \((f_i, g_j)\) \(i,j \leq 2\) qualify as the accepting answers to the queries \(F[L], F[L']\) respectively.

What is the “soundness guarantee” of the test? In other words, what can be said about tables \(F[\cdot]\) that pass the test with noticeable probability? One is tempted to speculate that if the test passes with probability \(\delta\), then the given table \(F[\cdot]\) has a “good” consistency with some global linear function \(f\), in the sense that \(F[L] = f|_L\) for \(C\) fraction of the \(\ell\)-dimensional subspaces \(L\), for some \(C\) depending on \(\delta\). The linear function \(f\) would then serve as a “decoding” of the given table \(F[\cdot]\) and one could even “list-decode”, i.e. make a list of all linear functions \(f\) that have a good consistency with \(F[\cdot]\), along with an upper bound on the list-size that depends (only) on \(\delta\).

This naive speculation, however, turns out to be false. Counter-examples are presented in Section 3.2.1. Nevertheless, any table \(F[\cdot]\) that passes the test with noticeable probability must have a global structure in a more complex sense:

\[\delta\text{ is thought of as a constant, }\ell\text{ as a sufficiently large integer after choosing }\delta,\text{ and the global dimension }\dim(V)\text{ as a sufficiently large integer after choosing }\ell.\]
Consistency Theorem (Informal): There are integers \( q, r \geq 0 \) and constant \( C > 0 \) depending only on \( \delta > 0 \) such that the following holds. Given a table \( F[\cdot] \) that passes the test with probability \( \delta \), there exists a subspace \( Q \) of dimension \( q \) (referred to as “zoom-in”), a subspace \( W \supseteq Q \) of co-dimension \( r \) (referred to as “zoom-out”), and a linear function \( f : W \to \{0, 1\} \) on \( W \), such that \( F[L] = f|_L \) for at least \( C \) fraction of \( \ell \)-dimensional subspaces \( L \) such that \( Q \subseteq L \subseteq W \).

In short, the table \( F[\cdot] \) agrees with a global linear function \( f \) on subspaces \( L \) such that \( Q \subseteq L \subseteq W \) for some “successful” choice of subspaces \( Q, W \) that have constant dimension and co-dimension respectively.

The Consistency Theorem is proved in Chapter 4. In this chapter we prove that the Consistency Theorem implies Theorem 1.2.6.

The theorem, as stated, is not sufficient for our purpose. The reduction, at the Outer PCP level, involves a certain 2-Prover-1-Round Game and the analysis at the Inner and Outer PCP levels needs to be coordinated so as to analyze the overall composed PCP. This necessitates the two provers at the Outer PCP level to “agree” on the choice of zoom-in and zoom-out spaces \( Q \) and \( W \) respectively that are “successful”. It turn out that to agree on the zoom-in space \( Q \), the following strengthening of the Consistency Theorem is required:

Consistency Theorem 2 (Informal): For every \( \delta > 0 \) there are integers \( q, r \geq 0 \) and constant \( C > 0 \), such that the following holds: Given a table \( F[\cdot] \) that passes the test with probability \( \delta \), for \( \alpha(\ell) \) fraction of subspaces \( Q \) of dimension \( q \), there exists a subspace \( W \supseteq Q \) of co-dimension \( r \), and a linear function \( f : W \to \{0, 1\} \) on \( W \), such that \( F[L] = f|_L \) for at least \( C \) fraction of \( \ell \)-dimensional subspaces \( L \) such that \( Q \subseteq L \subseteq W \).

The two theorems are morally equivalent. Roughly speaking, the reason is that for each \( Q \), only a tiny fraction of \( L \)’s contain it, and in particular they do not contribute much to the consistency of \( F \). Thus, if one can “ignore” them, one could sequentially obtain more \( Q \)’s, until their cover at least \( \approx \delta \) of the \( L \)’s.

Let us return to the issue of coordinating between the analysis at the Inner and Outer PCP levels. Since a constant fraction \( \alpha(\ell) \) of the zoom-in spaces \( Q \) are “successful”, the verifier in the 2-Prover-1-Round Game at the Outer PCP level, can simply send \( Q \) as “shared advice” to both the provers and the hypothesis states that the advice is successful with probability \( \alpha(\ell) \).

Coordinate zoom-outs requires more effort. Given the zoom-in space \( Q \), each prover makes a list of all successful zoom-outs \( W_1, \ldots, W_M \) from their own viewpoint, selects one of these zoom-outs at random, and then hopes to agree with the other prover on a common successful zoom-out. For this to work, firstly, the list needs to be “short”, and secondly, there needs to be a zoom-out space \( W \) that is successful for both the provers simultaneously (and hence appears in the lists for both). The latter issue involves delving into the specifics of the PCP composition and moreover, the resolution of the former issue gives a hint towards a resolution of the latter. For these reasons, here we focus only on the former issue, i.e. to upper bound the list size. As happens to be the case, any upper bound \( M = M(\ell) \) will work as long as it is independent of the global dimension \( k \). However, there are cases in which no such upper bound exists. For instance, it is certainly possible that every zoom-out space \( W \) of co-dimension \( r \) is successful. To circumvent this difficulty, we show that if there are too many \((> 2^{2r+2}\ell)\) zoom-out spaces of co-dimension \( r \) that are successful, then there exists a zoom-out space of co-dimension less than \( r \) that is successful! Thus, if we were to make a list of zoom-out spaces of minimal co-dimension that are successful, then it is indeed the case that their list is short (of size at most \( 2^{2r+2}\ell \)).

This completes the high level overview of the Inner PCP. For a more detailed presentation, see in Section 3.2.

---

11 The situation is rather peculiar. If at least a constant fraction \( \alpha(\ell) \) of the zoom-outs were successful, we would be OK, as the zoom-out space \( W \) could then be sent to the provers as advice. If at most a constant number \( M(\ell) \) of zoom-outs were successful, we would again be OK, as the list would be short. Thus we are OK if the list is either too large or too short, with the problematic case being in-between.
The Outer PCP

The Outer PCP is constructed from hard instances of linear constraint satisfaction problems. More specifically, the $\text{Gap3Lin}$ problem: given a set of variables $X$ and a set of equations $Eq$, each equation of the form $x_{i1} + x_{i2} + x_{i3} = b$ over $\mathbb{F}_2$, the goal is to find an assignment $A: X \rightarrow \mathbb{F}_2$ that satisfies as many of the equations as possible. It was shown by Håstad that for every $\varepsilon > 0$, $\text{Gap3Lin}[1 - \varepsilon, s^* = \frac{1}{2} + \varepsilon]$ is NP-hard.

Given a $3\text{Lin}$ instance $(X, Eq)$, we construct the Outer PCP of our reduction (which is then composed with the Inner PCP previously described). It is natural to describe the Outer PCP as an active 2-Prover-1-Round game, in which 2 provers try to convince a verifier that the given $3\text{Lin}$ instance is $(1 - \varepsilon)$-satisfiable. The game depends on a parameter $k$ that should be thought of as large constant (say, larger than doubly exponential in $\ell$) and a smoothness parameter $\beta$ such that $\frac{1}{k} \ll \beta \ll \frac{1}{\sqrt{k}}$; for concreteness $\beta = \frac{\log \log k}{k}$.

We remark that for the soundness analysis we need to incorporate the “advice feature” previously described, however it is not needed for the construction itself so we omit it from this informal overview.

The game proceeds as follows:

1. The verifier chooses $k$ equations $e_1, \ldots, e_k$ at random. Let $U$ be the set of variables that appear in these $k$ equations. The verifier sends $U$ as the question of the first prover.

2. The verifier chooses $V \subseteq U$ as follows: for each $1 \leq i \leq k$, the verifier adds the 3 variables of $e_i$ to $V$ with probability $1 - \beta$, and otherwise adds only one of the variables of $e_i$ chosen uniformly. The verifier sends $V$ as the question of the second prover.

3. The provers are expected to answer with strings $s_U \in \{0, 1\}^U$, $s_V \in \{0, 1\}^V$ indicating an assignment to the variables they received.

4. The prover accepts if and only if the answer of the first prover $s_U$ satisfies the equations $e_1, \ldots, e_k$ and the two provers gave consistent answers on their shared variables, i.e. $s_U|_{U \cap V} = s_V|_{U \cap V}$.

It is easy to see that if the $3\text{Lin}$ instance $(X, Eq)$ is $(1 - \varepsilon)$ satisfiable, then there are strategies for the provers that succeed in the above game with probability $1 - k\varepsilon$. By a careful application of the Parallel Repetition Theorem it can be shown that if the instance is at most $s^*$ satisfiable, then success probability of the provers in the above game is at most $2^{-\Omega(s^*)(\beta \ell)}$.

This completes the high level overview of the Outer PCP. For a more detailed presentation, see in Section 3.3.

Composition of the Outer PCP and Inner PCP

We first view the above game as a constraint satisfaction problem. We only retain the questions $U$ to the first prover, and put a constraint between two questions $U, U'$ if there is a question $V$ to the second prover that is possible for both $U$ and $U'$. The constraint on the edge $(U, U')$ is that the answers $s_U, s_{U'}$ satisfy the equations used to form $U$ and $U'$, and that they are consistent, i.e. $s_U|_{U \cap U'} = s_{U'}|_{U \cap U'}$.

The composition amounts to replacing each question $U$ with the Grassmann Graph $G(\{0, 1\}^U, \ell)$. The answer $s_U$ is naturally identified with the linear function $f_{s_U} : \{0, 1\}^U \rightarrow \{0, 1\}$ given by $f_{s_U}(x) = \langle s_U, x \rangle$; the intention is that the answer $s_U$ will be encoded by the codeword corresponding to $f_{s_U}$ in the Grassmann Code, i.e. the coloring that gives each $\ell$-dimensional subspace the function $f_{s_U}|_L$.

Next, we describe the constraints of the composed PCP. These constraints are supposed to simulate the tests in the Outer PCP using the Grassmann Graph constraints. Recall that the Outer PCP constraints were divided into two types: (1) “side conditions”, i.e. ensuring that $s_V$ is a satisfying assignment to the equations that were used to form $U$, and (2) consistency, i.e. $s_U|_{U \cap U'} = s_{U'}|_{U \cap U'}$ for any $U, U'$ that are adjacent. The first set of constraints is taken care of by employing the folding technique, which we omit from this overview.
For the second set of constraints, we use the edges of the Grassmann Graph “between” blocks. More precisely, looking at $G(\{0,1\}^U, \ell)$, $G(\{0,1\}^V, \ell)$ we connect a vertex $L$ of the first Grassmann Graph and a vertex $L'$ of the second Grassmann Graph by an edge if $\dim(L \cap L') = \ell - 1$. The constraint on that edge is that $F[L]|_{L \cap L'} = F[L']|_{L \cap L'}$.

This completes the overview of the composition and the reduction. We remark that there are more technical details that need to be taken into account that were omitted from the discussion above. The full reduction can is in Section 3.4.

3.2 The Inner PCP: Grassmann Graph and Encoding, 2-to-2 Linearity Test, Zooming In and Out

In this section we describe all the key component of our Inner PCP, the Grassmann graph, and phrase and motivate the Consistency Theorem on it (whose proof is deferred to a later chapter).

Definition 3.2.1 (Grassmann graph). Let $V$ be an $n$-dimensional vector space over a field of size $q$ and $1 \leq \ell \leq n - 1$ be an integer. The Grassmann graph $G(V, \ell)_q$ is defined as follows.

- The vertex set corresponds to all $\ell$-dimensional subspaces $L \subseteq V$.
- The edge set corresponds to all pairs $(L, L')$ of $\ell$-dimensional subspaces such that $\dim(L \cap L') = \ell - 1$.

The following easy fact introduces the important notation of $q$-binomial coefficients (also called Gaussian binomial coefficients).

Fact 3.2.2. Suppose $1 \leq \ell \leq n$.  

1. The number of vertices in the graph $G(V, \ell)_q$ is the $\ell$th $q$-binomial coefficient

$$\left[ \begin{array}{c} n \\ \ell \end{array} \right]_q \overset{\text{def}}{=} \prod_{i=0}^{\ell-1} \frac{q^n - q^i}{q^\ell - q^i}.$$  

2. The graph is regular with degree $d = \frac{q^n - q^{\ell - 1}}{q^{\ell} - q^\ell} \left[ \begin{array}{c} n \\ \ell \end{array} \right]_q$.

In this chapter, we will mostly be concerned with Grassmann graph over $F_2$, i.e. the case $q = 2$, in which we omit the $q$-subscript.

Remark 3.2.3. We will introduce several parameters $\delta, \delta', q, r, C$ that are all treated as constants, the parameter $\ell$ describing the Grassmann graph $G(V, \ell)_q$ is treated as a sufficiently large integer once the earlier parameters have been fixed, and the parameter $n = \dim(V)$ is treated as a sufficiently large integer once $\ell$ is fixed.

The Grassmann Encoding The Grassmann Graph can be used to construct an encoding scheme for linear functions. A linear function $f: V \to \{0,1\}$, is encoded by an assignment $F$ to the vertices of the graph, that assigns $f|_L$ to the subspace $L$ — that is, the restriction of $f$ to $L$. Since $f|_L$ is a linear function on $L$ and there are precisely $2^\ell$ linear functions on $L$, one can consider $f|_L$ as a symbol from the alphabet $[2^\ell] = \{1, \ldots, 2^\ell\}$. Thus clearly, the alphabet size of the Grassmann Encoding is $2^\ell$, and the word length is $\left[ \begin{array}{c} n \\ \ell \end{array} \right]$ — the number of $\ell$-dimensional subspaces of a given $n$-dimensional vector space over $F_2$. The relative distance of the...
encoding scheme $\approx 1 - 2^{-\ell}$: any two distinct linear functions $f, g: V \to \{0, 1\}$ agree on $\approx \frac{1}{2}$ of the vectors in $V$, and so they agree on an $\ell$-dimensional subspace with probability $\approx 2^{-\ell}$.

The edges of the Grassmann graph, naturally corresponds to local consistency checks, attempting to test whether a supposed encoding is indeed a valid codeword. The input to the test is a table $F[:]$ that assigns a linear function $F[L]$ on $L$ for every $\ell$-dimensional subspace $L \subseteq V$. The test attempts to check whether $F[:]$ is a valid encoding of some (global) linear function $f : V \to \{0, 1\}$.

**The Grassmann Linearity Test**

- Pick an edge $(L, L')$ of the Grassmann Graph at random and read $F[L], F[L']$.
- Accept if and only if $F[L]|_{L \cap L'} = F[L']|_{L \cap L'}$, i.e. if and only if $F[L], F[L']$ are consistent on the $(\ell - 1)$-dimensional intersection $L \cap L'$.

The test is 2-to-2 in the following sense. Fix two adjacent vertices $L, L'$. Any linear function $h : L \cap L' \to \{0, 1\}$ has two extensions to $L$ and two extensions to $L'$, since these subspaces contain $L \cap L'$ and have only one extra dimension. Thus, if $L$ is assigned one of $h$’s (two) extensions to it, $L'$ must be assigned one of $h$’s extensions to it as well.

The test has “perfect completeness”, meaning that if $F[:]$ is a legal encoding of some (global) linear function $f : V \to \{0, 1\}, F[L] = f|_L$ for all $L$, then the test accepts with probability 1. A natural inverse question arises from the “soundness” point of view: let $F[:]$ is a table that passes the Grassmann Test with constant probability $\delta > 0$ (we call such table $\delta$-consistent), is it necessarily the case that $F[:]$ “corresponds” to some global linear function? This question is quite vague at the moment, since one could interpret “corresponds” in more than one way.

**Definition 3.2.4.** Let $F : G(V, \ell) \to [2^\ell]$ assign, to each $\ell$-dimensional subspace $L \in G(V, \ell)$, a linear function $F[L]$. The table $F[:]$ is said to be $\delta$-edge-consistent (or simply $\delta$-consistent), if

$$\Pr_{(L, L') \in G(V, \ell)} [F[L]|_{L \cap L'} = F[L']|_{L \cap L'}] \geq \delta.$$ 

Let us consider the most simplistic interpretation of the inverse question: must a $\delta$-consistent assignment have non-negligible agreement with a global linear function? At first glance, one expects the answer to be positive, especially if one draws an analogy with the “line versus point” and “plane versus plane” low degree tests studied in literature [RS96, AS03, RS97]. Stated more precisely:

**Speculation 3.2.5.** For every constant $\delta > 0$, there exists a constant $\delta' > 0$, such that the following holds. Given a table $F[:]$ that is $\delta$-consistent, there exists a global linear function $f : V \to \{0, 1\}$, such that $F[L] \equiv f|_L$ for $\delta'$-fraction of the vertices in the Grassmann Graph.

It turns out that this speculation is false, as shown by the examples below.

**3.2.1 Subspace Example and Zooming-in**

**Subspace Example**

Here is a counter-example to Speculation 3.2.5. Let $Z \subseteq V$ be a fixed subspace of dimension $n - \ell$ and consider the set $S$ of vertices of the Grassmann graph,

$$S = \{L \mid L \subseteq V, \dim(L) = \ell, \dim(L \cap Z) = 1\}.$$ 

It is not difficult to see that $S$ constitutes a constant fraction, say $\beta$, of vertices of the Grassmann graph. We associate, with every $z \in Z, z \neq 0$, a global linear function $f_z : V \to \{0, 1\}$ chosen randomly. A table $F[:]$ is defined as follows:
If $L \in \mathcal{S}$, then $F[L] = f_z|_L$, where $z \in Z, z \neq 0$ is the unique point such that $L \cap Z = \text{Span}(z)$.

If $L \not\in \mathcal{S}$, then $F[L]$ is defined at random.

We observe that the assignment $F[\cdot]$ is consistent on a constant fraction of edges incident to $\mathcal{S}$. Indeed, fix any $L \in \mathcal{S}$, $L \cap Z = \text{Span}(z)$, and consider a random edge $(L, L')$ incident to $L$. The subspace $L'$ can be sampled by first choosing a random $(\ell - 1)$-dimensional subspace $R \subseteq L$, and then extending it with a random vector $w \not\in L$ to $L' = R \oplus \text{Span}(w)$. With probability $\approx \frac{1}{2}$ over the choice of $R$, we have $z \in R$, and if $w \not\in L \oplus Z$. In this case, $L' = R \oplus \text{Span}(w)$ satisfies $L' \cap Z = \text{Span}(z)$ and hence $L' \in \mathcal{S}$. The assignment $F[L']$ then equals $f_z|_{L'}$, while the assignment $F[L]$ equals $f_z|_{L}$ and the two assignments are consistent. To summarize, $\approx \frac{1}{2}$ fraction of the edges incident to $\mathcal{S}$ are satisfied by the assignment $F[\cdot]$, which constitute $\approx \frac{\beta^2}{4}$ fraction of all edges of the Grassmann graph. This $F[\cdot]$ is consistent on $z$ fraction of all edges incident to $\mathcal{S}$. Iterating this, one is able to assign $\Theta(1)$ fraction of the vertices, have $\Theta(1)$ consistency and no global structure.

The notion of “zoom-in” attempts to circumvent the counter-example above. Since we will propose another idea called a “zoom-out”, let us introduce a notation that handles both. For subspaces $Q \subseteq W \subseteq V$, thought of as having a constant dimension and co-dimension respectively inside the global space $V$, let

$$\text{Zoom}[Q, W] = \{L \mid L \subseteq V, \dim(L) = \ell, Q \subseteq L \subseteq W\}.$$  

In words, this is the set of $\ell$-dimensional subspaces that contain $Q$ and are contained in $W$. The set of all $\ell$-dimensional subspaces containing $Q$ is then $\text{Zoom}[Q, V]$ whereas the set of all $\ell$-dimensional subspaces, i.e. the vertex set of the Grassmann graph, is $\text{Zoom}[\{0\}, V]$. Before we continue, let take note of the following easy, nevertheless important, observation: the induced subgraph of on $\text{Zoom}[Q, W]$ is isomorphic to a lower order Grassmann graph $G(V', \ell')$ where $V' = W/Q$ is the quotient space, $\dim(V') = \dim(W) - \dim(Q)$, and $\ell' = \ell - \dim(Q)$.

In the counter-example above, $F[\cdot]$ is a table that is $\approx \Theta(1)$-consistent, but there is no global linear function that is $\Omega(1)$-consistent with $F[\cdot]$. However, suppose we fix some $z^* \in Z, z^* \neq 0$ and focus on the set $\text{Zoom}[\text{Span}(z^*), V]$, i.e. the set of $\ell$-dimensional subspaces containing $z^*$. We recall that $\mathcal{S}$ is the set of $\ell$-dimensional subspaces, of density $\beta$, that have one-dimensional intersection with $Z$.

- It is not difficult to see that $\mathcal{S}' = \mathcal{S} \cap \text{Zoom}[\text{Span}(z^*), V]$ has constant density inside the “focus set” $\text{Zoom}[\text{Span}(z^*), V]$. In fact, its density $\beta'$ is slightly larger than $\beta$, the density of $\mathcal{S}$ inside $\text{Zoom}[\{0\}, V]$.

\[\text{We used here the fact that if a set of vertices in the Grassmann Graph has density } \beta, \text{then it contains at least } \beta^2 \text{ fraction of the edges of the graph.}\]
• The assignment $F[\cdot]$ to vertices in $S'$ is precisely the restriction of the global linear function $f_{z'}$. I.e. after zooming into the one-dimensional subspace $\text{Span}(z^*)$, the assignment $F[\cdot]$ does have a good (= $\beta'$) consistency with some global linear function. The zoom-in space $\text{Span}(z^*)$ is said to be “successful” in this sense.

• The fraction of one-dimensional spaces $\text{Span}(z^*), z^* \in Z, z^* \neq 0$ is $\approx 2^{-\ell}$ fraction of all the one-dimensional subspaces of $V$.

These observations lead to the speculation below, stating that any $\delta$-consistent assignment $F[\cdot]$ must have a good consistency with a global linear function after zooming into some constant dimensional subspace. Moreover, there is a “reasonable” fraction (that may depend on $\ell$) of the zoom-ins are successful (the global linear function may depend on the choice of the zoom-in space).

**Speculation 3.2.6.** For every constant $\delta > 0$, there is an integer $q \geq 0$, a constant $\delta' > 0$, and a function $\alpha(\cdot) > 0$ of an integer parameter, such that the following holds. Given a table $F$ that is $\delta$-consistent, for $\alpha(\ell)$ fraction of $q$-dimensional subspaces $Q$, there exists a global linear function $f_Q : V \to \{0, 1\}$, such that $F[L] = f_Q|L$ for $\delta'$ fraction of the vertices in $\text{Zoom}[Q, V]$.

### 3.2.2 Hyperplane Example and Zooming-Out

It turns out that Speculation 3.2.6 is false as well, as shown by the example below.

**Hyperplane Example**

We assume that the dimension of the global space $V$ is $\gg 2^\ell$. Let $m = \gamma 2^\ell$ for a small constant $\gamma \left( \gamma = \frac{1}{20} \right)$ works. Let $W_1, \ldots, W_m$ be hyperplanes (= subspaces of co-dimension one) in space $V$ that are in general position, meaning the intersection of any $k$ of them, for $1 \leq k \leq m$, has co-dimension $k$. Let $f_i : W_i \to \{0, 1\}$ be linear functions on these hyperplanes such that for $1 \leq i \neq j \leq m$, $f_i$ and $f_j$ are different on $W_i \cap W_j$.

Let $S$ be the set of $\ell$-dimensional subspaces $L$ that are contained in precisely one of the $W_i, 1 \leq i \leq m$. It will be convenient to write $S = \bigcup_{i=1}^m S_i$ where $S_i$ is the set of subspaces $L$ that are contained in $W_i$ and not contained in any other $W_j, j \neq i$. We claim that $S$ has density at least $\frac{\gamma}{2}$. Indeed, choosing $L$ at random, the probability that $L \subseteq W_i$ for some fixed $i$ equals (essentially) $2^{-\ell}$ and the probability that $L \subseteq W_i \cap W_j$ for some fixed $i \neq j$ is (essentially) $2^{-2\ell}$. By inclusion-exclusion principle, the probability that $L$ is contained in precisely one of the $W_i, 1 \leq i \leq m$, is at least

$$m \cdot 2^{-\ell} - \binom{m}{2} \cdot 2^{-2\ell} = \gamma - \frac{\gamma^2}{2} \geq \frac{\gamma}{2}.$$

Let $E_i$ denote the set of edges inside $S_i$ and $E = \bigcup_{i=1}^m E_i$. We claim that $E$ constitutes at least $\frac{\gamma}{16}$ fraction of all the edges. This can be seen as follows. An argument as above shows that the probability that a random $(\ell + 1)$-dimensional subspace $R$ is contained in precisely one of the $W_i, 1 \leq i \leq m$, is at least

$$m \cdot 2^{-(\ell+1)} - \binom{m}{2} \cdot 2^{-2(\ell+1)} = \gamma - \frac{\gamma^2}{8} \geq \frac{\gamma}{4}.$$

A random edge $(L, L')$ can be chosen by first choosing a random $(\ell + 1)$-dimensional subspace $R$ and letting $L, L'$ to be distinct $\ell$-dimensional subspaces of $R$. As noted, with probability at least $\frac{\gamma}{16}$, $R$ is contained in precisely one $W_i, 1 \leq i \leq m$, say $W_{i_0}$. Now consider the $2^{\ell+1} - 1$ distinct $\ell$-dimensional subspaces of $R$. Each $W_j, 1 \leq j \neq i_0 \leq m$, can contain at most one of these subspaces, because containing two of them
would mean containing $R$. This implies that at least half of the \(\ell\)-dimensional subspaces of $R$ are contained in $W_0$, but not contained in any other $W_j$. Thus for a random pair of distinct $\ell$-dimensional subspaces $L, L'$ of $R$, with probability \(\frac{1}{4}\), both $L, L' \in S_0$, and hence the edge $(L, L') \in E_0$. This proves the claim.

We are now ready to define the table $F[\cdot]$. For $L \in S = \bigcup_{i=1}^{m} S_i$, if $L \in S_0$, let $F[L] = f_{i_0}|L$. For $L \notin S$, $F[L]$ is defined randomly. Clearly, $F[\cdot]$ satisfies all edges in $E$ and hence is $\frac{1}{16}$-consistent. However, we show that for constant $q$, there is no \(q\)-dimensional subspace $Q$ for which $F[\cdot]$ is consistent with some global linear function on \(\Omega(1)\) fraction of $\ell$-dimensional subspaces containing $Q$, exhibiting a counter-example to Speculation 3.2.6.

Fix any $q$-dimensional subspace $Q$ and let $\text{Zoom}(Q, V)$ be the set of all $\ell$-dimensional subspaces containing $Q$. These will be the only subspaces under consideration henceforth. Let $f : V \to \{0, 1\}$ be any global linear function. We look at the consistency between $f$ and $F[\cdot]$ on the subspaces in $\text{Zoom}(Q, V)$. Let $L_i$ be the set of those $L \in \text{Zoom}(Q, V)$ that are contained in $W_i$. Clearly, if $Q \notin W_i$, then $L_i = \emptyset$ and otherwise, $L_i$ has some fixed size depending on $q, \ell, \text{dim}(V)$. Let $T_i \subseteq L_i$ be the set of those $L \in \text{Zoom}(Q, V)$ that are contained in $W_i$ but not in any other $W_j$, $j \neq i$.

Firstly, since $F[\cdot]$ is defined randomly outside $\bigcup_{i=1}^{m} T_i$, non-trivial consistency between $f$ and $F[\cdot]$, if any, has to be on $\bigcup_{i=1}^{m} T_i$. Secondly, since $F[\cdot]$ is defined according to distinct functions $f_i, f_j$ on $T_i, T_j$ respectively, the consistency between $f$ and $F[\cdot]$ is non-negligible on at most one $T_i$. Suppose that $Q$ is contained in exactly $k$ of the $W_i$‘s, say $W_1, \ldots, W_k$. We consider two cases depending on how large $k$ is, that show the consistency between $f$ and $F[\cdot]$ is at most $2^{-\frac{\ell}{2}}$ in both the cases, exhibiting the counter-example.

- (Case when $k \leq 2^\frac{\ell}{2}$). We note that $L_{k+1} = \cdots = L_k$ is the density of $\bigcup_{i=1}^{k} L_i$ inside $\text{Zoom}(Q, V)$ is at most $k \cdot 2^{\ell - \ell} \leq 2^{\ell - \frac{\ell}{2}}$. Since the consistency between $f$ and $F[\cdot]$ has to be on $\bigcup_{i=1}^{k} T_i \subseteq \bigcup_{i=1}^{k} L_i$, the consistency is upper bounded by $2^{\frac{\ell}{2}}$.

- (Case when $k \geq 2^\frac{\ell}{2}$). Since by symmetry, all $T_i$, $i = 1, \ldots, k$, have the same size and are pairwise disjoint, and the consistency is non-negligible on at most one $T_i$, the consistency is upper bounded by $\frac{1}{k} \leq 2^{-\frac{\ell}{2}}$.

**Zooming-out**

We propose to circumvent the counter-example above using an idea of a “zoom-out” wherein one focuses on $\ell$-dimensional subspaces that are contained in a subspace of constant co-dimension. Indeed, in the example above, we can choose any of the hyperplanes $W_i$ and focus on $\text{Zoom}([\{0\}, W_i])$, the subset of those $L$ that are contained in $W_i$. By definition, there does exist a global linear function, namely $f_i$, that is consistent with $F[\cdot]$ on $S_i \subseteq \text{Zoom}([\{0\}, W_i])$. As noted, $S_i$ has density at least $\frac{\ell}{2}$ inside $\text{Zoom}([\{0\}, W_i])$, and hence $f_i$ is $\frac{\ell}{2}$-consistent with $F[\cdot]$ on $\text{Zoom}([\{0\}, W_i])$. We observe in addition that only the hyperplanes $W_1, \ldots, W_m$ (and further subspaces of them of constant co-dimension if one wishes) furnish a “successful” zoom-out.

It is certainly possible to combine the counter-examples in Section 3.2.1 and Section 3.2.2 so that both the zoom-in and the zoom-out are needed to circumvent the combined example. Our final speculation, which is actually a theorem stated next, proposes that there always exist a zoom-in of constant dimension, and a zoom-out on constant co-dimension respectively that are successful.

**Theorem 3.2.7.** For every constant $\delta > 0$, there exist integers $r, q \geq 0$, a constant $C > 0$, such that for all sufficiently large integers $n$, the following holds. Let $F[\cdot]$ be an assignment to the Grassmann graph $G(V, \ell)$, $\text{dim}(V) = n$, satisfying at least $\delta$ fraction of the edges.
Then there exist a $q$-dimensional subspaces $Q \subseteq V$, a subspace $W, Q \subseteq W \subseteq V$ of co-dimension $r$, and a global linear function $g_{Q,W} : W \rightarrow \{0,1\}$ such that (note the conditional probability)

$$\Pr_{L \in G(V,\ell)} [g_{Q,W}|L \equiv F[L] \mid Q \subseteq L \subseteq W] \geq C.$$  

(3.1)

The vigilant reader may notice that we have dropped the requirement that there is a “reasonable” fraction (that may depend on $\ell$) of zoom-ins $Q$, for which there exists a zoom-out that is successful. It turns out that one can get this additional property from Theorem 3.2.7 by a simple argument, as described below.

**Theorem 3.2.8.** For every constant $\delta > 0$, there exist integers $r, q > 0$, a constant $C > 0$, and a function $\alpha(\cdot) > 0$ of an integer parameter such that for all sufficiently large integers $n$, satisfying at least $\delta$ fraction of the edges. Then for at least $\alpha(\ell)$ fraction of the $q$-dimensional subspaces $Q \subseteq V$, there exists a subspace $W, Q \subseteq W \subseteq V$ of co-dimension $r$, and a global linear function $g_{Q,W} : W \rightarrow \{0,1\}$ such that (note the conditional probability)

$$\Pr_{L \in G(V,\ell)} [g_{Q,W}|L \equiv F[L] \mid Q \subseteq L \subseteq W] \geq C.$$  

(3.2)

**Proof.** The proof strategy is as follows: apply the Theorem 3.2.7 sequentially, finding a single $Q$ at a time. Each time a $Q$ is found, we “randomize” the assignment on $L \cap Q$ and continue. The two main point are (a) each newly found $Q$ only contains a small fraction of the $L$’s, hence as long as we have not found enough $Q$ our assignment is still highly consistent, and (b) the re-randomization phase does not contribute much to the $Q$’s that are found in subsequent steps.

Let $r, q, C$ be from Theorem 3.2.7 for $\delta/2$, $\tilde{F} = F$. We prove the Theorem for parameters $r, q, C/2$ and $\alpha(\ell) = 2^{-\ell^2 - 2}$.

Denote by $Q$ the set of $Q$ found so far, $N = |Q|$, and by $X$ the set of $L$’s whose assignment was re-randomized. At each step we apply Theorem 3.2.7 to obtain $Q$ such that there exists $W$ and $g_{Q,W}$ such that Equation (3.13) holds. We then define

$$X \leftarrow X \cup \{L \mid Q \subseteq L, \dim(L) = \ell\},$$

and reassign the spaces of $X$ on $\tilde{F}$ in a manner to be described later.

**Claim 3.2.9.** At the end of the process, $N \geq \frac{n}{q\ell} - 2^{\ell^2 - qn}$.

**Proof.** Note each $Q$ causes at most

$$\frac{n^\ell}{n^{\ell-q}} \leq 2^{n(\ell-q)} = 2^{\ell^2 - qn}$$

fraction of the $L$ spaces to be reassigned. Therefore overall during the process, at most $N \cdot 2^{\ell^2 - qn}$ fraction of $L$’s were reassigned. Hence at most $2N \cdot 2^{\ell^2 - qn}$ fraction of the edges in $G(V,\ell)$, at least one of their endpoints was reassigned. When the process is stuck, $\tilde{F}$ is at most $\delta/2$ consistent, and so it must be the case that

$$2N \cdot 2^{\ell^2 - qn} \geq \delta/2$$

$$\Rightarrow N \geq \delta 2^{n\ell^2 - 2} \geq \delta 2^{\ell^2 - 2} \left[\frac{n}{q}\right].$$

\[\square\]
Next we argue that each newly found $Q$, has $W, g_{Q, W}$ for which Equation (3.13) holds, albeit with $C/2$, for $F$. This is achieved by claiming that at most $21^{-\ell}$ of the consistency with $\tilde{F}$ comes from spaces that were reassigned. The latter is shown by considering two cases: if at most $21^{-\ell}$ fraction of $Q \subseteq L \subseteq W$ are in $X$, it is obvious. Otherwise

**Claim 3.2.10.** With probability $1 - o(1)$, for every $Q$ of dimension $q$, $W$ of dimension $r$ such that at least $2^{-\ell}$ fraction of $\{ L \mid Q \subseteq L \subseteq W \}$ is in $X$, for every linear function $g_{Q, W} : W \rightarrow \{0, 1\}$,

$$\Pr_{L \in R^X, Q \subseteq L \subseteq W} \left[ g_{Q, W} \mid L \equiv \tilde{F}[L] \right] \leq 21^{-\ell}$$

**Proof.** Fix such $Q, W, g_{Q, W}$, and denote $A = \{ L \mid Q \subseteq L \subseteq W \} \cap X$. For each $L \in A$ define the indicator random variable $Z_L$ which is $1$ iff $g_{Q, W} \mid L \equiv \tilde{F}[L]$. Then its expectation is $2^{-\ell}$, and, using Chernoff bound, the required probability is bounded by

$$\Pr_{L \in R^A} \left[ \frac{1}{|A|} \sum_{L \in A} Z_L \geq 21^{-\ell} \right] \leq \Pr_{L \in R^A} \left[ \frac{1}{|A|} \sum_{L \in A} Z_L - 2^{-\ell} \right] \leq 2^{-\frac{1}{2}2^{-2\ell}|A|}. $$

Note that $|A| \geq 2^{-\ell} |\{ L \mid Q \subseteq L \subseteq W \}| \geq 2^{-\ell} 2^{(\ell-q)q}2^{(n-r-\ell)}$. Therefore, using a union bound over $Q, W, g_{Q, W}$, the probability there exists such bad triplet is at most

$$2^{qn}2^{r(n-r)}2^{2^{-\ell}2^{(\ell-q)q}2^{(n-r-\ell)}} = o(1).$$

In particular, the previous claim implies there exist a reassignment (and we will pick such one in the reassignment phase) of vertices on $X$, such that on each newly found $Q, W$, at most $21^{-\ell}$ of the agreement with $\tilde{F}$ comes from $X$. Hence at least $C - 21^{-\ell} \geq C/2$ of the agreement comes outside $X$, i.e. Equation (3.13) holds with $C/2$. $\square$

### 3.3 The Outer PCP Game

Our Outer PCP is a carefully constructed 2-Prover-1-Round Game from a regular instance of the 3-Lin problem. Recall that an instance $(X, Eq)$ of the 3-Lin problem consists of a set of $\mathbb{F}_2$-valued variables $X$ and a set of equations $Eq$, where each equation containing three (distinct) variables. The instance is regular if every variable appears in at most 5 equations, and two distinct equations share at most one variable. Starting with a 3-Lin instance given by Håstad’s reduction [Hås01], a standard sequence of transformations can turn the instance into a regular one, while preserving the near-perfect completeness and keeping the soundness bounded away from $1$.\(^{14}\) To summarize:

**Theorem 3.3.1.** There exists an absolute constant $\frac{1}{2} < s^* < 1$ such that for every constant $\varepsilon > 0$, the Gap$3\text{Lin}(1 - \varepsilon, s^*)$ problem on regular instances is NP-hard.

Let $(X, Eq)$ be an instance of Gap$3\text{Lin}(1 - \varepsilon, s^*)$ as in Theorem 3.3.1. We intend to construct a 2-Prover-1-Round Game that is used as our Outer PCP. Instead of taking a passive view of 2-Prover-1-Round Game as a constraint satisfaction problem as in Definition 1.2.1, it is more intuitive to take an equivalent active view in terms of two provers and a probabilistic verifier. The two provers wish to convince the verifier that the 3-Lin instance is near-satisfiable. Since our construction has multiple subtle features, we present it incrementally, adding one feature at a time.

\(^{14}\)Starting with Hastad’s instances [Hås01], one reduces the number of appearances of each variable to constant using the expander technique of [PY91] (also in [Aro94, Theorem 6.2]). Similarly to Feige [Fei98], one can further reduce this constant to 5. To get the condition on distinct equations, replace each equation $e : x + y + z = b$ the equations $x + y_e + z_e = b$, $x + y + z_e = b$, $x_e + y + z = b$, where $x, y, z$, $x_e, y_e, z_e$ are new variables.
3.3.1 Equation vs Variable Game

We start with a standard “equation vs variable” game. In this game, the verifier chooses an equation \( e \in \text{Eq} \) uniformly at random, sends it to the first prover, chooses a variable \( x \) randomly from the three variables occurring in the equation \( e \) and sends it to the second prover. The provers are expected to provide an \( \mathbb{F}_2 \)-value for each of the variables they receive. The verifier accepts if and only if the first prover provides a satisfying assignment to \( e \) and if both provers give \( x \) the same value.

**Completeness:** Suppose there is an assignment to \( (X, \text{Eq}) \) that satisfies \( 1 - \varepsilon \) fraction of the equations. The provers can answer according to this assignment and the verifier accepts with probability at least \( 1 - \varepsilon \).

**Soundness:** Suppose no assignment to \( (X, \text{Eq}) \) satisfies more than \( s^* \) fraction of the equations. The strategy of the second prover is simply an assignment to all the variables. This assignment fails to satisfy \( 1 - s^* \) fraction of the equations. For every equation that fails, the second prover either has to give inconsistent answer to at least one of its variables or answer with an unsatisfying assignment to the equation. Thus the provers cannot make the verifier accept with probability more than \( 1 - \frac{1-s^*}{3} \) (i.e. bounded away from 1).

3.3.2 Smooth Equation vs Variable Game

We modify the equation vs variable game slightly and call it a smooth game.\(^{15}\) Let \( \beta \in (0, 1) \) be a smoothness parameter. The verifier sends an equation \( e \) to the first prover as before. To the second prover however, the verifier sends a random variable \( x \) occurring in \( e \) with probability \( \beta \), and sends the equation \( e \) with probability \( 1 - \beta \) (hence asking the same question to both provers).

**Completeness:** As before, the completeness is at least \( 1 - \varepsilon \).

**Soundness:** The new game is effectively a trivial game with probability \( 1 - \beta \) and is same as the equation vs variable game with probability \( \beta \). Hence the soundness is at most \( 1 - \Omega(\beta) \), where the \( \Omega \)-notation hides the dependence on \( s^* \) (which is an absolute constant anyways).

3.3.3 Smooth Equation vs Variable Game with Advice

We require an additional feature from our game that we call “advice”. In a sense, this “advice” acts like publicly shared randomness, which depends on the questions of the provers. Nevertheless, this advice cannot considerably help the provers.

As before, the verifier picks an equation \( e \) at random, say \( x_{i_1} + x_{i_2} + x_{i_3} = b_i \), and sends it to the first prover. With probability \( 1 - \beta \), the second prover receives the equation \( e \) as well, and otherwise a single variable from the equation \( e \) chosen at random. Let \( V \subseteq \{x_{i_1}, x_{i_2}, x_{i_3}\} \) be the set of variables sent to the second prover (so \( |V| \) is 1 or 3). The verifier chooses an advice vector \( a \in \{0, 1\}^V \) at random. If \( |V| = 3 \), define \( a^* = a \), and if \( |V| = 1 \), let \( a^* \) be obtained from \( a \) by padding with 0 in place of \( \{x_{i_1}, x_{i_2}, x_{i_3}\} \setminus V \). The verifier sends the first prover the vector \( a^* \) and the second prover the vector \( a \). As before, the provers are expected to provide a value for each of the variable they receive.

Call this game \( G_{\beta,1} \). The extra advice could give the first prover a hint as to which variables the second prover receives. For example, if the first prover’s advice vector is \( a^* = (0, 0, 1) \), she knows that the second prover has received either all three variables or (just) the variable \( x_{i_3} \). However, when the first prover receives the vector \( (0, 0, 0) \), she does not know whether the second prover has received all three variables along with advice \( a = (0, 0, 0) \) or a single variable, whose identity she does not know, along with advice \( a = (0) \). It is clear from this discussion that:

\(^{15}\)Smoothness refers to the property of a game wherein for a fixed question and two distinct answers to the first prover, w.h.p. over the choice of the question to the second prover, the second prover’s answers must be distinct for the verifier to accept. The game described is smooth provided \( \beta \ll 1 \).
Completeness: The completeness of game $G_{\beta,1}$ is at least $1 - \varepsilon$.

Soundness: The soundness of game $G_{\beta,1}$ is at most $1 - \Omega(\beta)$.

We further generalize to the game $G_{\beta,q}$ for any integer $q \geq 0$ where instead of sampling and sending the provers one pair $(a^*, a)$ respectively, the verifier samples independently, $q$ pairs $(a_1^*, a_1), \ldots, (a_q^*, a_q)$, and sends the list $[a_1^*, \ldots, a_q^*]$ to the first prover and the list $[a_1, \ldots, a_q]$ to the second prover. It is not difficult to see that:

Completeness: The completeness of game $G_{\beta,q}$ is at least $1 - \varepsilon$.

Soundness: The soundness of game $G_{\beta,q}$ is at most $1 - \Omega(2^q \beta)$.

Intuitively, the verifier rejects when the second prover is sent a single variable (which happens with probability $1/k$) along with the advice-list $[(0), \ldots, (0)]$ (which happens with probability $1/2^q$).

Remark 3.3.2. The soundness of the $(2^q \beta)$-fold parallel repetition game $G_{\beta,q}^{2^q/\beta}$ is upper bounded by an absolute constant less than 1. Intuitively, in $2^q$ “trials”, with constant probability, there is a “coordinate” on which the second prover receives a single variable along with the advice-list $[(0), \ldots, (0)]$, and then the verifier rejects with a constant probability.

3.3.4 The Final Game (Outer PCP)

Finally, our Outer PCP is a $k$-fold parallel repetition of the game $G_{\beta,q}$, i.e. the game $G_{\beta,q}^k$.

Completeness: The completeness of game $G_{\beta,q}^k$ is at least $1 - k\varepsilon$.

Soundness: The soundness of game $G_{\beta,q}^k$ is at most $2^{-\Omega(\beta k/2^q)}$. The game can be considered as $\frac{\beta k}{2^q}$-fold parallel repetition of the game $G_{\beta,q}^{2^q/\beta}$ which has constant soundness as per Remark 3.3.2. One can then apply the parallel repetition theorem for projection games with no dependency on the answer size as in [Rao11, DS14].

Remark 3.3.3. Let $U, V$ be the questions sent to the first and the second prover in the game $G_{\beta,q}^k$ not taking into account the “advice” yet. Thus $U$ is a set of $3k$ variables and $V \subseteq U$ with expected size $E[|V|] = 3k - 2\beta k$. With a careful look, it can be seen that the advice-list for the first prover is a list $[x_1, \ldots, x_q]$ with $\forall 1 \leq i \leq q, x_i \in \{0,1\}^U$. Similarly, the advice-list for the second prover is a list $[y_1, \ldots, y_q]$ with $\forall 1 \leq i \leq q, y_i \in \{0,1\}^V$. Moreover, if one regards the space $\{0,1\}^V$ as a subspace of $\{0,1\}^U$ in a natural manner, then $\forall 1 \leq i \leq q, x_i = y_i$. Thus the advice is to be interpreted as a list of $q$ points in $\{0,1\}^V$ that is sent to both provers.

3.4 The Reduction to 2-to-2 Games

In this section we present our reduction towards proving Theorem 1.2.6. The soundness analysis of the reduction is presented in Section 3.6.1.

3.4.1 Construction of the Transitive Game $G_{2:2}$

Given an instance $(X, E_q)$ of regular 3Lin, we construct an instance of 2-to-2 games. Towards this end, we consider the Outer PCP Game described in Section 3.3 with parameters $k, \beta, q$ and ignore the advice feature therein.

\[ \]
Let $\mathcal{U}$ denote the set of questions asked to the first prover. Specifically, $\mathcal{U}$ is the set of all $k$-tuples of equations $U = (e_1, \ldots, e_k)$ from the regular Gap3Lin instance $(X, \text{eq})$. For our purposes, it will be convenient to retain only those “legitimate” $U = (e_1, \ldots, e_k)$ such that (a) the equations $e_1, \ldots, e_k$ are distinct and do not share variables and (b) for any pair of variables $x \in e_i$ and $y \in e_j$, $i \neq j$, $x, y$ do not appear together in any equation in the instance $(X, \text{eq})$. Due to regularity of the instance $(X, \text{eq})$, every variable appears in a constant number of equations and hence the fraction of $U$ that are not legitimate is negligible, i.e. $O(k^2|X|^\epsilon)$, and dropping these does not affect our analysis. We assume henceforth that $\mathcal{U}$ contains only the legitimate tuples $U$.

The Outer PCP verifier picks a $k$-tuple $U = (e_1, \ldots, e_k) \in \mathcal{U}$ uniformly at random and then constructs a $k$-tuple $V$ such that independently for $1 \leq i \leq k$, the $i^{th}$ element of $V$ is the equation $e_i$ with probability $1 - \beta$ and is a variable in the equation $e_i$ with probability $\beta$. In the following construction, the questions $U$ to the first prover appear explicitly whereas the role of the questions $V$ to the second prover is only implicit.

We now describe the Transitive 2-to-2 Game $G_{2:2} = (V(G_{2:2}), E(G_{2:2}), \Sigma, \Phi)$. We regard $U \in \mathcal{U}$ both as a tuple of $k$ equations $(e_1, \ldots, e_k)$, and as the set of $3k$ variables appearing in these equations, say $(x_{11}, x_{12}, x_{13}, \ldots, x_{k1}, x_{k2}, x_{k3})$. For each equation $e_i$, define a vector $v_{e_i} \in \{0, 1\}^L$ that has 1 on coordinates corresponding to variables in $e_i$ and 0 on the rest. Denote $H_U = \text{Span}\{v_{e_1}, \ldots, v_{e_k}\}$ referred to as the space of side condition. Let $b_1, \ldots, b_k \in \{0, 1\}$ be the “right hand sides” of the equations, i.e. the equation $e_i$ is $x_{i1} + x_{i2} + x_{i3} = b_i$. Let $\psi_{H_U} : H_U \to \{0, 1\}$ be the linear function on $H_U$ defined by $\psi_{H_U}(v_{e_i}) = b_i$ for $1 \leq i \leq k$. The pair $(H_U, \psi_{H_U})$ is referred to as the side condition. Define

$$\mathcal{L}_U = \left\{L \subseteq \{0, 1\}^U \mid \dim(L) = \ell, \; L \cap H_U = \{0\} \right\}.$$ 

We note that for $L \in \mathcal{L}_U$, we have $L \cap H_U = \{0\}$ and $\dim(L \oplus H_U) = \ell + k$. Also, $|U| = 3k$, $\dim(H_U) = k$, $\dim(L) = \ell$. The fraction of $\ell$-spaces $L \subseteq \{0, 1\}^U$ such that $L \cap H_U \neq \{0\}$ is negligible ($\approx 2^{\ell - 2k}$).

**Vertices of $G_{2:2}$:** The game $G_{2:2}$ has a block of vertices $\text{Block}[U]$ for every $U \in \mathcal{U}$ defined as

$$\text{Block}[U] = \{ L \oplus H_U \mid L \in \mathcal{L}_U \}.$$ 

The vertex set of $G_{2:2}$ is the (disjoint) union of all blocks:

$$V(G_{2:2}) = \bigcup_{U \in \mathcal{U}} \text{Block}[U].$$ 

**Colors of $G_{2:2}$:** The set of colors $\Sigma$ has size $|\Sigma| = 2^\ell$. For a vertex $L \oplus H_U$, its color set $\Sigma$ is identified with

$$\{ \phi : L \oplus H_U \to \{0, 1\} \mid \phi \text{ is linear}, \; \phi|_{H_U} = \psi_{H_U} \}.$$ 

In words, the vertex $L \oplus H_U$ is to be assigned a linear function $\phi : L \oplus H_U \to \{0, 1\}$ that “respects the side condition”, meaning the restriction of $\phi$ to $H_U$ coincides with the given $\psi_{H_U}$. Since the values of $\phi$ are already determined on $H_U$, there are exactly $2^\ell$ eligible linear functions $\phi$.

**Edges and Constraints of $G_{2:2}$:** Before formally defining the edges and constraints of the game $G_{2:2}$, we stress a notational (and perhaps conceptual) point. $X$ is the set of all variables in the Gap3Lin instance, so $U \subseteq X$ and $\{0, 1\}^U$ is a subspace of $\{0, 1\}^X$ in a natural manner. Every subspace under consideration can be considered as a subspace of $\{0, 1\}^X$ and we can freely take the intersections or direct sums of subspaces. For instance if $U_1, U_2$ are two sets of variables and $L_1 \subseteq \{0, 1\}^{U_1}$, $L_2 \subseteq \{0, 1\}^{U_2}$ are subspaces, we can consider both $L_1, L_2$ as subspaces of $\{0, 1\}^{U_1 \cup U_2}$ (which in turn is a subspace of $\{0, 1\}^X$) and then the subspaces $L_1 \cap L_2, L_1 \oplus L_2$ are well defined.
We are ready to define the edges and the constraints of $G_{2:2}$. For $U, U' \in \mathcal{U}$ (allowing the possibility that $U = U'$), we describe the edges between their respective blocks. There is an edge between vertices $L \oplus H_U, L' \oplus H_{U'}$ if either of the following two conditions holds:

$$\dim(L \oplus H_U \oplus H_{U'}) = \dim(L' \oplus H_U \oplus H_{U'}) = \dim(L \oplus L' \oplus H_U \oplus H_{U'}), \quad (3.3)$$

in which case, the constraint is 1-to-1, or

$$\dim(L \oplus H_U \oplus H_{U'}) = \dim(L' \oplus H_U \oplus H_{U'}) = \dim(L \oplus L' \oplus H_U \oplus H_{U'}) - 1, \quad (3.4)$$

in which case, the constraint is 2-to-2. We recommend reading the proofs of Lemmas 3.4.1, 3.4.2, 3.4.3 to start gaining some intuition. We first consider the 1-to-1 constraints.

1-to-1 Constraints: From Lemma 3.4.1, we always have

$$\dim(L \oplus H_U \oplus H_{U'}) = \dim(L' \oplus H_U \oplus H_{U'}).$$

If, in addition, this dimension is same as that of $L \oplus L' \oplus H_U \oplus H_{U'}$ which contains both the spaces above, then all the three spaces must be identical, i.e. $L \oplus H_U \oplus H_{U'} = L' \oplus H_U \oplus H_{U'} = L \oplus L' \oplus H_U \oplus H_{U'} = Z$, say. From Lemma 3.4.3, there is a 1-to-1 correspondence between linear functions on $L \oplus H_U$ (that respect side condition on $H_U$) and linear functions on $L \oplus H_U \oplus H_{U'} = Z$ (that respect side condition on both $H_U, H_{U'}$), and the same holds between $L' \oplus H_{U'}$ and $L' \oplus H_U \oplus H_{U'} = Z$. This gives a 1-to-1 mapping between linear functions on $L \oplus H_U$ and $L' \oplus H_{U'}$ (respecting the relevant side conditions) which is regarded as the 1-to-1 constraint on the “coloring” of $L \oplus H_U$ and $L' \oplus H_{U'}$.

2-to-2 Constraints: As before, from Lemma 3.4.1, we always have

$$\dim(A = L \oplus H_U \oplus H_{U'}) = \dim(A' = L' \oplus H_U \oplus H_{U'}) = d \text{ (say)}.$$  

Now suppose that $Z = L \oplus L' \oplus H_U \oplus H_{U'}$, $\dim(Z) = d + 1$. Since $Z = A \oplus A'$, it follows that $\dim(A \cap A') = d - 1$. Thus, it is possible to choose a basis $I$ for $A \cap A'$ and $v \in L, v' \in L'$ so that $I \cup \{v\}$ is a basis for $A$, $I \cup \{v'\}$ is a basis for $A'$, and $I \cup \{v, v'\}$ is a basis for $Z$. In the following, all linear functions considered are supposed to respect the side condition on $H_U$ or $H_{U'}$ or both, depending on whether the relevant space contains $H_U, H_{U'}$ or both.

Every linear function $f$ on $A \cap A' = \text{Span}(I)$ has exactly two extensions $f_1, f_2$ to $A = \text{Span}(I \cup \{v\})$, depending on their value on $v$, and has exactly two extensions $f'_1, f'_2$ to $A' = \text{Span}(I \cup \{v'\})$, depending on their value on $v'$. Moreover, by Lemma 3.4.3, there is a 1-to-1 mapping between linear functions on $A$, and linear functions on $L \oplus H_U$. Denote by $\tilde{f}_1, \tilde{f}_2$ the functions corresponding to $f_1, f_2$ respectively via this mapping. Similarly, there is a 1-to-1 mapping from linear functions on $A'$ to linear functions on $L' \oplus H_{U'}$, and let $\tilde{f}'_1, \tilde{f}'_2$ be the functions corresponding to $f'_1, f'_2$ via this mapping. This gives a 2-to-2 constraint between $L \oplus H_U$ and $L' \oplus H_{U'}$ that matches the pair $(\tilde{f}_1, \tilde{f}_2)$ with the pair $(\tilde{f}'_1, \tilde{f}'_2)$. This completes the description of the game $G_{2:2}$.

The final construction is simply a weighted version of $G_{2:2}$, however to define this weight we must first prove additional properties of the construction called “transitivity” – this is done in the following section.

Auxiliary Lemmas

**Lemma 3.4.1.** Let $U, U' \in \mathcal{U}$ and $\text{Eq}[U], \text{Eq}[U']$ denote the sets of equations (or number) in $U, U'$ respectively. Then, for $L \in \mathcal{L}_U$,

$$\dim(L \oplus H_U \oplus H_{U'}) = \ell + 2k - |\text{Eq}[U] \cap \text{Eq}[U']|.$$
Proof. Let Eq[$U'$] = \{e'_1, \ldots, e'_k\} and recall that $H_{U'} = \text{Span}(v_{e'_1}, \ldots, v_{e'_k})$. Let $C$ denote the “current space” that is initialized to $C = L \oplus H_U$ and has dimension $\ell + k$. We consider equations $e'_1, \ldots, e'_k \in \text{Eq}[U']$ one by one and check whether “adding” $v_{e'_i}$ to the current space increases its dimension. If the equation $e'_i \in \text{Eq}[U]$, then $v_{e'_i} \in H_U$ already, and hence $\dim(C \oplus \text{Span}(v_{e'_i})) = \dim(C)$. Otherwise $e'_i \notin \text{Eq}[U]$ and shares at most one variable with $U \cup (U' \setminus e'_i)$. This is where we use the fact that $U, U'$ are “legitimate” tuples in the sense described in the first paragraph of the current section. Thus $v_{e'_i}$ is linearly independent of $L \oplus H_U \oplus \bigoplus_{j \neq i} \text{Span}(v_{e'_j})$. Hence $C \oplus \text{Span}(v_{e'_i})$ has dimension 1 larger than that of $C$. Carrying the argument for $i = 1, \ldots, k$ shows that in the end $C = L \oplus H_U \oplus H_{U'}$ and $\dim(C)$ is as desired. \qed

**Lemma 3.4.2.** Let $U, U' \in U$ and $L \in \mathcal{L}_U$, $L' \in \mathcal{L}_{U'}$. Then
\[
\dim(L \oplus H_U \oplus H_{U'}) = \dim(L' \oplus H_U \oplus H_{U'}). \]

**Proof.** From Lemma 3.4.1, both the dimensions are equal to $\ell + 2k - |\text{Eq}[U] \cap \text{Eq}[U']|$. \qed

**Lemma 3.4.3.** Let $U, U' \in U$ and $L \in \mathcal{L}_U$. Then any linear function on $L \oplus H_U$ that respects the side condition on $H_U$, has a unique extension to $L \oplus H_U \oplus H_{U'}$ that respects the side condition on both $H_U$ and $H_{U'}$.

**Proof.** Let $f$ be a linear function on $L \oplus H_U$ that respects the side condition on $H_U$. Clearly, it has at most one extension to $L \oplus H_U \oplus H_{U'}$ that respects the side condition on both $H_U$ and $H_{U'}$, so the main point is to show that there indeed is such an extension. Similar to the proof of Lemma 3.4.1, let $C$ denote the current space, $g$ denote the current linear function on $C$, so that initially $C = L \oplus H_U, g = f$ and at each step, $g$ respects the side condition on $H_U$ and the side condition due to equations $e'_1, \ldots, e'_{i-1}$ considered so far. Consider the equation $e'_i$. If $e'_i \in \text{Eq}[U]$ then $C \oplus \text{Span}(v_{e'_i}) = C$ and we keep $g$ unchanged and proceed next. If $e'_i \notin \text{Eq}[U]$, then as in the proof of Lemma 3.4.1, $v_{e'_i}$ is linearly independent of $L \oplus H_U \oplus \bigoplus_{j \neq i} \text{Span}(v_{e'_j})$. Hence $C \oplus \text{Span}(v_{e'_i})$ has dimension 1 larger than that of $C$ and the function $g$ can be safely extended to vector $v_{e'_i}$ as required. To be precise, one sets $g(v_{e'_i}) = b'_i$ where $b'_i$ is the “right hand side” of the equation $e'_i$ and then extends $g$ linearly to $C \oplus \text{Span}(v_{e'_i})$. Carrying the argument for $i = 1, \ldots, k$, completes the proof. \qed

### 3.4.2 The Game $G_{2,2}$ is a Transitive Game

To proceed with the construction, we must first prove a structural property of the 2-to-2 instance constructed so far. We begin with a definition.

**Definition 3.4.4.** A Transitive 2-to-2 Game is a game $G = (V, E, \Phi, \Sigma)$ (see Definition 1.2.1) where

- Each constraint $\Phi(u, v)$ is a 2-to-2 constraint or a 1-to-1 constraint.

- Transitivity: If there is a 1-to-1 constraint $\Phi(u, v)$ and a 1-to-1 or a 2-to-2 constraint $\Phi(v, w)$, then there is also a constraint $\Phi(u, w)$. The constraint $\Phi(u, w)$ is either 1-to-1 or 2-to-2 depending on whether $\Phi(v, w)$ is 1-to-1 or 2-to-2 respectively. Moreover, the constraint $\Phi(u, w)$ is a composition of constraints $\Phi(u, v)$ and $\Phi(v, w)$, i.e. for every $a, b, c \in \Sigma$,

\[
(a, b) \in \Phi(u, v), (b, c) \in \Phi(v, w) \Rightarrow (a, c) \in \Phi(u, w). \]

Given a Transitive 2-to-2 Game $G = (V, E, \Phi, \Sigma)$, one can define an equivalence relation on the set of its variables where two variables $u, v$ are equivalent if either $u = v$ or if there is a 1-to-1 constraint $\Phi(u, v)$. This equivalence relation partitioned $V = \text{Clique}_1 \cup \ldots \cup \text{Clique}_m$ into equivalency classes, referred to as
cliques. The 2-to-2 constraints are across the cliques. For every $1 \leq i \neq j \leq m$, either there is no constraint across variables in $C_i$ and $C_j$ or there is a constraint between every pair of variables, one in $C_i$ and one in $C_j$.

The following Theorem captures the additional structure we require from the game. We encourage the reader to skip the proof at first reading to get a complete picture of the construction.

**Theorem 3.4.5.** The game $G_{2:2}$ is a transitive game in the sense of Definition 3.4.4.

**Proof.** The proof is divided into two lemmas, the first showing that if there is a 1-to-1 constraint between $u$, $v$ and some constraint between $v$, $w$ (1-to-1 or 2-to-2), then there is a constraints between $u$ and $w$. The second lemma shows the “constraint composition property”.

**Lemma 3.4.6.** Suppose $L_1 \oplus H_{U_1}, L_2 \oplus H_{U_2}$ have a 1-to-1 constraint between them in $G_{2:2}$, and $L_2 \oplus H_{U_2}, L_3 \oplus H_{U_3}$ have a 1-to-1 or a 2-to-2 constraint between them. Then there is a constraint between $L_1 \oplus H_{U_1}, L_3 \oplus H_{U_3}$, and it is 1-to-1 or a 2-to-2 depending on whether the constraint between $L_2 \oplus H_{U_2}, L_3 \oplus H_{U_3}$ is 1-to-1 or 2-to-2 respectively.

**Proof.** Since there is 1-to-1 constraint between $L_1 \oplus H_{U_1}, L_2 \oplus H_{U_2}$, we have

$$L_1 \oplus H_{U_1} \oplus H_{U_2} = L_2 \oplus H_{U_1} \oplus H_{U_2}. \quad (3.5)$$

We first consider the case when the constraint between $L_2 \oplus H_{U_2}, L_3 \oplus H_{U_3}$ is also 1-to-1. This gives

$$L_2 \oplus H_{U_2} \oplus H_{U_3} = L_3 \oplus H_{U_2} \oplus H_{U_3}. \quad (3.6)$$

Combining the above equations gives (“add” $H_{U_1}$ to Equation (3.6) and do a “substitution” using Equation (3.5))

$$L_1 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3} = L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}. \quad (3.7)$$

Now we would like to “remove” $H_{U_2}$ from both the sides so as to obtain $L_1 \oplus H_{U_1} \oplus H_{U_3} = L_3 \oplus H_{U_1} \oplus H_{U_3}$ and implying that there is a 1-to-1 constraint between $L_1 \oplus H_{U_1}$ and $L_3 \oplus H_{U_3}$. This “removal” can be done for the following reason. Write $H_{U_2} = A \oplus B$, $A \cap B = \{0\}$ where (a) $A$ is the span of all vectors $v_e$ such that the equation $e$ occurs in $U_2$, but also occurs in $U_1$ or $U_3$, and hence $A \subseteq H_{U_1} \cup H_{U_3}$. (b) $B$ is the span of all vectors $v_e$ such that the equation $e$ occurs in $U_2$, but not in $U_1$ nor in $U_3$. Any such equation $e$ shares at most one variable with $U_1$ and at most one variable with $U_3$ and no variable with $U_2 \setminus e$. Hence there is a variable that is “private” to $e$, meaning it does not occur in $U_1 \cup U_3 \cup (U_2 \setminus e)$. In particular, the “private” variables of the equations contributing to $B$ are distinct. Thus the intersection of $B$ and $L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_3} \subseteq \{0, 1\}^{U_1} \oplus \{0, 1\}^{U_3}$ is $\{0\}$. To summarize, we can write Equation (3.7) as

$$L_1 \oplus (H_{U_1} \oplus H_{U_3} \oplus A) \oplus B = L_3 \oplus (H_{U_1} \oplus H_{U_3} \oplus A) \oplus B,$$

which simplifies to

$$(L_1 \oplus H_{U_1} \oplus H_{U_3}) \oplus B = (L_3 \oplus H_{U_1} \oplus H_{U_3}) \oplus B,$$

and we can now safely “remove” $B$, using Fact 3.8.4.

We now consider the case when the constraint between $L_2 \oplus H_{U_2}, L_3 \oplus H_{U_3}$ is 2-to-2. We have Equation (3.5) as before, but instead of Equation (3.6), we now have

$$\dim(L_2 \oplus H_{U_2} \oplus H_{U_3}) = \dim(L_3 \oplus H_{U_2} \oplus H_{U_3}) = \dim(L_2 \oplus L_3 \oplus H_{U_2} \oplus H_{U_3}) - 1. \quad (3.8)$$

We claim that one can “add” $H_{U_1}$ to all three “sums” in Equation (3.8). Arguing as earlier, one can write $H_{U_2} = A' \oplus B'$, $A' \cap B' = \{0\}$ where $A' \subseteq H_{U_2} \cup H_{U_3}$ and $B'$ is linearly independent of $\{0, 1\}^{U_2} \oplus \{0, 1\}^{U_3}$. Thus “adding” $H_{U_1}$ to all three “sums”, increases the dimension of each “sum” by precisely $\dim(B')$. Thus

$$\dim(L_2 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \dim(L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \dim(L_2 \oplus L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) - 1.$$
Using Equation (3.5) and “substituting”, we get
\[ \dim(L_1 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \dim(L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \dim(L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) - 1. \]

Now, arguing as earlier again, we “remove” $H_{U_2}$ from all three “sums”. One can write $H_{U_2} = A \oplus B$, $A \cap B = \{0\}$ where $A \subseteq H_{U_1} \oplus H_{U_3}$ and $B$ intersects $L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_3}$ only at $\{0\}$. Thus “removing” $H_{U_2}$ from all three “sums”, decreases the dimension of each “sum” by precisely $\dim(B)$. Thus
\[ \dim(L_1 \oplus H_{U_1} \oplus H_{U_3}) = \dim(L_3 \oplus H_{U_1} \oplus H_{U_3}) = \dim(L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_3}) - 1, \]

implying that there is a 2-to-2 constraint between $L_1 \oplus H_{U_1}$ and $L_3 \oplus H_{U_3}$.

**Lemma 3.4.7.** Let $s_1 = L_1 \oplus H_{U_1}$, $s_2 = L_2 \oplus H_{U_2}$, $s_3 = L_3 \oplus H_{U_3}$ be vertices in $G_{2,2}$ such that there is 1-to-1 constraint between $(s_1, s_2)$ and a constraint between $(s_2, s_3)$. Then the constraint between $(s_1, s_3)$ (as guaranteed by Lemma 3.4.6) is a composition of the constraints between $(s_1, s_2)$ and $(s_2, s_3)$.

Specifically, if linear functions (respecting relevant side conditions) $f$ on $L_1 \oplus H_{U_1}$, $g$ on $L_2 \oplus H_{U_2}$, and $h$ on $L_3 \oplus H_{U_3}$ are such that $(f, g)$ satisfy $(s_1, s_2)$ and $(g, h)$ satisfy $(s_2, s_3)$, then $(f, h)$ satisfy $(s_1, s_3)$.

**Proof.** In the following, whenever we construct a linear function on a certain space, it will always respect the side condition contained in that space. Since $(g, h)$ satisfy the constraint $(s_2, s_3)$, there is a linear function $\beta$ on $W = L_2 \oplus L_3 \oplus H_{U_2} \oplus H_{U_3}$ that respects side conditions on $H_{U_2}$ and $H_{U_3}$ and
\[ g = \beta|_{L_2 \oplus H_{U_2}}, \quad h = \beta|_{L_3 \oplus H_{U_3}}. \]

Let $Z = W \oplus H_{U_1}$ and extend the linear function $\beta$ on $W$ uniquely to a linear function $\gamma$ on $Z$ so as to respect the side condition $H_{U_1}$. This is possible because every equation in $U_1$ that does not appear in $U_2$ or $U_3$ has a “private variable”, as in the proof of the previous lemma. We note that
\[ \gamma|_{L_2 \oplus H_{U_2}} = (\gamma|_W)|_{L_2 \oplus H_{U_2}} = \beta|_{L_2 \oplus H_{U_2}} = g, \]
\[ \gamma|_{L_3 \oplus H_{U_3}} = (\gamma|_W)|_{L_3 \oplus H_{U_3}} = \beta|_{L_3 \oplus H_{U_3}} = h. \]

Since there is a 1-to-1 constraint $(s_1, s_2)$, we have
\[ L_1 \oplus H_{U_1} \oplus H_{U_2} = L_2 \oplus H_{U_1} \oplus H_{U_2}. \]

It therefore holds that
\[ Z = L_2 \oplus L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3} \supseteq L_2 \oplus H_{U_1} \oplus H_{U_2} = L_1 \oplus H_{U_1} \oplus H_{U_2} \supseteq L_1 \oplus H_{U_1}. \]

Since $\gamma$ is an assignment on $Z \gamma|_{L_1 \oplus H_{U_1}}$ and $\gamma|_{L_2 \oplus H_{U_2}} = g$ satisfy the constraint $(s_1, s_2)$. However $(f, g)$ is supposed to satisfy this 1-to-1 constraint, and hence we must have $f = \gamma|_{L_1 \oplus H_{U_1}}$. Now we have $f = \gamma|_{L_1 \oplus H_{U_1}}$ and $h = \gamma|_{L_3 \oplus H_{U_3}}$, and thus $f, h$ are restrictions of a common linear function satisfying the side-conditions. In particular, $(f, h)$ satisfy the constraint $(s_1, s_3)$.

\[ \Box \]
3.4.3 Construction of the Final 2-to-2 Games Instance

We now define the final 2-to-2 Games instance $G_{\text{final}}$. As stated before, $G_{\text{final}}$ is simply a weighted version of $G_{2:2}$ in the following sense:

**Definition 3.4.8.** Let $G = (V, E, \Phi, \Sigma)$ be an instance of a Transitive 2-to-2 Game with the partition $V = C_1 \cup \ldots \cup C_m$ into cliques. In a weighted version, (only) the 2-to-2 constraints are assigned non-negative weights (the weight function is omitted from the notation) so that

- The total weight of all the 2-to-2 constraints equals 1.
- For any pair $1 \leq i \neq j \leq m$, if there are some (and hence all) constraints across cliques $C_i$ and $C_j$, the weights of all these constraints, $|C_i| \cdot |C_j|$ many of them, are all equal.

For an assignment $A : V \rightarrow \Sigma$, $A$ is clique-consistent if it satisfies all the 1-to-1 constraints (which form the cliques $C_1, \ldots, C_m$). The fraction of constraints satisfied by $A$ refers to the weighted fraction of (only) the 2-to-2 constraints satisfied by $A$.

It is well known that hardness for weighted instance can be converted into hardness of unweighted instances [CST01], and thus to prove Theorem 1.2.6 we only need to show hardness for weighted 2-to-2 instances.

Fix a tuple $U \in \mathcal{U}$ and let $H_U$ be the corresponding side condition. As before,

$$
\mathcal{L}_U = \left\{ L \subseteq \{0, 1\}^U \mid \dim(L) = \ell, \ L \cap H_U = \emptyset \right\}, \quad \text{Block}[U] = \{ L \oplus H_U \mid L \in \mathcal{L}_U \}.
$$

We define an equivalence relation $\sim$ on $\text{Block}[U]$ as follows: $L \oplus H_U \sim L' \oplus H_U$ if $L \oplus H_U = L' \oplus H_U$. Let

$$
\text{Block}[U] = C_1 \cup C_2 \cup \ldots \cup C_m
$$

be the partition of $\text{Block}[U]$ into the corresponding equivalence classes. We call two equivalency classes $C_i, C_j, i \neq j$ “adjacent” if there are $L \in C_i, L' \in C_j$ such that $\dim(L \cap L') = \ell - 1$. Relating to the construction of the game $G_{2:2}$, it is easily seen that there are 1-to-1 constraints between all pairs of vertices inside class $C_i$ whereas there are 2-to-2 constraints between all pairs of vertices across “adjacent” classes $C_i, C_j$. In particular, the classes $C_1, \ldots, C_m$ are contained in (distinct) “cliques” $\text{Clique}_1, \ldots, \text{Clique}_m$ respectively in the transitive game $G_{2:2}$. The weighted game $G_{\text{final}}$ is defined as follows.

- Pick $U \in \mathcal{U}$ at random. Let $\text{Block}[U] = C_1 \cup C_2 \cup \ldots \cup C_m$ be the partition into equivalence classes, with the classes contained in cliques $\text{Clique}_1, \ldots, \text{Clique}_m$ respectively.
- Pick a random pair $L \oplus H_U, L' \oplus H_U$ with $\dim(L \cap L') = \ell - 1$. Suppose

$$
L \oplus H_U \in C_i \subseteq \text{Clique}_i, \quad L' \oplus H_U \in C_j \subseteq \text{Clique}_j.
$$

- Pick a random 2-to-2 constraint between vertices in $\text{Clique}_i$ and $\text{Clique}_j$ respectively, among the $|\text{Clique}_i| \cdot |\text{Clique}_j|$ constraints between these cliques.

**Remark 3.4.9.** The vertex set of $G_{2:2}$ is the disjoint union of all blocks $\text{Block}[U], \ U \in \mathcal{U}$ and all blocks have the same size. Therefore the distribution of the vertex $L \oplus H_U$ above is uniform over the vertex set of $G_{2:2}$. The “left” endpoint of the 2-to-2 constraint chosen is a uniformly random vertex in the clique of the vertex $L \oplus H_U$. Since the cliques partition the overall vertex set, it follows that the left endpoint (and similarly the right endpoint) of the constraint is uniformly distributed over the vertex set of $G_{2:2}$. Combining this with the Cauchy-Schwarz inequality, one gets that any set of vertices of weight $\delta$ contains $\Omega(\delta^2)$ fraction of the edges.
Remark 3.4.10. After picking the tuple $U \in \mathcal{U}$, even though the 2-to-2 constraint picked thereafter is “morally” local to (i.e. inside) the block $\text{Block}[U]$, the cliques $\text{Clique}_1, \ldots, \text{Clique}_m$ actually extend much beyond $\text{Block}[U]$ and hence a random constraint between a pair of these cliques is unlikely to be inside $\text{Block}[U]$.

Remark 3.4.11. The construction can be additionally modified as was done in [DKK+18a]:

- The construction can be made into a 2-to-1 game, by replacing the Grassmann graph with the bipartite graph whose sides are $\ell - 1$ dimensional subspaces and $\ell$ dimensional subspaces, and $(A, B)$ is an edge if $A \subseteq B$. Some care has to be taken with the side conditions and the edge weighting.

- One can in fact fold the 1-to-1 cliques, however since we feel the edge weighting appears more natural without this additional folding.

3.4.4 Completeness

We now prove the completeness property of Theorem 1.2.6. Suppose the Gap3Lin instance $(X, \text{eq})$ has an assignment $\sigma : X \rightarrow \{0, 1\}$ that satisfies $1 - \varepsilon$ fraction of the equations. Let $\mathcal{U}_{\text{sat}} \subseteq \mathcal{U}$ be the set of all tuples $U = (e_1, \ldots, e_k)$ such that all the $k$ equations $e_1, \ldots, e_k$ are satisfied. Clearly, $|\mathcal{U}_{\text{sat}}| \geq (1 - k\varepsilon)|\mathcal{U}| \geq (1 - \frac{k}{2})|\mathcal{U}|$ if $\varepsilon$ is chosen to be sufficiently small (see Section 3.4.5).

For every $U \in \mathcal{U}_{\text{sat}}$, let $\sigma[U]$ denote the restriction of $\sigma$ to $U$ and let $f_{\sigma[U]} : \{0, 1\}^U \rightarrow \{0, 1\}$ be the linear “inner product” function $x \rightarrow \langle \sigma[U], x \rangle$. Since $\sigma[U]$ satisfies all equations inside $U$, the linear function $f_{\sigma[U]}$ respects the side condition $H_U$. Now we assign to every vertex $L \oplus H_U$ in $\text{Block}[U]$, the linear function $f_{\sigma[U]}|_{L \oplus H_U}$. We show that this assignment satisfies all constraints whose both endpoints have been assigned. Indeed if $(L \oplus H_U, L' \oplus H_{U'})$ is a constraint such that both endpoints are assigned, then the constraint is satisfied since all spaces are assigned using the same global assignment $\sigma$. Thus all constraints inside $\bigcup_{U \in \mathcal{U}_{\text{sat}}} \text{Block}[U]$ are satisfied, which constitutes $1 - \frac{k}{2}$ fraction of the vertices of the game $G_{\text{final}}$. By Remark 3.4.9, for a random 2-to-2 constraint, both its endpoints are in $\bigcup_{U \in \mathcal{U}_{\text{sat}}} \text{Block}[U]$ with probability $1 - \delta$ and the constraint is satisfied.

3.4.5 Setting of the Parameters

Let $(X, \text{eq})$ be an instance of regular Gap3Lin$(1 - \varepsilon, s^*)$ as in Theorem 3.3.1. We will use the game $G^{\otimes k}_{\beta, r}$ in Section 3.3.4 as the “Outer PCP”. Since there are several parameters involved, we specify the (tedious) order in which the parameters are chosen.

Let $\delta > 0$ be the parameter required in Theorem 1.2.6. Let $q, r, C, \alpha(\cdot)$ be as given in Theorem 3.2.8 for $\delta/2$, and let $\ell$ be an integer large enough so that Theorem 3.2.8 holds for all sufficiently large integers $k$ ($= n$ therein). Choose $\beta = \log \log k / \ell$.

The soundness analysis of the reduction shows that a $\delta$-consistent assignment to the 2-to-2 Game yields strategies for the provers in the Outer PCP Game with success probability $p(\delta, \ell)$ only depending on $\delta$ and $\ell$, whereas the soundness of the Outer PCP is $2^{-\Omega(\beta k/2^q)} = 2^{-\Omega(\log \log k)}$. By taking $k$ to be large enough, the soundness of the Outer PCP is ensured to be smaller than $p(\delta, \ell)$, implying that the 2-to-2 Game does not have a $\delta$-consistent assignment.

Finally we choose $\varepsilon$ in the 3Lin instance to be $\frac{\delta}{2\ell}$.

3.5 Preliminary steps towards Soundness Analysis

In this section we present several essential components for the soundness analysis in Theorem 1.2.6: the Covering Property and the upper-bound on the number of minimal co-dimension successful zoom-outs.
also show how to incorporate the side conditions into Theorem 3.2.8.

### 3.5.1 Covering Property

We state a basic version of the covering property, and then two refinements of it that will be required in our proof. Throughout this section, the notations and parameters are as in the Outer PCP described in Section 3.4.

Let \( U \) be the set of \( 3k \) variables in a fixed set of \( k \) equations. We recall that in the Outer PCP game, the verifier chooses \( V \subseteq U \) randomly by picking from each equation independently (a) with probability \( \beta \), one of the variables from the equation and (b) with probability \( 1 - \beta \), all three variables from the equation. We consider \( \{0, 1\}^V \) as a subspace of \( \{0, 1\}^U \) in a natural manner. The basic covering property, first introduced by [KS13], states that the statistical distance between the following two distributions over one-dimensional subspaces of \( \{0, 1\}^U \) is small, i.e. at most \( O(\beta \sqrt{k}) \)

- Choose a random one-dimensional subspace \( P \subseteq \{0, 1\}^U \).
- Choose \( V \subseteq U \) as described above, choose a random one-dimensional subspace \( P' \subseteq \{0, 1\}^V \) and regard it as a subspace of \( \{0, 1\}^U \).

Our basic Covering Property shows that two distributions over \( \ell \)-dimensional subspaces of \( \{0, 1\}^U \) are close in statistical distance.

**Definition 3.5.1.** Let \( U \) be the set of \( 3k \) variables in a fixed set of \( k \) equations. Let \( \beta \) be the smoothness parameter and \( V \subseteq U \) be chosen at random. Let \( \ell \geq 1 \) be an integer. Let \( \mathcal{L}, \mathcal{L}' \) be distributions over \( \ell \)-dimensional subspaces of \( \{0, 1\}^U \) sampled as follows.

- \( \mathcal{L} \): Choose a uniformly random \( \ell \)-dimensional subspace of \( \{0, 1\}^U \).
- \( \mathcal{L}' \): Choose \( V \subseteq U \) as in the Outer PCP, choose a uniformly random \( \ell \)-dimensional subspace of \( \{0, 1\}^V \) and regard it as a subspace of \( \{0, 1\}^U \).

**Lemma 3.5.2.** Suppose \( 2^{\ell} \beta \leq \frac{1}{8} \). Let \( \mathcal{L}, \mathcal{L}' \) be distributions over \( \ell \)-dimensional subspaces over \( \{0, 1\}^U \) sampled as in Definition 3.5.1. Then the statistical distance between \( \mathcal{L}, \mathcal{L}' \) is bounded as

\[
\text{SD}(\mathcal{L}, \mathcal{L}') \leq \beta \sqrt{k} \cdot 2^{\ell+4}.
\]

The proof of this lemma is along the lines of the the Covering Property of Khot and Safra [KS13], using the Hellinger Distance between distributions. The proof is deferred to Section 3.8.1.

The following lemma is a variant of the above covering property that is compatible with the notion of “zoom-ins”.

**Lemma 3.5.3.** Let \( 0 \leq q \leq \ell - 1 \) be an integer. Let \( Q \) be \( q \)-dimensional subspace of \( \{0, 1\}^U \). Let \( \mathcal{L}_Q \) and \( \mathcal{L}'_Q \) be distributions \( \mathcal{L} \) and \( \mathcal{L}' \) conditioned on the event that a sampled \( \ell \)-subspace \( L \) contains \( Q \). Suppose \( 2^{\ell} \beta \leq \frac{1}{8} \). Then for at least \( 1 - \sqrt{\beta} k^{\frac{3}{2}} \) fraction of \( Q \),

\[
\text{SD}(\mathcal{L}_Q, \mathcal{L}'_Q) \leq \sqrt{\beta} k^{\frac{3}{2}} \cdot 2^{\ell+5}.
\]  

We defer the (straightforward) derivation of Lemma 3.5.3 from Lemma 3.5.2 to Section 3.8.1.

Finally, we need a version of the covering property that is also compatible with zoom-outs. Note that \( \mathcal{L}_Q \) is a uniform distribution on \( \text{Zoom}([Q, \{0, 1\}^U]) \), that \( \mathcal{L}_{Q,W} \) is a uniform distribution on \( \text{Zoom}([Q, W]) \) and that for subspace \( W \) with co-dimension \( r \), \( \text{Zoom}([Q, W]) \) contains roughly \( 2^{-r(\ell-q)} \) fraction of the subspaces in \( \text{Zoom}(Q, \{0, 1\}^U) \). This observation, together with Lemma 3.5.3, immediately yields the lemma below. We omit the full proof.
The "pairwise independence" means that the subspaces have focused on the same zoom-in set by introducing the idea of advice in the Outer PCP Game. Thus, we shall assume from now on the provers wish to use Theorem $G$ onto the Grassmann graph $G$. Suppose the provers from the Outer PCP game (from Section 3.5.2 The Number of Successful Zoom-outs of Minimal Co-dimension is Bounded) receive questions $V \subseteq U$, and they have $\delta$-consistent tables to the Grassmann graph $G(\{0, 1\}^U, \ell)$. Observe that this naturally induces an assignment onto the Grassmann graph $G(\{0, 1\}^V, \ell)$, which we will further assume to also be $\delta$-consistent. The provers wish to use Theorem 3.2.8 to get global functions $f_U, f_V$ that correspond to answers to their questions, however there are several obstacles in their way. For this strategy to have any chance, the provers must use the same zoom-in subspace $Q$, and a matching zoom-out space, in the sense that $W_V = W_U \cap \{0, 1\}^V$ where $W_V, W_U$ are the zoom-out spaces the smaller and larger prover choose respectively.

In previous sections, we have already described how the provers may agree on their zoom-in subspace by introducing the idea of advice in the Outer PCP Game. Thus, we shall assume from now on the provers have focused on the same zoom-in set $Q$ and ignore it henceforth, i.e. assume $Q = \{0\}$. This simplifies notations and only comes at the cost of introducing $2^n$ factors in some arguments, which is thought of as a constant anyway.

The goal of this section is to resolve the issue of agreeing on a zoom-out space. Denote by $r$ be the co-dimension of the zoom-out space. As a first attempt for the provers to correlate their zoom-out space, one may try to simply let the verifier sample a successful co-dimension $r$ subspace $W$, and send it to the provers. However, since the number of successful zoom-outs might be small, a prover may receive a lot of information about the question of the other prover by the identity of the zoom-out space, thereby destroying the soundness guarantee of the Outer PCP Game. What can we do then, if the number of zoom-outs is not that large?

If the number of successful zoom out of co-dimension $r$ is very small, i.e. upper bounded by some function of $\ell$ (but not of $n = \dim(V)$), then each prover can simply sample a successful zoom-out, and since there is a bounded number of successful zoom-outs, the probability they choose a matching zoom-out subspace is non-negligible.

However, there are cases in which such upper bound does not hold; when a table $F[\cdot]$ is, as intended, derived from a global linear function $f$, then every zoom-out space $W$ of co-dimension $r$ is successful along with the linear function $f|_W$ on it. We propose to circumvent this difficulty as follows.

The key observation is that if there are too many ($\geq 2^{|r^2|\ell}$) subspaces $W$ of co-dimension $r$ that are successful zoom-outs, then there is a subspace $W^*$ of co-dimension less than $r$ that serves as a successful zoom-out. Thus, if $r^*$ is the minimal co-dimension in which a successful zoom-out exists, then there is indeed an upper bound of $2^{|r^*|^2\ell}$ on the number of successful zoom-outs in that co-dimension. The provers can agree (by selecting $0 \leq r^* \leq r$ at random) to consider zoom-outs in that co-dimension and then the number of successful zoom-outs is bounded, and the issue is resolved! This observation is stated formally as Lemma 3.5.7. Lemmas 3.5.5, 3.5.6 are used as auxiliary lemmas in its proof.

The argument uses “pairwise independent” collection of subspaces $W_1, \ldots, W_N$ of a global space $V$. The “pairwise independence” means that the subspaces $W_i, 1 \leq i \leq N$, all have the same co-dimension, say $r$, and for all pairs, $1 \leq i \neq j \leq N$, the intersection $W_i \cap W_j$ has co-dimension $2r$. It is easy to see that

\begin{equation}
SD(L_{Q,W}, L'_{Q,W}) \leq \sqrt{3} k^{\frac{1}{2}} \cdot 2^{\ell+5} \cdot 2^{r(q-\ell)+5}.
\end{equation}

3.5.2 The Number of Successful Zoom-outs of Minimal Co-dimension is Bounded

Suppose the provers from the Outer PCP game (from Section 3.3) receive questions $V \subseteq U$, and they have $\delta$-consistent tables to the Grassmann graph $G(\{0, 1\}^U, \ell)$. Observe that this naturally induces an assignment onto the Grassmann graph $G(\{0, 1\}^V, \ell)$, which we will further assume to also be $\delta$-consistent. The provers wish to use Theorem 3.2.8 to get global functions $f_U, f_V$ that correspond to answers to their questions, however there are several obstacles in their way. For this strategy to have any chance, the provers must use the same zoom-in subspace $Q$, and a matching zoom-out space, in the sense that $W_V = W_U \cap \{0, 1\}^V$ where $W_V, W_U$ are the zoom-out spaces the smaller and larger prover choose respectively.

In previous sections, we have already described how the provers may agree on their zoom-in subspace by introducing the idea of advice in the Outer PCP Game. Thus, we shall assume from now on the provers have focused on the same zoom-in set $Q$ and ignore it henceforth, i.e. assume $Q = \{0\}$. This simplifies notations and only comes at the cost of introducing $2^n$ factors in some arguments, which is thought of as a constant anyway.

The goal of this section is to resolve the issue of agreeing on a zoom-out space. Denote by $r$ be the co-dimension of the zoom-out space. As a first attempt for the provers to correlate their zoom-out space, one may try to simply let the verifier sample a successful co-dimension $r$ subspace $W$, and send it to the provers. However, since the number of successful zoom-outs might be small, a prover may receive a lot of information about the question of the other prover by the identity of the zoom-out space, thereby destroying the soundness guarantee of the Outer PCP Game. What can we do then, if the number of zoom-outs is not that large?

If the number of successful zoom out of co-dimension $r$ is very small, i.e. upper bounded by some function of $\ell$ (but not of $n = \dim(V)$), then each prover can simply sample a successful zoom-out, and since there is a bounded number of successful zoom-outs, the probability they choose a matching zoom-out subspace is non-negligible.

However, there are cases in which such upper bound does not hold; when a table $F[\cdot]$ is, as intended, derived from a global linear function $f$, then every zoom-out space $W$ of co-dimension $r$ is successful along with the linear function $f|_W$ on it. We propose to circumvent this difficulty as follows.

The key observation is that if there are too many ($\geq 2^{|r^2|\ell}$) subspaces $W$ of co-dimension $r$ that are successful zoom-outs, then there is a subspace $W^*$ of co-dimension less than $r$ that serves as a successful zoom-out. Thus, if $r^*$ is the minimal co-dimension in which a successful zoom-out exists, then there is indeed an upper bound of $2^{|r^*|^2\ell}$ on the number of successful zoom-outs in that co-dimension. The provers can agree (by selecting $0 \leq r^* \leq r$ at random) to consider zoom-outs in that co-dimension and then the number of successful zoom-outs is bounded, and the issue is resolved! This observation is stated formally as Lemma 3.5.7. Lemmas 3.5.5, 3.5.6 are used as auxiliary lemmas in its proof.

The argument uses “pairwise independent” collection of subspaces $W_1, \ldots, W_N$ of a global space $V$. The “pairwise independence” means that the subspaces $W_i, 1 \leq i \leq N$, all have the same co-dimension, say $r$, and for all pairs, $1 \leq i \neq j \leq N$, the intersection $W_i \cap W_j$ has co-dimension $2r$. It is easy to see that

\begin{equation}
SD(L_{Q,W}, L'_{Q,W}) \leq \sqrt{3} k^{\frac{1}{2}} \cdot 2^{\ell+5} \cdot 2^{r(q-\ell)+5}.
\end{equation}
for a random choice of an \( \ell \)-dimensional subspace \( L \subseteq V \), the events \( L \subseteq W_i, 1 \leq i \leq N \), are (essentially) pairwise independent. Indeed, for \( 1 \leq i \neq j \leq N \),

\[
\Pr [L \subseteq W_i] \approx 2^{-\ell}, \quad \Pr [L \subseteq W_i \land L \subseteq W_j] = \Pr [L \subseteq W_i \cap W_j] \approx 2^{-2\ell},
\]

where the approximations above are off by an additive \( 2^{\ell-n} \) which is negligible as \( n \gg \ell \gg r \). This (near-) pairwise independence turns out to be very convenient in our arguments.

We begin by proving Lemmas 3.5.5, 3.5.6 that, roughly speaking, state that every large collection of subspaces of a fixed co-dimension contains a large sub-collection that is pairwise independent, possibly after “redefining” the global space.

**Lemma 3.5.5.** Let \( W_1, \ldots, W_N \) be (distinct) subspaces of \( V \) of co-dimension \( r \). Then for any integer \( m \geq 1 \),

- Either there are \( m \) of these subspaces, say \( W_1, \ldots, W_m \), such that for every pair, \( 1 \leq i \neq j \leq m \), their intersection \( W_i \cap W_j \) has co-dimension \( 2r \).
- Or there is a subspace \( V' \subseteq V \) of co-dimension \( 1 \) that contains \( N' = \frac{N}{m^2r} \) of these subspaces, say \( W_1, \ldots, W_{N'} \).

**Proof.** We first note that for any \( 1 \leq i \neq j \leq N \), the co-dimension of \( W_i \cap W_j \) is at most \( 2r \). Consider a graph whose vertices are \( W_1, \ldots, W_N \), and \( (W_i, W_j) \), \( i \neq j \) is an edge if and only if \( W_i \cap W_j \) has co-dimension less than \( 2r \). In this graph, if all vertices have degree at most \( \frac{N}{m} \), then there is an independent set of size \( m \), satisfying the first conclusion of the lemma. We may assume therefore that there is a vertex, say \( W_N \), that has \( \frac{N}{m} \) neighbors, say \( W_1, \ldots, W_{N'} \). For \( 1 \leq i \leq \frac{N}{m} \), since \( W_i \cap W_N \) has co-dimension less than \( 2r \),

\[
\dim(W_i \oplus W_N) = \dim(W_i) + \dim(W_N) - \dim(W_i \cap W_N) \leq \dim(V) - 1.
\]

Hence there is at least one subspace of \( V \) of co-dimension \( 1 \) that contains \( W_N \) and also contains \( W_i \). On the other hand, there are precisely \( 2^r - 1 \) subspaces of \( V \) of co-dimension \( 1 \) that contain \( W_N \). Therefore, by pigeon-hole principle, one of them, say \( V' \), contains at least \( N' = \frac{N}{m^2r} \) of the subspaces from the list \( W_1, \ldots, W_{N'} \).

The lemma above is used as follows. The goal is to “reach” the first conclusion of the lemma. If \( r = 1 \), since any two distinct subspaces of co-dimension \( 1 \) have an intersection of co-dimension \( 2 \), the first conclusion of the lemma holds automatically. So assume \( r \geq 2 \). If the first conclusion of the lemma does not hold, then there are \( W_1, \ldots, W_{N'} \) subspaces of \( V' \), whose co-dimension in \( V' \) is 1. We “redefine” the global space to be \( V' \) and focus on \( W_1, \ldots, W_{N'} \) whose co-dimension with respect to the new global space is \( r' = r - 1 \). The argument is applied iteratively, reducing the “current” co-dimension \( r' \) of the “current” \( W_1, \ldots, W_{N'} \) with respect to the “current” global space \( V' \) by 1 in each step. Finally, either the first conclusion of the lemma is reached or the “current” co-dimension \( r' \) has dropped to 1, in which case, the first conclusion of the lemma holds automatically. With a careful choice of quantitative parameters, one gets the following lemma.

**Lemma 3.5.6.** Let \( W_1, \ldots, W_N \) be (distinct) subspaces of \( V \) of co-dimension \( r \). Let \( m \) be an integer such that \((m \cdot 2^r)^r \leq N\). Then for some integer \( 1 \leq s \leq r \), there exist \( m \) of these subspaces, say \( W_1, \ldots, W_m \), all contained in some subspace \( U \subseteq V \) of co-dimension at most \( r - s \) such that

- The co-dimension of every \( W_i, 1 \leq i \leq m \), with respect to the space \( U \), is \( s \).
- The co-dimension of every \( W_i \cap W_j, 1 \leq i \neq j \leq m \), with respect to the space \( U \), is \( 2s \).
Finally, we present the key lemma that allows us to move from a large number of successful zoom-out spaces with co-dimension \( r \) to a successful zoom-out space of co-dimension less than \( r \).

**Lemma 3.5.7.** Let \( \delta > 0 \) be a constant, \( r \geq 1 \) be a (constant) integer, \( \ell \) be a large enough integer and \( n \) be even a larger, and large enough, integer. Let \( F[\cdot] \) be a table that assigns, to every vertex \( L \) of the Grassmann graph \( G(V, \ell) \), \( \dim(V) = n \), a linear function \( F[L] \) on \( L \). Suppose there are \( N \geq 2^{8r^2\ell} \) distinct subspaces of co-dimension \( r \), \( W_1, \ldots, W_N \), and linear functions \( g_i : W_i \to \{0, 1\} \) on them, such that for every \( 1 \leq i \leq N \), \( F[L] = g_i|_L \) for at least \( \delta \) fraction of subspaces \( L \in \text{Zoom}[\{0\}, W_i] \). Then there exists a subspace \( U \) of co-dimension less than \( r \) and a linear function \( g : U \to \{0, 1\} \) on it such that

- \( F[L] = g|_L \) for at least \( 10^{-9}\delta^{12} \) fraction of subspaces \( L \in \text{Zoom}[\{0\}, U] \).
- \( U \) contains at least \( d = 2^\ell \) of the \( W_i \), say \( W_1, \ldots, W_d \), such that for \( 1 \leq i \leq d \), \( g|_{W_i} = g_i \).

To prove this lemma we require the following lemma, whose proof is deferred to Section 3.8.2.

**Lemma 3.5.8.** Let \( F[\cdot] \) be a table that assigns, to every \( \ell \)-dimensional subspace \( L \) of an \( n \)-dimensional space \( V \), a linear function \( F[L] \) on \( L \). Suppose for \( \beta \) fraction of pairs \((L, L')\) such that \( \dim(L \cap L') = b \), \( b = \frac{\ell}{10} \), \( F[L] \) and \( F[L'] \) agree on \( L \cap L' \). Then there is a global linear function \( g : V \to \{0, 1\} \) such that \( F[L] = g|_L \) for at least \( \beta \) fraction of \( L \subseteq V \).

**Proof of Lemma 3.5.7.** Applying Lemma 3.5.6, there exists a subspace \( U \) of co-dimension \( r - s \), \( 1 \leq s \leq r \), that contains (by re-indexing) subspaces \( W_1, \ldots, W_m \) such that for all \( 1 \leq i \neq j \leq m \), \( W_i \) has co-dimension \( s \) inside \( U \) and \( W_i \cap W_j \) has co-dimension \( 2s \) inside \( U \). Lemma 3.5.6 gives a lower bound \( m \geq N^{\frac{r^2}{2\ell}} \geq 2^{4r\ell} \). We assume that \( m = \gamma \cdot 2^{r(2\ell-b)} \) (ignoring the rest) where \( \gamma = \frac{\beta}{2} \). Henceforth \( U \) is treated as the global space and all subspaces are subspaces of \( U \).

The basic idea is simple. We claim that for a random pair of \( \ell \)-dimensional subspaces \((L, L')\) in \( U \) such that \( \dim(L \cap L') = b \) where \( b = \frac{\ell}{10} \), say, with good probability \( F[L], F[L'] \) are consistent on \( L \cap L' \). Using Lemma 3.5.8, it then follows that \( F[\cdot] \) has a good agreement with a global linear function. The claim holds for the following reason: it is likely that both \( L, L' \subseteq W_i \) for some \( 1 \leq i \leq m \), since \( m \) is large, and since \( F[\cdot] \) has a good agreement with a linear function \( g_i \) on \( W_i \), it is likely that \( F[L] = g_i|_L \), as well as \( F[L'] = g_i|_{L'} \), and therefore \( F[L], F[L'] \) are consistent on \( L \cap L' \). The formal argument has to take some care of “overcounting issues”, which is the reason a proper size sub-collection was chosen, and is presented below.

Consider a random choice of \( \ell \)-dimensional subspaces \( L, L' \) such that \( \dim(L \cap L') = b \) where \( b = \frac{\ell}{10} \). We note that the pair \((L, L')\) can be chosen by first choosing a random \((2\ell-b)\)-dimensional subspace \( R \) and then choosing random \( \ell \)-dimensional subspaces \( L, L' \subseteq R \) with \( b \)-dimensional intersection. The choice of \( L, L' \) after choosing \( R \) is essentially independent; choosing them independently, it does hold that \( \dim(L \cap L') = b \) except with probability \( 2^{-\Omega(b)} \). Fix an index \( 1 \leq i \leq m \). It will be convenient to define events \( E_i, \mathcal{P}_i, \mathcal{S}_i \) such that \( \mathcal{P}_i \subseteq \mathcal{E}_i, \mathcal{S}_i \subseteq \mathcal{E}_i \) as follows.

- Let \( \mathcal{E}_i \) be the event that \( R \subseteq W_i \) (and in particular \( L, L' \subseteq W_i \)). Then by Fact 3.5.10
  \[ \Pr[\mathcal{E}_i] = \Pr[R \subseteq W_i] \geq (1 - o(1))2^{-s(2\ell-b)}. \]

- Let \( \mathcal{P}_i \) be the event that both \( L, L' \subseteq W_i \), but for \( 1 \leq j \neq i \leq m \), \( W_j \) does not contain both \( L, L' \) (i.e. the pair \((L, L')\) is private to \( W_i \)). Then by Fact 3.5.10
  \[ \Pr[\mathcal{P}_i] \geq \Pr[\mathcal{E}_i] - \sum_{j \neq i} \Pr[\mathcal{E}_i \cap \mathcal{E}_j] \geq (1-o(1))2^{-s(2\ell-b)} - (m-1)2^{-2s(2\ell-b)} \geq (1-o(1))2^{-s(2\ell-b)}. \]

We note that the events \( \mathcal{P}_i, 1 \leq i \leq m \), are disjoint.
• Let $\mathcal{S}_i$ be the event that both $L, L' \subseteq W_i$ and moreover that $F[L] = g_i|_L$, $F[L'] = g_i|_{L'}$ (and in particular $F[L], F[L']$ are consistent on $L \cap L'$). We noted that $R \subseteq W_i$ with probability $\geq (1 - o(1))2^{-s(2\ell - b)}$ and then denoting by $p(R)$, the fraction of $\ell$-dimensional subspaces $L \subseteq R$ for which $F[L] = g_i|_L$,
\[
\Pr[\mathcal{S}_i] = (1 - o(1))2^{-s(2\ell - b)} \cdot \mathbb{E}_{R \subseteq W_i} [p(R)^2] \geq (1 - o(1))2^{-s(2\ell - b)} \cdot \mathbb{E}_{R \subseteq W_i} [p(R)]^2 \geq (1 - o(1))2^{-s(2\ell - b)} \cdot \delta^2.
\]
• In particular,
\[
\Pr[\mathcal{S}_i \land \mathcal{P}_i] = \Pr[\mathcal{S}_i \land \mathcal{P}_i | R \subseteq W_i] \Pr[R \subseteq W_i] \geq (\delta^2 - \gamma - o(1)) \cdot (1 - o(1))2^{-s(2\ell - b)} \geq (1 - o(1)) \frac{\delta^2}{2} \cdot 2^{-s(2\ell - b)}.
\]

The probability that $F[L], F[L']$ are consistent on $L \cap L'$ is at least $(\mathcal{S}_i \land \mathcal{P}_i$ are disjoint)
\[
\Pr[\bigvee_{i=1}^m \mathcal{S}_i \land \mathcal{P}_i] = \sum_{i=1}^m \Pr[\mathcal{S}_i \land \mathcal{P}_i] \geq m(1 - o(1)) \cdot \frac{\delta^2}{2} \cdot 2^{-s(2\ell - b)} = (1 - o(1)) \gamma \cdot \frac{\delta^2}{2} = (1 - o(1)) \frac{\delta^4}{4}. \tag{3.11}
\]

By Lemma 3.5.8, there is a global linear function $g : U \to \{0, 1\}$ that agrees with $F[\cdot]$ on $\frac{\delta^{12}}{1000}$ fraction of $\ell$-dimensional subspaces. This proves the first conclusion of Lemma 3.5.7. We now make a more careful argument and prove both the conclusions of the lemma together.

Let $f_1, \ldots, f_k$ be the list of all global linear functions on $U$ that have $10^{-9}\delta^{12}$ agreement with $F[\cdot]$. We note that the event $\bigvee_{i=1}^m \mathcal{S}_i \land \mathcal{P}_i$ implies the event that $F[L], F[L']$ are consistent on $L \cap L'$. By Lemma 3.5.9, $k \leq 2 \frac{10^9}{\delta^{12}}$ and
\[
\Pr[(\bigvee_{i=1}^m \mathcal{S}_i \land \mathcal{P}_i) \land F[L] \not\in \{f_1|_L, \ldots, f_k|_L\} \land F[L'] \not\in \{f_1|_{L'}, \ldots, f_k|_{L'}\}] \leq \frac{\delta^4}{100}. \tag{3.12}
\]

From Equations (3.11), (3.12) and noting that the roles of $L, L'$ are symmetric,
\[
\Pr[(\bigvee_{i=1}^m \mathcal{S}_i \land \mathcal{P}_i) \land F[L] \in \{f_1|_L, \ldots, f_k|_L\}] \geq \frac{\delta^4}{16}.
\]

Noting again that $\mathcal{S}_i \land \mathcal{P}_i$ are disjoint, the above equation implies
\[
\sum_{i=1}^m \Pr[\mathcal{S}_i \land \mathcal{P}_i \land F[L] \in \{f_1|_L, \ldots, f_k|_L\}] \geq \frac{\delta^4}{16}.
\]

Replacing the event $\mathcal{S}_i \land \mathcal{P}_i$ by its implication $\mathcal{E}_i \land F[L] = g_i|_L$ and further relaxing to implication $g_i|_L \in \{f_1|_L, \ldots, f_k|_L\}$, we have
\[
\sum_{i=1}^m \Pr[\mathcal{E}_i \land g_i|_L \in \{f_1|_L, \ldots, f_k|_L\}] \geq \frac{\delta^4}{16}.
\]

Noting that $m = \gamma \cdot 2^{s(2\ell - b)}$ and $\Pr[\mathcal{E}_i] \leq 2^{-s(2\ell - b)}$ (Fact 3.5.10), we rewrite as
\[
\frac{1}{m} \sum_{i=1}^m \Pr[g_i|_L \in \{f_1|_L, \ldots, f_k|_L\} | \mathcal{E}_i] \geq \frac{1}{\gamma} \cdot \frac{\delta^4}{16} = \frac{\delta^2}{8}.
\]
In the above inequality, $E_i$ is the event that $R, L, L' \subseteq W_i$, but $R, L'$ have no role, so we can rewrite as

$$\frac{1}{m} \sum_{i=1}^{m} \Pr [g_i|L \in \{f_1|L, \ldots, f_k|L\} | L \subseteq W_i] \geq \frac{\delta^2}{8}.$$ 

Now we can finish the proof: By an averaging argument, for at least $d^* = \frac{\delta^2}{16} \cdot m$ of the indices $1 \leq i \leq m$, say for the indices $i = 1, \ldots, d^*$, we have

$$\Pr [g_i|L \in \{f_1|L, \ldots, f_k|L\} | L \subseteq W_i] \geq \frac{\delta^2}{16}.$$ 

Clearly, it must then be the case that $g_i$ is identically equal to one of the functions $f_1|W_i, \ldots, f_k|W_i$, since otherwise $g_i$ would agree with any $f_j|W_i$ for at most $2^{-\ell}$ fraction of $L \subseteq W_i$ and then one could take a union bound over $1 \leq j \leq k$. Finally, we conclude that one of the $k$ functions, $f_1, \ldots, f_k$, call it $g$, satisfies $g|W_i \equiv g_i$ for at least $d = \frac{d^*}{k}$ of the indices $1 \leq i \leq d^*$, say the indices $i = 1, \ldots, d$. We note that $d \geq \Omega(\delta^{14}) \cdot m \geq 2^\ell$.

**Auxiliary Lemmas**

**Lemma 3.5.9.** Let $F[\cdot]$ be a table that assigns, to every $\ell$-dimensional subspace $L$ of an $n$-dimensional space $V$, a linear function $F[L]$ on $L$. Suppose $\ell$ is a sufficiently large integer, $b = \ell_{\beta}$ and $n \geq 2\ell$. Let $g_1, \ldots, g_m$ be the list of all global linear functions on $V$ that have $\beta$-agreement with $F[\cdot]$, i.e. for every $1 \leq i \leq m$, $F[L] = g_i|L$ for at least $\beta$ fraction of subspaces $L \subseteq V$ and moreover every such global linear function appears in the list. Then $m \leq \frac{\beta^{2}}{\beta^2 - 2^{-\ell}}$ and the probability

$$\Pr_{L,L' \in \dim(L \cap L') = b} [F[L]|L \cap L' = F[L']|L \cap L' \wedge F[L] \not\in \{g_1|L, \ldots, g_m|L\} \wedge F[L'] \not\in \{g_1|L', \ldots, g_m|L'\}]$$

is at most $10\sqrt{\beta}$.

**Proof:** The upper bound on $m$ is due to Blinovskiy [Bli86], and is proven in Section 3.8 for completeness. Now assume, on the contrary, that the probability in the statement of the lemma is at least $10\sqrt{\beta}$. We define another table $F^*[\cdot]$ where $F^*[L] = F[L]$ if $F[L] \not\in \{g_1|L, \ldots, g_m|L\}$ (let $L^*$ denote the set of such $L$) and otherwise $F^*[L]$ is defined as a random linear function on $L$. The assumption implies that for $10\sqrt{\beta}$ fraction of pairs $(L, L')$, $\dim(L \cap L') = b$, $F^*[L], F^*[L']$ are consistent on $L \cap L'$. By Lemma 3.5.8, there exists a global linear function $g : V \to \{0, 1\}$ that agrees with $F^*[\cdot]$ on at least $3\beta$ fraction of subspaces $L \subseteq V$. Since $F^*[\cdot]$ is defined at random outside $L^*$, this agreement must essentially be on $L^*$ (one could have used a Chernoff bound and taken a union bound over all global linear functions beforehand). However, $F[\cdot]$ and $F^*[\cdot]$ agree on $L^*$ and hence $g$ agrees with $F[\cdot]$ at $\beta$ fraction of $L \subseteq V$. This is a contradiction since $g$ is distinct from $g_1, \ldots, g_m$; indeed for any $L \in L^*$ such that $F[L] = g|L$, we have $g|L \not\in \{g_1|L, \ldots, g_m|L\}$. 

**Fact 3.5.10.** Let $W$ be subspace of co-dimension $r$ of a space $U$. Then

$$2^{-r\ell}(1 - 2^{r+\ell+1-n}) \leq \Pr_{X \in G(U, \ell)} [X \subseteq W] \leq 2^{-r\ell}.$$ 

**Proof:**

$$\Pr_{X \in G(U, \ell)} [X \subseteq W] = \left[\frac{n-r}{\ell}\right] = \prod_{i=0}^{\ell-1} \frac{2^{n-r} - 2^i}{2^{n} - 2^i} = 2^{-r\ell} \prod_{i=0}^{\ell-1} \left(1 - \frac{2^{i+r} - 2^i}{2^{n} - 2^i}\right).$$

The last product is at most 1 and at least $1 - 2^{r+\ell+1-n}$. 

74
3.5.3 Incorporating Side Condition

Let $G(V, \ell)$ be the Grassmann graph with $\dim(V) = n$. In this section, we augment the Grassmann Linearity Test to incorporate the so-called “side condition”. As we saw in the reduction in Section 3.4, this feature arises naturally at the Outer PCP level and needs to be incorporated at the Inner PCP level so as to facilitate the composition.

**Definition 3.5.11.** A side condition is a pair $(H, \psi_H)$ where $H \subseteq V$ and $\psi_H : H \to \{0, 1\}$ is a linear function. For a subspace $H \subseteq Y \subseteq V$, a linear function $f : Y \to \{0, 1\}$ is said to respect the side condition $(H, \psi_H)$ if $f|_H \equiv \psi_H$. We will assume that $h = \dim(H) \leq \frac{n}{2}$.

The reference to the pair $(H, \psi_H)$ will often be omitted while referring to the side condition. As we did in Section 3.4, it is convenient to restrict the vertex set of the Grassmann graph to those $L \subseteq V$, $\dim(L) = \ell$ such that $L \cap H = \{0\}$. Let $L$ denote the set of such vertices. It constitutes at least $1 - 2^{\ell + h - n}$ fraction of the vertices of the Grassmann graph and we ignore the vertices outside $L$ henceforth without affecting any of our analysis.

Define an equivalence relation $\sim$ on $L$ as follows: $L \sim L'$ if and only if $L \oplus H = L' \oplus H$. Let

$$L = C_1 \cup C_2 \cup \ldots \cup C_m$$

be the partition of $L$ into the corresponding equivalence classes and let $\text{Space}(C_i) \subseteq V$ denote the “common” subspace $L \oplus H$ over all $L \in C_i$. It is easily seen that all equivalence classes are of the same size. We call two equivalence classes $C_i, C_j, i \neq j$ “adjacent” if there are $L \in C_i, L' \in C_j$ such that $\dim(L \cap L') = \ell - 1$, i.e. if there is some edge of the Grassmann graph between $C_i, C_j$.

**Remark 3.5.12.** It is easy to see that for a fixed class $C_i$, a class $C_j$ adjacent to it can be chosen at random as follows: Pick an $(\ell - 1)$-dimensional subspace $D \subseteq \text{Space}(C_i)$ such that $D \cap H = \{0\}$. Let $L' \oplus H$ be a random $\ell$-dimensional subspace of (the global space) $V$ such that $D \subseteq L' \not\subseteq \text{Space}(C_i)$ and let $C_j$ be the equivalence class of $L'$.

Let $G_H$ denote the “augmented Grassmann graph” whose vertex set is $L = \cup_{i=1}^m C_i$ and whenever two equivalence classes $C_i, C_j$ are adjacent, in the augmented graph $G_H$, there is a complete bipartite graph between $C_i, C_j$. Our linearity test will have a constraint for every edge of the augmented graph. To motivate the test, we make a couple of observations:

- There are precisely $2^\ell$ linear functions on $\text{Space}(C_i)$ that respect the side condition. For $L \in C_i$, every linear function on $L$ (there are precisely $2^\ell$ of them) has a unique extension to $\text{Space}(C_i)$ that respects the side condition.

- If $C_i, C_j$ are adjacent then there are precisely $2^{\ell+1}$ linear functions on $\text{Space}(C_i) \oplus \text{Space}(C_j)$ that respect the side condition. Every linear function on $\text{Space}(C_i)$ (that respects the side condition) has exactly two extensions to $\text{Space}(C_i) \oplus \text{Space}(C_j)$ (that respect the side condition).

We denote by $[2^\ell]$ the set of all linear functions on an $\ell$-dimensional space. We are ready to describe the linearity test and the hypothesis describing its soundness guarantee. The input to the test is a table $F : L = \cup_{i=1}^m C_i \rightarrow [2^\ell]$ that assigns, to every $\ell$-space $L \in L$, a linear function on it. The intention is that there is a (global) linear function $f : V \to \{0, 1\}$ that respects the side condition and for every $L \in L$, $F[L] = f|_L$. The test is designed so that any “intended” assignment $F[\cdot]$ passes the test with probability 1.

**Grassmann Linearity Test with Side Condition**

- Pick a random pair $C_i, C_j$ of adjacent classes and a random $L \in C_i, L' \in C_j$.  

75
• Accept if and only if $F[L], F[L']$ are consistent, in the sense that there is a linear function on $\text{Space}(C_i) \oplus \text{Space}(C_j)$ that extends both and respects the side condition (if so, such a function is unique).

We emphasize that for every “answer” $F[L]$, it has a unique extension to $\text{Space}(C_i)$ followed by exactly two extensions to $\text{Space}(C_i) \oplus \text{Space}(C_j)$ (respecting the side condition), and hence exactly two “answers” to $F[L']$ for the test to accept. Thus this is a 2-to-2 test. While the test makes sense for any table $F[]$, we will need to consider only those tables that “respect the side condition”:

**Definition 3.5.13.** A table $F : \mathcal{L} = \bigcup_{i=1}^{m} C_i \rightarrow [2^{|C|}]$ respects side condition if for every $1 \leq i \leq m$, there is a linear function $g_i$ on $\text{Space}(C_i)$ that respects the side condition and for every $L \in C_i$, $F[L] = g_i|_L$.

We note that a table $F[\cdot]$ that respects side condition is consistent on every equivalence class $C_i$, but not necessarily consistent across different classes. The following theorem provides the soundness guarantee for the test.

**Theorem 3.5.14.** For every $\delta > 0$, there exist integers $r, q \geq 0$, a constant $C > 0$, and a function $\alpha(\cdot) > 0$ of an integer parameter such that for all sufficiently large integers $\ell$, for all sufficiently large integers $n$, the following holds.

Let $F[\cdot]$ be an assignment to the Grassmann graph $G(V, \ell)$, $\dim(V) = n$ that respects a side condition $(H, \psi_H)$ and passes the Grassmann linearity test with the side condition with probability at least $\delta$. Then for at least $\alpha(\ell)$ fraction of the $q$-dimensional subspaces $Q \subseteq V$, there exists a subspace $W \subseteq V$ of co-dimension $r$ that contains $Q \oplus H$, and a linear function $g_{Q,W} : W \rightarrow \{0, 1\}$ respecting the side condition such that

$$\Pr_{L \in G(V,\ell)} \left[ g_{Q,W}|_L \equiv F[L] \mid Q \subseteq L \subseteq W \right] \geq C. \tag{3.13}$$

The rest of the section is devoted to proving the above theorem. We show that if $F[\cdot]$ respects the side condition, then the probabilities that it passes the Grassmann Linearity Test with or without the side condition are essentially the same. Using Theorem 3.2.8, we get, for a good fraction of subspaces $Q$, a subspace $W$ and a linear function $g_{Q,W}$ on it that has a good agreement with $F[\cdot]$ on $\text{Zoom}[Q, W]$. Since $F[\cdot]$ respects side condition, we argue that $W$ can be assumed to contain $H$ and then that $g_{Q,W}$ respects the side condition. Towards a formal proof, we start with several preliminary observations.

• The test without the side condition is performed on the edges of the Grassmann graph $G$ whereas the test with the side condition is performed on the edges of the augmented Grassmann graph $G_H$.

• For classes $C_i, C_j, i \neq j$ that are non-adjacent, there are no edges in $G$ or in $G_H$ across these classes.

• For classes $C_i, C_j, i \neq j$ that are adjacent, there are some edges across them in $G$ (and their number is independent of the specific pair $C_i, C_j$) and a complete bipartite graph across them in $G_H$.

• If $F[\cdot]$ is an assignment that respects the side condition, then for any adjacent classes $C_i, C_j, i \neq j$, either all edges in $G$ as well as in $G_H$ across them are satisfied by $F[\cdot]$ or no edge in $G$ or $G_H$ across them is satisfied by $F[\cdot]$.

• Finally, the Grassmann graph $G$ may have some edges inside the classes $C_i$, but their number is negligible compared to the edges across distinct classes.

Theorem 3.5.14 follows quickly from these observations. Let $F[\cdot]$ be a table respecting the side condition that passes the Grassmann Linearity Test with the side condition with probability $\delta$. The observations above imply that $F[\cdot]$ passes the Grassmann Linearity Test (without the side condition) with essentially the same probability $\delta$. Applying Theorem 3.2.8, there exist parameters $r, q, C, \alpha(\cdot)$ as therein so that when $\ell$ and
then \( n \) are large enough, for \( \alpha(\ell) \) fraction of \( q \)-dimensional subspaces \( Q \subseteq V \), there is a subspace \( W \subseteq V \) containing \( Q \) and with co-dimension \( r \), and a linear function \( g_{Q,W} : W \to \{0, 1\} \) such that
\[
\Pr_{L \in G(V, \ell)} \left[ g_{Q,W}|_L \equiv F[L] \mid Q \subseteq L \subseteq W \right] \geq C. \tag{3.14}
\]

We may assume that \( Q \cap H = \{0\} \) as this is the case except with probability \( 2^{q+h-n} \ll \alpha(\ell) \). At this point, it would have been convenient if \( W \) already contained \( H \) and \( g_{Q,W} \) satisfied the side condition \((H, \psi_H)\). However, this is not automatically guaranteed by Theorem 3.2.8 which is not concerned with the issue of the side condition. Nevertheless, this is easily arranged as follows. Let \( H' = W \cap H \). We show first that \( g_{Q,W}|_{H'} \equiv \psi_H|_{H'} \), i.e. \( g_{Q,W} \) does respect the side condition as far as \( H' \), the “portion of \( H \) inside \( W \)”, is concerned. Assume by the way of contradiction that \( g_{Q,W}|_{H'} \neq \psi_H|_{H'} \). Partition the set of \( \ell \)-spaces \( L \) such that \( Q \subseteq L \subseteq W \) into classes (almost all of them satisfy \( L \cap H' = \{0\} \), so assume as such)
\[
C_1', \ldots, C_m',
\]
via the equivalence relation \( L \sim L' \) if and only if \( L \oplus H' = L' \oplus H' \). Let \( \text{Space}(C_i') \) be the common space \( L \oplus H' \) over \( L \in C_i' \). For any \( L, L' \in C_i' \), we have \( L \oplus H' = L' \oplus H' \), hence \( L \oplus H = L' \oplus H \), and hence the entire class \( C_i' \) is contained in some class \( C_s \) defined earlier. The assignment \( F[\cdot] \) over the class \( C_s \) is the restriction of some side condition respecting function \( g_s \) on \( \text{Space}(C_s) \supseteq \text{Space}(C_i') \). On their restriction to \( H' \), \( g_{Q,W} \) disagrees with \( \psi_H \) by the assumption and \( g_s \) agrees with \( \psi_H \) since it respects the side condition. Thus \( g_{Q,W} \) and \( g_s \) disagree on their restriction to \( \text{Space}(C_i') \supseteq H' \). Therefore, for a random \( L \in C_i' \), one could have the agreement \( g_{Q,W}|_L = g_s|_L = F[L] \) with probability at most \( 2^{q-\ell} \). Since this happens for every index \( i \) and \( \ell \) could have been chosen large enough, we get a contradiction to Equation (3.14).

It holds therefore that \( g_{Q,W}|_{H'} = \psi_H|_{H'} \). The same argument, as presented above, then shows that for each class \( C_i' \), the agreement between \( g_{Q,W} \) and \( F[\cdot] \) is either perfect or is negligible (i.e. at most \( 2^{q-\ell} \)). The two scenarios are determined by whether \( g_{Q,W} \) and \( g_s \) (where \( C_s \) is the class containing \( C_i' \)) agree on \( \text{Space}(C_i') \) or not. Now \( g_{Q,W} \) can be extended uniquely to \( \tilde{W} = W \oplus H \), call the new function \( g_{Q,\tilde{W}} \), so as to satisfy the “full” side condition \((H, \psi_H)\). Depending on whether \( g_{Q,W} \) and \( g_s \) agree on \( \text{Space}(C_i') \) or not, \( g_{Q,\tilde{W}} \) and \( g_s \) agree on \( \text{Space}(C_s) \) or not. Hence Equation (3.14) implies that (the same statement with \( W \) replaced by \( \tilde{W} \))
\[
\Pr_{L \in G(V, \ell)} \left[ g_{Q,\tilde{W}}|_L \equiv F[L] \mid Q \subseteq L \subseteq \tilde{W} \right] \geq \frac{C}{2}. \tag{3.15}
\]
To summarize, there is a subspace \( \tilde{W} \) containing \( Q \oplus H \) of co-dimension at most \( r \) and a linear function \( g_{Q,\tilde{W}} \) on it that respects the side condition so that Equation (3.15) is satisfied. This completes the proof of Theorem 3.5.14.

### 3.6 Soundness Analysis

In this section, we prove the soundness property of Theorem 1.2.6 and complete its proof (modulo Theorem 3.2.7). Starting with a \( \delta \)-consistent assignment \( F[\cdot] \), the proof first transforms it to a \( \delta \)-consistent assignment that is also clique consistent and then uses it to derive provers’ strategies for the Outer PCP Game from Section 3.3. Finally, it is shown that these strategies help the provers to win the Outer PCP Game with probability larger than its soundness error, thereby concluding there is no \( \delta \)-consistent assignment.

#### 3.6.1 Clique Consistent Assignment

Let \( F[\cdot] \) be an assignment to the game \( G_{\text{final}} \) that satisfies \( \delta \) fraction of the constraints.
Recall that the vertices of $G_{\text{final}}$ may be partitioned into $\text{Clique}_1 \cup \ldots \cup \text{Clique}_m$, and between any two cliques we either have a complete bipartite graph with equals weighted, or an empty graph.

**Lemma 3.6.1.** There exists a $\delta$-consistent assignment $F'$ that is clique consistent, i.e. for every $i$ assigns the vertices of $\text{Clique}_i$ consistent assignment according to the 1-to-1 constraints.

**Proof.** Consider the following distribution over clique consistent assignment. For each $\text{Clique}_i$, pick $L \in \text{Clique}_i$ randomly and define $F'[L']$ to be the assignment that $F[L]$ induces on $L'$ according to the 1-to-1 constraint between them, for every $L' \in \text{Clique}_i$. Consider the weight of constraints $F'$ satisfies expectedly. Denoting by $\delta(\text{Clique}_i, \text{Clique}_j)$ the weight of constraints satisfied by $F$ between $\text{Clique}_i$, $\text{Clique}_j$, we have that with probability $\delta(\text{Clique}_i, \text{Clique}_j)$, $F'$ satisfies all of the constraints between $\text{Clique}_i$, $\text{Clique}_j$. Therefore, the expected weight of constraints $F'$ satisfies is

$$\sum_{i,j} p_{i,j} \delta(\text{Clique}_i, \text{Clique}_j),$$

where $p_{i,j}$ is the probability the chosen constraint $(L, L')$ is between the $i$, $j$ cliques.

On the other hand, this expression is exactly the weight of constraints $F$ satisfies, since in the test once the cliques $i, j$ have been chosen, a random constraint between them is picked. In particular this expression is at least $\delta$, and $F'$ expectedly satisfies $\delta$ weight of the constraints. In particular, there exists $\delta$-consistent $F'$ (which is clique consistent).

By the above lemma, we assume henceforth that $F$ itself is clique-consistent. We remark that $F$ still respects the side conditions.

For each $U$, consider Block[$U$] and in particular $p(U)$ the fraction of constraints that are satisfied inside Block[$U$].

**Lemma 3.6.2.** $E_U [p(U)] \geq \delta$.

**Proof.** To sample a constraint in the 2-to-2 game, we first pick $U$ uniformly at random, then pick an edge $x = L \oplus H_U$, $y = L' \oplus H_U$ in it randomly and finally pick a $(x', y')$ random constraint between the cliques of $L \oplus H_U$ and $L' \oplus H_U$. Note that by the clique consistency and transitivity of the constraints (Definition 3.4.4), the constraint $(x, y)$ is satisfied by $F$ if and only if the constraint $(x', y')$ is satisfied. Therefore,

$$E_U [p(U)] = \Pr_{U,L,L'} [F[x], F[y] \text{ consistent}] = \Pr_{U,L,L',x',y'} [F[x'], F[y'] \text{ consistent}] \geq \delta.$$

Finally, we recall that in the Outer PCP Game the first prover (the “larger” prover) receives as a question, a set $U$ of $3k$ variables and as an advice, a $q$-dimensional subspace $Q \subseteq \{0, 1\}^U$ (in fact $Q \subseteq \{0, 1\}^V$). The second prover (the “smaller” prover) receives as a question a subset $V \subseteq U$ of variables of expected size $3k - 2\beta k$ and the same advice $Q \subseteq \{0, 1\}^V$.

### 3.6.2 A Strategy for the First (Larger) Prover

As before, let $p(U)$ denote the fraction of constraints satisfied after picking $U$ so that $E_U [p(U)] \geq \delta$. By an averaging argument, for at least $\frac{\delta}{2}$ fraction of the tuples $U$, we have $p(U) \geq \frac{\delta}{2}$. Call such a tuple $U$ good, and let $U_{\text{good}}$ be the set of good tuples, with $|U_{\text{good}}| \geq \frac{\delta}{2} \cdot |U|$.

Let the question to the first prover be $U \in U$ along with the advice $Q$. If $U \notin U_{\text{good}}$, the prover gives up, so let’s assume henceforth that $U \in U_{\text{good}}$. As observed, $p(U)$ is precisely the probability that the assignment $F[\cdot]$ on Block[$U$] passes the Grassmann linearity test with side condition $H_U$. Since $p(U) \geq \frac{\delta}{2}$, by Theorem
3.5.14 there are parameters \(q, r, C, \alpha(\cdot)\) such that, for at least \(\alpha(\ell)\) fraction of the \(q\)-dimensional subspaces \(Q \subseteq \{0, 1\}^U\), there exists \(W \subseteq \{0, 1\}^U\) of co-dimension \(r\) that contains \(Q \oplus H_U\) and a global linear function 
\[ g_{Q,W} : W \to \{0, 1\} \]
that respects the side condition \(H_U\) and\(^{18}\)
\[ \Pr_L \left[ g_{Q,W} | L \equiv F[L] \mid Q \subseteq L \subseteq W \right] \geq C. \] (3.16)

We call such a choice of \(Q\) “lucky” and let \(Q_{\text{lucky}}\) be the set of all lucky \(q\)-dimensional subspaces of \(\{0, 1\}^U\). Moreover, call a \(q\)-dimensional subspace \(Q\) “smooth” if it satisfies Equation (3.9) in Lemma 3.5.3, and let \(Q_{\text{smooth}}\) be the set of all smooth \(q\)-dimensional subspaces of \(\{0, 1\}^U\). If any of the three conditions \(Q \in Q_{\text{lucky}}, Q \in Q_{\text{smooth}}, Q \cap H_U = \{0\}\), does not happen, the prover gives up, so assume that all three conditions hold and fix \(Q\) henceforth.

Ideally, the prover would want to use the function \(g_{Q,W}\) to output her answer. However, as explained in Section 3.5.2, the number of zoom-outs \(W\) of co-dimension \(r\) that are eligible (i.e. satisfy Equation (3.16)) could be large. As proposed therein, to remedy this situation we consider the least co-dimension in which a successful zoom-out exists.\(^ {19}\)

Accordingly, let \(C_r = C\) and \(C_i = 10^{-9} \left(10^{-9}(\frac{c+1}{1})^{12}\right)\) for \(i = r - 1, \ldots, 0\). Let \(r \geq r^* \geq 0\) be the least integer such that there exists a subspace \(W^* \subseteq \{0, 1\}^U\) of co-dimension \(r^*\) that contains \(Q \oplus H_U\), and a linear function \(g_{Q,W^*} : W^* \to \{0, 1\}\) that respects the side condition \(H_U\) such that
\[ \Pr_L \left[ g_{Q,W^*} | L \equiv F[L] \mid Q \subseteq L \subseteq W^* \right] \geq C_{r^*}. \] (3.17)

By the conclusion preceding Equation (3.16), such a choice of \(r^*, W^*, g_{Q,W^*}\) exists. The prover picks one such \(W^*, g_{Q,W^*}\), extends \(g_{Q,W^*}\) arbitrarily to the whole space \(\{0, 1\}^U\), considers the extended function as an inner product function \(x \to \langle s_{Q,U}, x \rangle\) for a string \(s_{Q,U} \in \{0, 1\}^U\) and outputs \(s_{Q,U}\) as her answer. Since \(H_U \subseteq W^*\) and \(g_{Q,W^*}\) respects the side condition, it is guaranteed that \(s_{Q,U}\) satisfies the \(k\) linear equations that define the tuple \(U\).

3.6.3 A Strategy for the Second (Smaller) Prover

Let the question to the second prover be \(V\). The prover derives his strategy from a table \(\bar{F}[:]\) that assigns, to every \(\ell\)-dimensional subspace \(L \subseteq V\), a linear function \(\bar{F}[L]\) on it. The table \(\bar{F}[:]\) is really the same as the given table \(F[:]\), i.e. the given assignment to the game \(G_{\text{final}}\). We do make a distinction between the two tables to clarify one important point. Formally, the table \(F[:]\) is defined on all vertices
\[ L \oplus H_U, \ U \in \mathcal{U}, \ L \subseteq \{0, 1\}^U, \ \dim(L) = \ell. \]

For a given question \(V\) to the second prover, let \(U \supseteq V\) be an arbitrary question to the first prover. For all \(\ell\)-dimensional subspaces \(L \subseteq \{0, 1\}^V\), define
\[ \bar{F}[L] = F[L \oplus H_U]|_L. \]

Here we regard \(L\) as a subspace of \(\{0, 1\}^U\) and then the term \(F[L \oplus H_U]\) makes sense. The key point here is that this definition is independent of the choice of the question \(U \supseteq V\) and hence unambiguous! The reason is that for a given \(L \subseteq \{0, 1\}^V\), the vertices \(L \oplus H_U\) over all \(U \supseteq V\) are in the same clique of the game

\(^{18}\)Strictly speaking, one has to write \(F[L \oplus H_U]\) instead of \(F[L]\) in the inequality; we chose the latter for simplicity.

\(^{19}\)The point of the upper bound on the number of successful zoom-outs is not used here and will only come into play when analyzing the coordination between the provers’ strategies.
and a linear function. The uniform distribution on \( \Omega \) is consistent on cliques, \( F[\Omega \oplus H_V] \) is the same over all \( U \supseteq V \). We are therefore justified to replace the notation \( \bar{F}[] \) by \( F[] \) henceforth.

We now describe the second prover’s strategy on question \( V \) and advice \( Q \). It is useful to think of the question to the first prover \( U \in \mathcal{U}_{\text{good}} \) and the advice \( Q \in \Omega_{\text{lucky}} \cap Q_{\text{smooth}} \) as being fixed and the question to the second prover as a random question \( V \subseteq U \) along with the (same) advice \( Q \). We note that the first prover has the integer \( r^* \) already “in her mind” as in the previous section. By picking a random integer in \( \{0, 1, \ldots, r\} \), we may assume that the second prover also “knows” the integer \( r^* \) (he may sample it, and happens to choose the correct one with probability \( \frac{1}{r^*} \)).

Let \( H_V \) denote the side condition on \( V \), defined by the equations that are contained in \( V \). The second prover chooses, if possible, a subspace \( \mathcal{W}_{\text{second}} \subseteq \{0, 1\}^V \) of co-dimension \( r^* \) containing \( Q \oplus H_V \) and a linear function \( g_{Q, \mathcal{W}_{\text{second}}} : \mathcal{W}_{\text{second}} \to \{0, 1\} \) that respects the side condition \( H_V \) and such that

\[
\Pr_L \left[ h_{Q, \mathcal{W}_{\text{second}}} \mid L \equiv F[L] \mid Q \subseteq L \subseteq \mathcal{W}_{\text{second}} \right] \geq \frac{C_{r^*}}{4}.
\]

If no \( \mathcal{W}_{\text{second}} \), \( g_{Q, \mathcal{W}_{\text{second}}} \) is “eligible”, the prover gives up. If several \( \mathcal{W}_{\text{second}} \), \( g_{Q, \mathcal{W}_{\text{second}}} \) are “eligible”, then one of them is picked at random (we will prove an upper bound on their number). The prover then extends \( g_{Q, \mathcal{W}_{\text{second}}} \) randomly to the whole space \( \{0, 1\}^V \) (there are only \( 2^{r^*} \) choices as \( \mathcal{W}_{\text{second}} \) has co-dimension \( r^* \)), considers the extended function as an inner product function \( y \to \langle s_{Q, V}, y \rangle \) for a string \( s_{Q, V} \in \{0, 1\}^V \) and outputs \( s_{Q, V} \) as his answer.

### 3.6.4 The Success Probability of the Provers

We now analyze the success probability of the provers. We assume that \( U \in \mathcal{U}_{\text{good}} \) which happens with probability \( \frac{\delta}{2} \). We also assume that \( Q \in \Omega_{\text{lucky}} \cap Q_{\text{smooth}} \) which happens with probability \( \alpha(\ell) \): by Theorem 3.5.14 we have \( Q \in \Omega_{\text{lucky}} \) with probability \( \geq \alpha(\ell) \), and by Lemma 3.5.3 \( Q \) is smooth except with probability \( \sqrt{\beta k^\frac{1}{2}} \). Thus it is sufficient to choose \( \beta, k \), as we indeed do, such that \( \sqrt{\beta k^\frac{1}{2}} \leq \frac{\alpha(\ell)}{2} \). The first prover has the integer \( r^* \) “in her mind” and answers according to the subspace \( W^* \) that contains \( Q \oplus H_U \) and a linear function \( g_{Q, W^*} \) on \( W^* \) that respects the side condition \( H_U \) and satisfies Equation (3.17). We think of the function \( g_{Q, W^*} \) itself as her answer; extending it to the whole space \( \{0, 1\}^U \) and considering it as an inner product function to arrive at the actual answer, are mere technicality. We now consider a random question \( V \subseteq U \) to the second prover along with the advice \( Q \). To achieve a good success probability, we need to show that

- Given the answer \( g_{Q, W^*} \) of the first prover, the second prover is likely to have a possible answer \( h_{Q, \mathcal{W}_{\text{second}}} \) that is consistent with it, i.e., such that \( \mathcal{W}_{\text{second}} \subseteq W^* \) and the restriction of \( g_{Q, W^*} \) to \( \mathcal{W}_{\text{second}} \) coincides with \( h_{Q, \mathcal{W}_{\text{second}}} \).

- The second prover is likely to have a reasonable upper bound on the number of his eligible answers (and the answer \( h_{Q, \mathcal{W}_{\text{second}}} \) that is consistent with the first prover’s answer, as in the preceding item, appears on this short list).

The first item follows immediately from the smoothness property. We recall Equation (3.17),

\[
\Pr_L \left[ g_{Q, W^*} \mid L \equiv F[L] \mid Q \subseteq L \subseteq W^* \right] \geq C_{r^*}.
\]

Since \( Q \) is smooth and \( W^* \) has co-dimension \( r^* \) inside \( \{0, 1\}^U \), by Lemmas 3.5.3, 3.5.4, we can replace the uniform distribution on \( \ell \)-spaces \( L \) such that \( Q \subseteq L \subseteq W^* \), up to a tiny statistical distance, by the
distribution that first samples a random question \( V \subseteq U \) and then samples uniformly an \( \ell \)-space \( L \) such that \( Q \subseteq L \subseteq W^* \cap \{0,1\}^V \). Thus the equation above implies that

\[
\mathbb{E}_V \left[ \Pr \left[ g_{Q,W^*}|L \equiv F[L] \mid Q \subseteq L \subseteq W_{\text{second}} \right] \right] \geq \frac{C_{r^*}}{2}.
\]

Hence for at least \( \frac{C_{r^*}}{4} \) fraction of the questions \( V \) to the second prover (let \( \mathcal{V}_{\text{consistent}} \) denote the set of such questions), we have

\[
\Pr \left[ g_{Q,W^*}|L \equiv F[L] \mid Q \subseteq L \subseteq W_{\text{second}} \right] \geq \frac{C_{r^*}}{4}.
\]

Thus \( W_{\text{second}}, h_{Q,W^*} \) is an “eligible” answer for the second prover that is, moreover, consistent with the first prover’s answer \( W^*, g_{Q,W^*} \). This happens whenever \( V \in \mathcal{V}_{\text{consistent}} \). By Lemma 3.6.3, we may also assume that the co-dimension of \( W_{\text{second}} \) inside \( \{0,1\}^V \) remains \( r^* \).

The second item requires more effort. We call a question \( V \) to the second prover “bad” (let \( \mathcal{B} \) denote the set of such questions), if he has too many, i.e. \( \geq 2^{8r^2 \ell} \), eligible zoom-outs. Assume by the way of contradiction that the fraction \( \eta = \Pr_V [V \in \mathcal{B}] \) of the bad set is non-negligible. We reach a contradiction by showing that then the first prover would have a successful zoom-out of one less co-dimension, contradicting the minimality of \( r^* \).

Specifically, suppose that for a question \( V \in \mathcal{B} \) to the second prover, there are subspaces \( W_1, \ldots, W_N \subseteq \{0,1\}^V, N = 2^{8r^2 \ell} \), of co-dimension \( r^* \), all containing \( Q \oplus H_V \), and linear functions \( h_{Q,W_1}, \ldots, h_{Q,W_N} \) on them, all respecting the side condition \( H_V \) and all having agreement with \( F[\cdot] \) on at least \( \frac{C_{r^*}}{4} \) fraction of \( \ell \)-spaces \( L \in \text{Zoom}(Q,W_i) \). By Lemma 3.5.7, there is a subspace \( W[V] \subseteq \{0,1\}^V \) such that

- \( W[V] \) has co-dimension \( s \leq r^* - 1 \) inside \( \{0,1\}^V \). Different \( V \in \mathcal{B} \) may lead to different values of \( s \), but by choosing the most popular value, and possibly losing a factor \( r \) in \( \eta = \Pr_V [V \in \mathcal{B}] \), we assume henceforth that all \( V \in \mathcal{B} \) lead to the same value \( s \).

- \( W[V] \) contains at least one \( W_{i_0} \) and hence contains \( Q \oplus H_V \).

- There is a linear function \( h_{Q,W[V]} \) on \( W[V] \) that has agreement with \( F[\cdot] \) on at least \( \theta = 10^{-9} \left( \frac{C_{r^*}}{4} \right)^{12} \) fraction of \( \ell \)-spaces \( L \in \text{Zoom}(Q,W[V]) \).

- \( h_{Q,W[V]} \) agrees with \( h_{Q,W_{i_0}} \) on \( W_{i_0} \) and since \( H_V \subseteq W_{i_0} \) and \( h_{Q,W_{i_0}} \) respects the side condition \( H_V \), so does \( h_{Q,W[V]} \).

We are almost at the end of the argument. Consider the collection

\[
\{W[V], \ \text{codim}(W[V]) = s, h_{Q,W[V]} \mid V \in \mathcal{B}\}, \quad \eta = \Pr_V [V \in \mathcal{B}].
\]

Applying Lemma 3.6.4, there is a sub-collection \( V_1, \ldots, V_m \in \mathcal{B} \) and integer \( \frac{1}{2} \beta k \leq t \leq 2 \beta k \), such that

- \( |V_1| = \ldots = |V_m| = 3k - 2t \).
- \( m \geq \frac{\eta}{2^{8(2t+s)\ell}} \geq 2^{8(2t+s)\ell} \). The second inequality is where we used the assumption that \( \eta \) is non-negligible.
- For \( 1 \leq i \neq j \leq m \), \( \text{dim}(W[V_i] \oplus W[V_j]) \geq 3k - s \) and \( \text{Eq}[V_i] \cup \text{Eq}[V_j] = \text{Eq}[U] \).

Now regard the subspaces \( W[V_i], 1 \leq i \leq m \), as subspaces of \( \{0,1\}^U \) of co-dimension \( 2t + s \). Applying Lemma 3.5.7, there is a subspace \( P \subseteq \{0,1\}^U \) that contains at least two \( W[V_{i_0}], W[V_{j_0}] \), and a linear function \( g_{Q,P} \) on \( P \) that is
• consistent with \( h_{Q,W[V_{i_0}]} \), \( h_{Q,W[V_{j_0}]} \) on \( W[V_{i_0}], W[V_{j_0}] \) respectively, and

• agrees with \( F[\cdot] \) on at least \( 10^{-9} \theta^{12} = C_{r^{-1}} \geq C_s \) fraction of \( \ell \)-spaces \( Q \subseteq L \subseteq P \).

We further note that

• \( \dim(P) \geq \dim(W[V_{i_0}] \oplus W[V_{j_0}]) \geq 3k - s \) and hence \( P \) has co-dimension at most \( s \) inside \( \{0, 1\}^U \).

• \( g_{Q,P} \) is consistent with \( h_{Q,W[V_{i_0}]} \), \( h_{Q,W[V_{j_0}]} \) on \( W[V_{i_0}], W[V_{j_0}] \) respectively, which respect side conditions \( H_{V_{i_0}}, H_{V_{j_0}} \) respectively, and hence \( g_{Q,P} \) respects the side condition \( H_{V_{i_0}} \oplus H_{V_{j_0}} = H_U \), where the last claim holds because \( \text{Eq}[V_{i_0}] \cup \text{Eq}[V_{j_0}] = \text{Eq}[U] \).

In short, \( P \) is a successful zoom-out of co-dimension \( s \leq r^* - 1 \) for the first prover, in the sense described in/before Equation (3.17), contradicting the minimality of \( r^* \).

This shows that the size \( \eta \) of the bad set \( B \) is negligible for appropriate choice of parameters. More precisely, choosing \( \beta = \frac{\log \log k}{k} \), \( t = \Theta(\beta k) = \Theta(\log \log k) \), the argument above gives an upper bound

\[
\eta \leq 2^{O(s)} \cdot \beta^3 k^2 \cdot 2^{O((2t+s)^2 \ell)} \leq \frac{1}{k} \cdot (\log \log k)^3 \cdot 2^{O((\log \log k)^2 \ell)} \leq \frac{C_{r^*}}{8} \leq \frac{1}{2} \cdot \Pr[V \in V_{\text{consistent}}],
\]

by choosing \( k \) large enough and using the lower bound on \( \Pr_V[V \in V_{\text{consistent}}] \) noted earlier. Therefore, on at least half of the questions in \( V_{\text{consistent}} \), the second prover has at least one answer that is eligible and consistent with the first prover’s answer, and at most \( 2^{-s^2 \ell} \) eligible zoom-outs. Therefore, the second prover chooses the correct zoom-out with probability at least \( 2^{-8s^2 \ell} \). Choosing his answer uniformly among the functions that have at least \( \frac{C_{r^*}}{k} \) consistency with \( F \) on \( Q \subseteq L \subseteq W_{\text{second}} \), he chooses the correct function with probability at least \( \frac{1}{m} \) where \( m \) is the total number of functions on \( W_{\text{second}} \) that have such consistency, which by Lemma 3.5.9 is at most \( \frac{C_{r^*}^2}{2^{s^2 \ell + q - \epsilon}} \).

Finally to be consistent with the first prover, the second prover must extend his function \( h_{Q,W_{\text{second}}} \) to \( \{0, 1\}^V \) in a consistent manner to the first prover’s function. This event occurs with probability at least \( 2^{-r} \), as the co-dimension of \( W_{\text{second}} \) inside \( \{0, 1\}^V \) is at most \( r \).

The overall success probability of the provers is at least

\[
\frac{\delta}{2} \cdot \frac{\alpha(\ell)}{2} \cdot \frac{C_{r^*}}{8} \cdot 2^{s^2 \ell} \cdot \frac{C_{r^*}^2}{16} 2^{4+q-\ell} 2^{-r}.
\]

3.6.5 Auxiliary Lemmas

The following lemma, not strictly necessary for our argument, does make the argument cleaner and its proof serves as a warm-up for the next lemma.

Lemma 3.6.3. Let \( U \) be a fixed question to the first prover in the Outer PCP, i.e. a set of \( 3k \) variables in some set of \( k \) equations. Let \( V \subseteq U \) be a question to the second prover chosen at random as in the Outer PCP. Let \( W \subseteq \{0, 1\}^U \) be a subspace of co-dimension \( s \). Then except with probability \( 2^{s^2 + 3\beta^2 k} \) over the choice of the question \( V \),

\[ \dim(W \cap \{0, 1\}^V) = |V| - s, \]

i.e. the co-dimension of \( W \cap \{0, 1\}^V \) inside \( \{0, 1\}^V \) remains \( s \).

Proof. We note first that except with probability \( 2^{-\Omega(\beta k)} \), we have \( |V| = 3k - b \) with \( b \leq 4\beta k \), and we assume this to be the case henceforth. Suppose \( W \subseteq \{0, 1\}^U \) is defined by linear conditions

\[ \langle u_1, x \rangle = 0, \ldots, \langle u_s, x \rangle = 0, \]

82
for some \( u_1, \ldots, u_s \in \{0, 1\}^U \). Denote by \( v_j \in \{0, 1\}^U \), the vector that has 1 in coordinate \( j \) and 0 elsewhere. Suppose \( \{0, 1\}^V \) regarded as a subspace of \( \{0, 1\}^U \) is defined by linear conditions

\[
\langle v_{i_1}, x \rangle = 0, \ldots, \langle v_{i_b}, x \rangle = 0,
\]

where \( U \setminus V = \{i_1, \ldots, i_b\} \). The event we are interested in is precisely the event that \( \{u_1, \ldots, u_s, v_{i_1}, \ldots, v_{i_b}\} \) are linearly independent. A linear dependency could arise only if some non-trivial linear combination \( u \) of \( \{u_1, \ldots, u_s\} \) has Hamming weight at most \( b \) and the non-zero coordinates of \( u \) are contained in \( \{i_1, \ldots, i_b\} \). Let \( S \) be the union of all non-zero coordinates of all \( u \) that are linear combinations of \( \{u_1, \ldots, u_s\} \) and have Hamming weight at most \( b \). Thus \( |S| \leq 2^s b \leq 2^s \cdot 4\beta k \). It follows that while choosing \( V \subseteq U \), if \( (U \setminus V) \cap S = \phi \), then the desired event holds, and this happens with probability at least \( 1 - \beta \cdot 2^s \cdot 4\beta k \). 

For a question \( U \) or \( V \) to the first or the second prover respectively, let \( \text{Eq}[U] \) or \( \text{Eq}[V] \) denote the set of equations contained in \( U \) or \( V \) respectively. The above lemma asserts, roughly speaking, that for any question \( U \) to the first prover and any small co-dimension space \( W \), the co-dimension of \( W \cap \{0, 1\}^V \) is preserved with high probability in \( V \) over the choice of \( V \) in the Outer PCP. The lemma below is a reverse type statement, stating that in any \( \{(V, W[V])\}_{V \in B} \) a sizable collection of questions to the second prover \( V \subseteq U \) and co-dimension \( s \) subspace \( W[V] \subseteq V \), there is a sub-collection \( B' \) in which for every two distinct \( V_1, V_2 \in B' \), the codimension of \( W[V_1] \cap W[V_2] \) inside \( U \) is at most \( s \).

**Lemma 3.6.4.** Let \( U \) be a fixed question in the Outer PCP, i.e. a set of \( 3k \) variables in some set of \( k \) equations. Let \( V \subseteq U \) be a question to the second prover chosen at random as in the Outer PCP. Given that the question to the first prover is \( U \), let \( B \) be some subset of the set of all questions to the second prover and let \( \eta = \Pr[V \in B] \). Suppose that for every \( V \in B \), a subspace \( W[V] \subseteq \{0, 1\}^V \) is given with co-dimension \( s \). Then there exist \( V_1, \ldots, V_m \in B \) such that

- \( m \geq \frac{\eta}{2^{2s + 2s + 3\beta k}} \).
- For some \( \frac{1}{2} \beta k \leq t \leq 2\beta k \), we have \( |V_1| = \ldots = |V_m| = 3k - 2t \).
- For \( 1 \leq i \neq j \leq m \), \( \dim(W[V_i] \cap W[V_j]) \geq 3k - s \) and \( \text{Eq}[V_i] \cup \text{Eq}[V_j] = \text{Eq}[U] \).

**Proof.** We find the desired \( V_1, \ldots, V_m \) iteratively. Having found \( V_1, \ldots, V_{m-1} \), as long as \( m \leq \frac{\eta}{2^{2s + 2s + 3\beta k}} \), we show that a randomly chosen \( V \subseteq U \) can be added as \( V = V_{m+1} \) with good probability.

We first note that the expected size of \( V \) is \( 3k - 2\beta k \) and except with probability \( 2^{-\Omega(\beta k)} \), the size of \( V \) is in the range \( [3k - 4\beta k, 3k - 3\beta k] \). The number of possible sizes in this range is \( 3\beta k \). Let the most popular size of \( V \in B \) be \( 3k - 2t \) for some \( \frac{1}{2} \beta k \leq t \leq 2\beta k \). We restrict \( B \) to those \( V \) that have size exactly \( 3k - 2t \), possibly losing a factor \( 3\beta k \) in its “size” \( \eta \). Let \( B' \) denote this restricted set and \( \eta' \geq \frac{\eta}{3\beta k} \) be its fractional size. Fix any \( V^* \in B' \) and consider the process of picking a random \( V \subseteq U \). We consider three events below, obtain bounds on their probability, and then argue that all three events happen with a good probability.

- \( \Pr[V \in B^*] = \eta^* \).
- \( \Pr[\text{Eq}[V^*] \cup \text{Eq}[V] = \text{Eq}[U]] \geq 1 - t\beta \).

The event under consideration holds except when \( V \) happens to include a single variable from some equation among the precisely \( t \) equations from which \( V^* \) includes a single variable. The probability of this happening for one given equation is \( \beta \) and a union bound is taken.
• Pr \[
\dim(W[V^*] \cap \{0, 1\}^V) = \dim(W[V^*]) - (3k - |V|) \geq 1 - 2^{s+\beta}k - 2^{-\Omega(\beta k)}.
\]

Here the argument is a bit subtle. Except with probability $2^{-\Omega(\beta k)}$, we have $|V| = 3k - b$ with $b \leq 4\beta k$. Denote by $v_j \in \{0, 1\}^U$, the vector that has 1 in coordinate $j$ and 0 elsewhere. Suppose that $W[V^*]$ considered as a subspace of $\{0, 1\}^U$ is defined by linear conditions

$$\langle v_j, x \rangle = 0, \ldots, \langle v_{j2t}, x \rangle = 0, \quad \langle u_1, x \rangle = 0, \ldots, \langle u_s, x \rangle = 0,$$

where $U \setminus V = \{j_1, \ldots, j_{2t}\}$ and $u_1, \ldots, u_s \in \{0, 1\}^{V^*}$. On the other hand, $\{0, 1\}^V$ considered as a subspace of $\{0, 1\}^U$ is defined by linear conditions

$$\langle v_i, x \rangle = 0, \ldots, \langle v_b, x \rangle = 0,$$

where $U \setminus V = \{i_1, \ldots, i_b\}$. The event we are interested in is precisely the event that the vectors $\{v_{j1}, \ldots, v_{j2t}, u_1, \ldots, u_s, v_{i1}, \ldots, v_{ib}\}$ are linearly independent. A linear dependency could arise only if either (1) $\{j_1, \ldots, j_{2t}\} \cap \{i_1, \ldots, i_b\} \neq \emptyset$ or (2) if some non-trivial linear combination $u$ of $\{u_1, \ldots, u_s\}$ has Hamming weight at most $2t + b$, and furthermore the non-zero coordinates of $u$ are contained in $\{j_1, \ldots, j_{2t}, i_1, \ldots, i_b\}$.

Let $\mathcal{S}$ be the set of coordinates containing $\{j_1, \ldots, j_{2t}\}$ and all non-zero coordinates of vectors $u$ that are linear combinations of $\{u_1, \ldots, u_s\}$ having Hamming weight at most $2t + b$. Note that if (1) or (2) holds, then $\mathcal{S} \cap (U \setminus V) \neq \emptyset$, so it suffices to upper bound the probability this happens. Clearly, $\mathcal{S}$ is a fixed set (depending only on $U$ and $V^*$ and $|\mathcal{S}| \leq 2t + 2s(2t + b) \leq 2^{s+4} \cdot \beta k$. It follows that while choosing $V \subseteq U$, the probability $\sequent{U \setminus V} \cap \mathcal{S} \neq \emptyset$ is at most $|\mathcal{S}| \cdot \frac{b}{k} \leq 2^{s+6} \beta^2 k$.

It follows that all three events hold with probability at least $\eta^* - 2^{s+7} \beta^2 k$. If so, we note further that $|V| = 3k - 2t$ and

$$\dim(W[W^*] \oplus W[V]) = \dim(W[V^*]) + \dim(W[V]) - \dim(W[V^*] \cap W[V]) \geq \dim(W[V^*]) + \dim(W[V]) - \dim(W[V^*] \cap \{0, 1\}^V) = 2 \cdot (3k - 2t - s) - (\dim(W[V^*]) - (3k - |V|)) = 2 \cdot (3k - 2t - s) - ((3k - 2t - s) - 2t) = 3k - s.$$

The lemma now follows immediately. We find the desired $V_1, \ldots, V_m$ iteratively. Having found $V_1, \ldots, V_{m^*}$, we choose $V \subseteq U$ at random. The argument above shows, by letting $V_1, \ldots, V_{m^*}$ play the role of $V^*$ and taking a union bound, that $V$ can be added as $V = V_{m^*+1}$ with probability $\eta^* - m^* \cdot 2^{s+7} \beta^2 k$. This works as long as $m^* \leq \frac{\eta^*}{2^{s+7} \beta^2 k} \leq \frac{\eta}{2^{s+7} \beta^2 k}$. \qed
The $\ell$-space vs $b$-space Linearity Test

For $\Omega = \{-1, 1\}^n$, let $B$ and $L$ denote the set of all $b$-dimensional and $\ell$-dimensional subspaces of $\Omega$. Let $A, F$ be tables that assign linear functions to $B \in B$ and $L \in L$ respectively, linear functions $A[B] : B \to \{-1, 1\}$, $F[L] : L \to \{-1, 1\}$ on the respective subspaces. The test picks a pair $(B, L)$ uniformly at random with $B \subseteq L$, $B \in B$, $L \in L$ and accepts if and only if

$$F[L]|_B \equiv A[B].$$

Our result is the following:

**Theorem 3.7.1.** Let $\Omega, B, L$ and parameters $n, \ell, 1 < b < \frac{\ell}{4}$ be as in the description of the test above. Let $A, F$ be tables that assign linear functions to $B \in B$ and $L \in L$ respectively. Suppose the tables pass the linearity test with probability at least $\frac{1}{2^b} + \varepsilon$ where $\varepsilon \geq 2^{2^{-b/4}}$, i.e.

$$\Pr_{B \subseteq L, B \in B, L \in L} [F[L]|_B \equiv A[B]] \geq \frac{1}{2^b} + \varepsilon.$$

Then there exists a global linear function $g : \Omega \to \{-1, 1\}$ that agrees with at least $\frac{\varepsilon^3}{300}$ fraction of the $\ell$-spaces, that is

$$\Pr_{L \in L} [F[L] \equiv g|L] \geq \frac{\varepsilon^3}{300}.$$

The rest of the section is devoted to proving Theorem 3.7.1.

We start by viewing the entire table $A[\cdot]$ as a function $f : \{-1, 1\}^n \to \{-1, 1\}^b$ as follows. In notation, for $(v_1, \ldots, v_b) \in \{-1, 1\}^n = \{-1, 1\}^n$,

$$f(v_1, \ldots, v_b) = (A[\text{Span}(v_1, \ldots, v_b)](v_1), \ldots, A[\text{Span}(v_1, \ldots, v_b)](v_b)).$$

In words, to evaluate $f(v_1, \ldots, v_b)$, one considers the $b$-space $B = \text{Span}(v_1, \ldots, v_b)$, and the linear function $A[B]$ on $B$. The linear function assigns, in particular, $\{-1, 1\}$-values to the vectors $v_1, \ldots, v_b$ respectively. The list of these $b$ values is defined to be $f(v_1, \ldots, v_b)$. Since the output of $f$ is a string of length $b$, we can think of $f$ as a collection of $\{-1, 1\}$-valued functions, $f_1, \ldots, f_b$, one for each output coordinate. In notation, $f_i : \{-1, 1\}^n \to \{-1, 1\}$ is defined as

$$f_i(v_1, \ldots, v_b) = A[\text{Span}(v_1, \ldots, v_b)](v_i).$$

We must make a couple of clarifying remarks. First, when the input vectors $\{v_1, \ldots, v_b\}$ are linearly dependent, then their span $B$ has dimension less than $b$ and $A[B]$ is undefined. However the fraction of such inputs is negligible (at most $2^{b-n}$) and on those inputs $f$ can be defined arbitrarily without affecting the analysis.
The Gowers Test

The main idea behind the analysis of the above linearity test is to use a “Gowers Test” as an auxiliary tool. We can relate the acceptance probability of the $\ell$-space vs $b$-space test to that of the acceptance probability of the Gowers test. The Gowers test allows us to conveniently switch from local considerations to global considerations.

Let $1^b$ denote the $b$-dimensional vector with all coordinates 1.

**Definition 3.7.2. [Gowers Test]** Given $h : \{-1, 1\}^{nb} \rightarrow \{-1, 1\}^b$, pick $x, y, z \in \{-1, 1\}^{nb}$ randomly and check if

$$h(x)h(y)h(z)h(x \cdot y \cdot z) = 1^b.$$  

For $u, v \in \{-1, 1\}^b$, the notation $uv$ stands for the coordinatewise product of $u, v$.

Represent a function $h : \{-1, 1\}^{nb} \rightarrow \{-1, 1\}^b$ as $h = (h_1, \ldots, h_b)$ where $h_i$ are the coordinate-wise functions. For $T \subseteq [b]$, let $h_T = \prod_{i \in T} h_i$ be the product functions. The lemma below expresses the probability of $h$ passing the Gowers test in terms of the Fourier coefficients of products of functions $h_T$.

**Lemma 3.7.3.** The probability that $h : \{-1, 1\}^{nb} \rightarrow \{-1, 1\}^b$ passes the Gowers Test is:

$$\Pr_{x, y, z \in \{-1, 1\}^{nb}} \left[ h(x)h(y)h(z)h(x \cdot y \cdot z) = 1^b \right] = \frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \sum_{S \subseteq [nb]} \widehat{h}_T(S).$$

**Proof.** For the test to pass, it must pass on every coordinate. Thus,

$$\Pr_{x, y, z \in \{-1, 1\}^{nb}} \left[ h(x)h(y)h(z)h(x \cdot y \cdot z) = 1^b \right] = \mathbb{E}_{x, y, z \in \{-1, 1\}^{nb}} \left[ \prod_{i=1}^b \frac{1 + h_i(x)h_i(y)h_i(z)h_i(x \cdot y \cdot z)}{2} \right]$$

$$= \frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \mathbb{E}_{x, y, z \in \{-1, 1\}^{nb}} \left[ h_T(x)h_T(y)h_T(z)h_T(x \cdot y \cdot z) \right]$$

$$= \frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \sum_{S \subseteq [nb]} \widehat{h}_T(S).$$

The main trick is that Lemma 3.7.3 is applied globally as well as locally and then the information gained from the two applications is combined. Globally, the lemma is applied to the function $f : \{-1, 1\}^{nb} \rightarrow \{-1, 1\}^b$ that (essentially) represents the entire assignment $\{A[B] | B \in \mathcal{B}\}$. Locally, for a fixed $\ell$-space $L$, the lemma is applied to the function $g : \{-1, 1\}^b \rightarrow \{-1, 1\}^b$ that represents, in a similar manner, the assignment $\{A[B] | B \subseteq L\}$ (i.e. only the assignment to $b$-spaces that are contained in $L$). We present the local application first.

Fix an $\ell$-space $L$. Locally, $L$ can be identified with $\{-1, 1\}^\ell$ and the linear function $F[L]$ on it can be identified with a Fourier character $\chi_S$ for some $S \subseteq [\ell]$. The assignment $\{A[B] | B \subseteq L\}$ can be represented, in a similar manner as before, by a function $g : \{-1, 1\}^b \rightarrow \{-1, 1\}^b$, $g = (g_1, \ldots, g_b)$ where for $(w_1, \ldots, w_b) \in \{-1, 1\}^b$, $g_i(w_1, \ldots, w_b) = A[\text{Span}(w_1, \ldots, w_b)](w_i)$. We note that $g$ really is the restriction of $f$ to $L^b$. As before, for $T \subseteq [b]$, let $g_T = \prod_{i \in T} g_i$ be the product functions. We now relate the probability that $g$ passes the Gowers test with the probability that the linearity
test passes for the fixed $L$, i.e. the probability that $F[L] \mid B = A[B]$ for a random $B \subseteq L$. Let $1 - \gamma$ be the probability that random vectors $w_1, \ldots, w_b \in \{-1, 1\}^\ell$ are linearly independent, so that $\gamma \leq 2^{b-\ell}$ is negligible. Thus the distribution of a random $b$-dimensional subspace of $L$ is $\gamma$-close to the distribution of the space spanned by $b$ randomly chosen vectors from $L$. We now have

$$
(1 - \gamma) \cdot \Pr_{B \subseteq L} [F[L] \mid B = A[B]] \leq \Pr_{w_1, \ldots, w_b \in \{-1, 1\}^\ell} \left[ \bigwedge_{i=1}^b F[L](w_i) = A[\text{Span}(w_1, \ldots, w_b)](w_1) \right] 
$$

$$
= \Pr_{w_1, \ldots, w_b \in \{-1, 1\}^\ell} \left[ \bigwedge_{i=1}^b \chi_S(w_i) = g_i(w_1, \ldots, w_b) \right] 
$$

$$
= \mathbb{E}_{w_1, \ldots, w_b \in \{-1, 1\}^\ell} \left[ \prod_{i=1}^b \frac{1 + \chi_S(w_i)g_i(w_1, \ldots, w_b)}{2} \right] 
$$

$$
= \frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \mathbb{E}_{w_1, \ldots, w_b \in \{-1, 1\}^\ell} \left[ g_T(w_1, \ldots, w_b) \prod_{i \in T} \chi_S(w_i) \right] 
$$

$$
= \frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \tilde{g}_T(S_T), 
$$

where $S_T \subseteq [b]$ is defined as $(S_T(1), \ldots, S_T(b))$ and $S_T(i) \subseteq [\ell]$ equals $S$ if $i \in T$ and equals $\phi$ if $i \notin T$. Hence noting that $\gamma \leq 2^{b-\ell}$,

$$
\Pr_{B \subseteq L} [F[L] \mid B = A[B]] \leq \frac{1}{2^b} + 2 \cdot 2^{b-\ell} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \tilde{g}_T(S_T). 
$$

Now we take average of this inequality over the choice of $L \in \mathcal{L}$ and note that the L.H.S. then equals the probability that the linearity test accepts (which is $\geq \frac{1}{2^b} + \epsilon$). This gives

$$
\frac{\epsilon}{2} \leq \frac{\epsilon}{2} - 2 \cdot 2^{b-\ell} \leq \mathbb{E}_{L \in \mathcal{L}} \left[ \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \tilde{g}_T(S_T) \right].
$$

We keep in mind that $g$ and $S$ depend on the choice of $L$. Using convexity of the function $x \rightarrow x^4$, we get

$$
\frac{\epsilon^4}{16} \leq \mathbb{E}_{L \in \mathcal{L}} \left[ \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \tilde{g}_T^4(S_T) \right].
$$

Applying Lemma 3.7.3 to $g : \{-1, 1\}^\ell \rightarrow \{-1, 1\}^b$, we get

$$
\frac{\epsilon^4}{16} \leq \mathbb{E}_{L \in \mathcal{L}} \left[ \Pr [g \text{ passes Gowers test}] \right].
$$

We relate the R.H.S. to the probability that $f$ passes the Gowers test, using the fact that $g$ really is the restriction of $f$ to $L^b$. Let $x = (x_1, \ldots, x_b), y = (y_1, \ldots, y_b), z = (z_1, \ldots, z_b)$ where $x_i, y_i, z_i$ are either in $L$ or in the global space $\{-1, 1\}^n$, as understood from the context. We would like to argue as

$$
\frac{\epsilon^4}{16} \leq \mathbb{E}_{L \in \mathcal{L}} \left[ \Pr [g \text{ passes Gowers test}] \right] 
$$

$$
= \Pr_{L \in \mathcal{L}, x, y, z \in L} \left[ g(x)g(y)g(z)g(x \cdot y \cdot z) = 1^b \right] 
$$

$$
\approx \Pr_{x, y, z \in \{-1, 1\}^n} \left[ f(x)f(y)f(z)f(x \cdot y \cdot z) = 1^b \right] 
$$

$$
= \Pr [f \text{ passes Gowers test}].
$$

87
This is an almost correct argument, except that the distribution $D$ of $x_i, y_i, z_i \in \{-1, 1\}^n$ is slightly different from the distribution $D'$ of $L \in \mathcal{L}$, $x_i, y_i, z_i \in L = \{-1, 1\}^\ell$ (i.e. first choosing $L$ at random and then choosing $x_i, y_i, z_i$ from inside $L$). The distributions are identical however if conditioned on the $3b$ vectors $x_i, y_i, z_i$ being linearly independent. The probability of this happening is at least $1 - 2^{3b-n}$ and $1 - 2^{3b-\ell}$ depending on the space they are chosen from. It follows that the statistical distance between the distributions is at most $3 \cdot 2^{3b-\ell}$ and the argument above is correct up to that much error. It follows that (provided $\varepsilon \geq 4 \cdot 2^{-b/4}$)

$$\frac{\varepsilon^4}{24} \leq \Pr \left[ f \text{ passes Gowers test} \right].$$

Applying Lemma 3.7.3 to $f$, 

$$\frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \phi} \sum_{S \subseteq [nb]} \tilde{f}_T(S) \geq \frac{\varepsilon^4}{24}.$$  

Noting that the sum of squares of Fourier coefficients of a boolean function equals 1, we see that there exists $T \subseteq [b], T \neq \phi$ and $S \subseteq [nb]$ such that $\tilde{f}_T(S) \geq \frac{\varepsilon^4}{32}$. We are almost done, by inspecting the coefficient $\tilde{f}_T(S)$. Let $S = (S_1, \ldots, S_6)$, $S_i \subseteq [n]$, and denote $B = \text{Span}(v_1, \ldots, v_b)$ below. By definition of Fourier coefficients and of the functions $f, f_T$,

$$\tilde{f}_T(S) = \mathbb{E}_{v_1, \ldots, v_b \in \{-1, 1\}^n} \left[ f_T(v_1, \ldots, v_b) \cdot \prod_{i=1}^b \chi_S(v_i) \right]$$

$$= \mathbb{E}_{v_1, \ldots, v_b \in \{-1, 1\}^n} \left[ \prod_{i \in T} A[B](v_i) \cdot \prod_{i=1}^b \chi_S(v_i) \right]$$

$$= \mathbb{E}_{B : \text{dim}(B) = b, \ v_1, \ldots, v_b \in B, \ \text{Rank}(v_1, \ldots, v_b) = b} \left[ \prod_{i \in T} A[B](v_i) \cdot \prod_{i=1}^b \chi_S(v_i) \right] \pm 2^{b-n},$$

where while choosing $v_1, \ldots, v_b \in \{-1, 1\}^n$, they are assumed to be linearly independent (introducing the negligible error term $2^{b-n}$) and then their choice is same as first choosing a random $b$-space $B$ and then letting $v_1, \ldots, v_b$ be a random basis of $B$. Regard $B = \{-1, 1\}^b$ and $A[B]$ as the linear function $\chi_{S[B]}$ for $S[B] \subseteq [b]$. The global function $\chi_{S_i}(v_i)$ where $S_i \subseteq [n], v_i \in \{-1, 1\}^n$, after restricting to $v_i \in B$, amounts to a linear function on $B$, say $\chi_{S_i \downarrow B}$ with $S_i \downarrow B \subseteq [b]$. Thus

$$\tilde{f}_T(S) = \mathbb{E}_{B : \text{dim}(B) = b, \ v_1, \ldots, v_b \in B, \ \text{Rank}(v_1, \ldots, v_b) = b} \left[ \prod_{i \in T} \chi_{S[B] \Delta S_i \downarrow B}(v_i) \cdot \prod_{i \in [b] \setminus T} \chi_{S_i \downarrow B}(v_i) \right] \pm 2^{b-n}.$$ 

Let us look at the expectation for a fixed $B$. Call $B$ good if

$$\forall \ i \in T, \ S_i \downarrow B = S[B], \quad \forall \ i \in [b] \setminus T, \ S_i \downarrow B = \phi,$$

(3.19)

and let $B'$ be the set of such good $B$. For a good $B$, the expectation equals 1 and from Lemma 3.8.1, the expectation is bounded by $2^{-b+1}$ in magnitude for a bad $B$. Thus

$$\tilde{f}_T(S) = \Pr_B \left[ B \in B' \right] \pm 2^{-b+1} \pm 2^{b-n}.$$ 

Since $|\tilde{f}_T(S)| \geq \frac{\varepsilon^4}{24}$, it follows that $\Pr_B \left[ B \in B' \right] \geq \frac{\varepsilon^4}{48}$ (since $\varepsilon \geq 2^{2-b/4} \geq 2^3-b/2$). Now we show that in fact for some $S^* \subseteq [n]$, for all $i \in T$, $S_i = S^*$ and for all $i \in [b] \setminus T$, $S_i = \phi$. This is because if this were not the case, for a random $b$-space $B$, Condition (3.19) holds with probability at most $2^{-b}$, upper bounding $\Pr_B \left[ B \in B' \right]$ by $2^{-b}$, a contradiction. It follows that $\chi_{S^*} : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is a global linear function that agrees with the given linear function $A[B]$ on $\geq \frac{\varepsilon^4}{48}$ fraction of $B$, $\text{dim}(B) = b$.  

88
Agreement with $\ell$-spaces

What we have concluded so far is that if tables $F, A$ pass the $\ell$-space vs $b$-space linearity test with probability $\geq \frac{1}{2^b} + \varepsilon$, then there is a global linear function $g : \{-1, 1\}^n \rightarrow \{-1, 1\}^n$ that agrees with $A[B]$ for $\geq \frac{\varepsilon^2}{80}$ fraction of $b$-spaces $B$. Theorem 3.7.1 however demands a good agreement with $F[L]$ for $\ell$-spaces $L$. This is easy to fix. Let

$$B^* = \left\{ B \mid \Pr_{L: B \subseteq L} [F[L]|_B = A[B]] \geq \frac{\varepsilon}{2} \right\}.$$ 

Since the linearity test succeeds with probability $\geq \frac{1}{2^b} + \varepsilon$, by an averaging argument, $|B^*| \geq \frac{\varepsilon}{2} \cdot |B|$. Now modify the table $A[\cdot]$ to table $A'[\cdot]$ so that $A'[B] = A[B]$ for $B \in B^*$ and $A'[B]$ is a random linear function on $B$ otherwise. Clearly, the tables $F, A'$ still pass the linearity test with probability $\geq 2^{-b} + \frac{\varepsilon}{2}$ and by the analysis so far, there is a global linear function $g$ that agrees with $A'[B]$ for $\geq \frac{\varepsilon^2}{80}$ fraction of $B \in B$. Since $A'[B]$ for $B \not\in B^*$ was defined at random, their contribution to consistency with $g$ is negligible, i.e. at most $2^{-b}$. Thus we have

$$\Pr_B [g|_B = A[B] \land B \in B^*] \geq \frac{\varepsilon^2}{80}.$$ 

For every $B \in B^*$, by definition, $A[B]$ is consistent with $F[L]$ for at least $\frac{\varepsilon}{2}$ fraction of $L$ containing $B$. Hence,

$$\Pr_{B \subseteq L} [g|_B = A[B] \land B \in B^* \land F[L]|_B = A[B]] \geq \frac{\varepsilon^3}{160}.$$ 

In particular,

$$\mathbb{E}_L \left[ \Pr_{B \subseteq L} [F[L]|_B = g|_B] \right] \geq \frac{\varepsilon^3}{160}.$$ 

This implies immediately that $g|_L = F[L]$ for at least $\frac{\varepsilon^3}{160} - 2^b \geq \frac{\varepsilon^3}{300}$ fraction of $L$, since for $L$ not satisfying this, the inside probability is at most $2^{-b}$.

### 3.8 Missing Proofs

#### 3.8.1 Covering Properties

**Proof of Lemma 3.5.2**

We recall that $U, |U| = 3k$ is a set of $3k$ variables in $k$ equations. A subset $V \subseteq U$ is chosen by picking independently for each equation, one of the variables in the equation with probability $\beta$ and all three variables in the equation with probability $1 - \beta$. The expected size of $V$ is $3k - 2\beta k$ and except with probability $2^{-\Omega(k)}$, we have $|V| \geq 2k$.

We note that choosing a uniformly random $\ell$-subspace $L$ of $\{0, 1\}^U$ (resp. $\{0, 1\}^V$) is equivalent to choosing uniformly a sequence of points $x_1, \ldots, x_\ell$ in $\{0, 1\}^U$ (resp. $\{0, 1\}^V$) that are linearly independent and letting $L = \text{Span}(x_1, \ldots, x_\ell)$. Since a uniformly random and independent sequence of points $x_1, \ldots, x_\ell$ in $\{0, 1\}^U$ (resp. in $\{0, 1\}^V$) is linearly independent except with probability $\leq 2^{\ell - \dim(U)}$ (resp. $\leq 2^{\ell - \dim(V)}$, see Fact 3.8.2), we might as well focus on such sequences of points. It is thus enough to bound the statistical distance between distributions $\mathcal{D}, \mathcal{D}'$ over $\{(0, 1)^U\}^\ell$ sampled as:

- $\mathcal{D}$: Choose uniformly and independently $x_1, \ldots, x_\ell \in \{0, 1\}^U$.
- $\mathcal{D}'$: Choose $V \subseteq U$, choose uniformly and independently $x'_1, \ldots, x'_\ell \in \{0, 1\}^V$ and regard them as points in $\{0, 1\}^U$ (by appending 0 in coordinates $U \setminus V$).
We now observe that since the process of choosing $V \subseteq U$ is independent over the $k$ equations, $\mathcal{D} = S^k$ and $\mathcal{D}' = S'^k$ where $S, S'$ are the “basic” distributions exactly as above, but with $k = 1, |U| = 3$. A bound on the statistical distance between $\mathcal{D}, \mathcal{D}'$ now follows in the same manner as in [KS13, Lemma 3.1], by bounding the Hellinger distance between $S, S'$, using the multiplicativity of the Hellinger distance to bound the Hellinger distance between $\mathcal{D}, \mathcal{D}'$ and finally, bounding the Hellinger distance in terms of the statistical distance. We observe how a bound on Hellinger distance between $S, S'$ also follows already from the proof of [KS13, Lemma 3.1]. Re-writing the sampling process for $S, S'$ for convenience (this is the special case $k = 1, \ell_U = 3$):

- $S$: Choose uniformly at random $x_1, \ldots, x_\ell \in \{0, 1\}^3$.

- $S'$: With probability $1 - \beta$, choose uniformly at random $x_1, \ldots, x_\ell \in \{0, 1\}^3$. Otherwise:
  Choose uniformly at random $b_1, \ldots, b_\ell \in \{0, 1\}$. Output with probability $\frac{\beta}{3}$ each,
  $$b_100, \ldots, b_\ell00, \text{ or } 0b_10, \ldots, 0b_\ell0, \text{ or } 00b_1, \ldots, 00b_\ell.$$  

The distributions $S, S'$ are over $\{0, 1\}^3$ which is equivalent to $\Sigma^3$ where $\Sigma = \{0, 1\}^\ell$ is the concatenation of the first (second, third, respectively) bit of each of the $\ell$ strings of length three. Note that $0^\ell \in \Sigma$. Denoting the uniform distribution over $\Sigma$ by $\text{Uniform}(\Sigma)$, it is seen that

$$S = (\text{Uniform}(\Sigma), \text{Uniform}(\Sigma), \text{Uniform}(\Sigma)),$$

i.e. three independent and uniform copies of $\Sigma$, whereas,

$$S' = \left(1 - \beta\right) (\text{Uniform}(\Sigma), \text{Uniform}(\Sigma), \text{Uniform}(\Sigma)) + \frac{\beta}{3} \left(\text{Uniform}(\Sigma), 0^\ell, 0^\ell\right) + \frac{\beta}{3} \left(0^\ell, \text{Uniform}(\Sigma), 0^\ell\right) + \frac{\beta}{3} \left(0^\ell, 0^\ell, \text{Uniform}(\Sigma)\right).$$

With this viewpoint, the Hellinger distance between $S, S'$ is calculated to be at most $4\beta^2 |\Sigma|^2$ in the proof of [KS13, Lemma 3.1]. The statistical distance between $\mathcal{D}, \mathcal{D}'$ is then at most $16\beta \sqrt{k} \cdot |\Sigma|$.

**Proof of Lemma 3.5.3**

Let $U, V, k, \beta, \ell, \mathcal{L}, \mathcal{L}', q, \mathcal{L}_Q, \mathcal{L}'_Q$ be as in Definition 3.5.1 and Lemmas 3.5.2, 3.5.3. Let $Q, Q'$ be distributions over $q$-dimensional subspaces of $\{0, 1\}^U$ that are analogous to $\mathcal{L}, \mathcal{L}'$ respectively (i.e. as in Definition 3.5.1, with parameter $q$ instead of $\ell$). It is easily observed that an equivalent way to sample from $Q$ (resp. $Q'$) is to sample an $\ell$-space $L$ from $\mathcal{L}$ (resp. $\mathcal{L}'$) and then sample a uniformly random $q$-dimensional subspace of $L$. We stress that $Q$ and $\mathcal{L}$ are uniform distributions on $q$-dimensional and $\ell$-dimensional subspaces of
\{0, 1\}^U$ respectively. We have the sequence of arguments

$$
\mathbb{E}_{Q \sim Q} \left[ \text{SD}(\mathcal{L}_Q, \mathcal{L}'_Q) \right] = \sum_Q \Pr[Q = Q] \sum_{L \subseteq Q} \left| \Pr[\mathcal{L}_Q = L] - \Pr[\mathcal{L}'_Q = L] \right|
$$

$$
= \sum_Q \sum_{L \subseteq L} \Pr[Q = Q] \cdot \Pr[\mathcal{L}_Q = L] - \Pr[Q = Q] \cdot \Pr[\mathcal{L}'_Q = L]
$$

$$
\leq \sum_Q \sum_{L \subseteq Q} \left| \Pr[Q = Q] \cdot \Pr[\mathcal{L}_Q = L] - \Pr[Q' = Q] \cdot \Pr[\mathcal{L}'_Q = L] \right| + \sum_Q \sum_{L \subseteq Q} \left| \Pr[Q' = Q] \cdot \Pr[\mathcal{L}'_Q = L] - \Pr[Q = Q] \cdot \Pr[\mathcal{L}'_Q = L] \right|
$$

$$
= \sum_L \left| \Pr[\mathcal{L} = L] - \Pr[\mathcal{L}' = L] \right| + \sum_Q \Pr[Q' = Q] - \Pr[Q = Q] \sum_{L \subseteq Q} \Pr[\mathcal{L}'_Q = L]
$$

$$
= \text{SD}(\mathcal{L}, \mathcal{L}') + \text{SD}(Q, Q'),
$$

where we used triangle inequality and the fact that sampling $Q \in Q$ and then $L \sim \mathcal{L}_Q$ is equivalent to sampling $L \sim \mathcal{L}$ and similarly, sampling $Q \in Q'$ and then $L \sim \mathcal{L}'_Q$ is equivalent to sampling $L \sim \mathcal{L}'$. Now Lemma 3.5.2 upper bounds both $\text{SD}(\mathcal{L}, \mathcal{L}')$ and $\text{SD}(Q, Q')$ by $\beta\sqrt{k} \cdot 2^{\ell+4}$ and hence

$$
\mathbb{E}_{Q \sim Q} \left[ \text{SD}(\mathcal{L}_Q, \mathcal{L}'_Q) \right] \leq \beta\sqrt{k} \cdot 2^{\ell+5}.
$$

Lemma 3.5.3 now follows by Markov inequality.

### 3.8.2 Proof of Lemma 3.5.8

The proof follows rather easily from Theorem 3.7.1 in Section 3.7, and is given below.

Define a table $A[\cdot]$ that assigns, to every $b$-dimensional subspace $B$, a linear function $A[B]$ on it, by letting $A[B] = F[L]|_B$ for a randomly chosen $\ell$-dimensional subspace $L$ containing $B$. It follows that, for a random pair $B \subseteq L$ of $b$-dimensional and $\ell$-dimensional subspaces respectively, $A[B]$ and $F[L]$ are consistent on $B$ with probability at least $\beta$. By Theorem 3.7.1, there is a global linear function $g : U \rightarrow \{0, 1\}$ that agrees with $F[\cdot]$ on $\beta^3 \cdot 300$ fraction of $\ell$-dimensional subspaces.

### 3.8.3 Proof of the Upper Bound in Lemma 3.5.9

Denote by $N$ the size of $\mathcal{L} \overset{\text{def}}{=} \{ L \in G(V, \ell) \mid Q \subseteq L \}$ and let $f_1, \ldots, f_m$ be all functions agreeing with $F[\cdot]$ on at least $C$ fraction of $L \in \mathcal{L}$. We construct a bipartite graph, where the left side consists of $f_1, \ldots, f_m$ and the right side consists of pairs $(L, \sigma)$ where $(L, \sigma) \in F[L]$. We connect $f_i$ and $(L, \sigma)$ by an edge if $f_i|_L \equiv \sigma$. Then the degree of each $f_i$ is at least $C \cdot N$ and the number of vertices on the right side is at most $jN$. Let us remove edges if necessary so that the degree of each $f_i$ is exactly $C \cdot N$.

Denote by $d(L, \sigma)$ the degree of $(L, \sigma)$ and let us count the number of triples $\{(f_i, f_j, (L, \sigma)) \mid i \neq j \text{ and } (f_i, (L, \sigma)), (f_j, (L, \sigma)) \}$ are both edges in the bipartite graph. Using Cauchy-Schwartz and noting that the number of vertices on the right side is at most $jN$ and $\sum_{L \in \mathcal{L}, \sigma \in F[L]} d(L, \sigma) = CmN$, the number of such triples is lower bounded as

$$
\sum \frac{d(L, \sigma)}{2} = \sum \frac{d(L, \sigma)^2}{2} - \frac{d(L, \sigma)}{2} \geq jN \cdot \left( \frac{CmN}{jN} \right)^2 - \frac{CmN}{2} = \frac{C^2m^2N}{2j} - \frac{CmN}{2}.
$$
On the other hand, since any distinct pair of functions $f_i, f_j$ agree on at most $2^{q-\ell}$ fraction of $L \in \mathcal{L}$, the number of such triples is at most $\binom{m}{2} 2^{q-\ell} N \leq \frac{m^2 2^{q-\ell} N}{2}$. Combining the two bounds gives $m \leq \frac{jC}{C^2 - j2^{q-\ell}}$.

### 3.8.4 Auxiliary Lemmas and Facts

**Lemma 3.8.1.** Let $s_1, \ldots, s_b \in \mathbb{F}_2^b$ such that at least one of them is non-zero. Let $v_1, \ldots, v_b \in \mathbb{F}_2^b$ be chosen at random. Then the following conditional expectation is bounded as:

$$
\mathbb{E}_{v_1, \ldots, v_b} \left[ \prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid \text{Rank}(v_1, \ldots, v_b) = b \right] \leq 2^{-b+1}.
$$

**Proof.** Note that without the conditioning, the expectation is clearly zero. The point is to prove the upper bound conditional on the event that $v_1, \ldots, v_b$ are linearly independent (and hence form a basis of $\mathbb{F}_2^b$). Assume w.l.o.g. that $s_1$ is non-zero. Let

$$
\mathcal{A} = \{ A = (v_1, \ldots, v_b) \mid \text{Rank}(A) = b \},
$$

so that we are interested in the expectation

$$
\mathbb{E}_A \left[ \prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid A \in \mathcal{A} \right].
$$

Let

$$
\mathcal{A}' = \{ A = (v_1, \ldots, v_b) \mid \text{Rank}(A) = b, \forall 2 \leq i \leq b, \langle s_1, v_i \rangle = 0 \}.
$$

It is easily seen that $|\mathcal{A}'| \leq 2^{-b+1} \cdot |\mathcal{A}|$. Imagine choosing $v_2, v_3, \ldots, v_b$ so that every $v_i$ is outside the span of the previously chosen ones. If we require (in addition) that every $v_i$ also lies in the hyperplane defined by the equation $\langle s_1, x \rangle = 0$, then at each step, this happens with probability at most $\frac{1}{2}$, showing the desired upper bound on $|\mathcal{A}'|$. Hence the two expectations

$$
\mathbb{E}_A \left[ \prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid A \in \mathcal{A} \right], \quad \mathbb{E}_A \left[ \prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid A \in \mathcal{A} \setminus \mathcal{A}' \right]
$$

differ by at most $2^{-b+1}$. We show that the latter is zero. For fixed $\alpha_2, \alpha_3, \ldots, \alpha_b \in \mathbb{F}_2$, consider the following bijection on $\mathcal{A} \setminus \mathcal{A}'$ (that adds to the first vector, a linear combination of others):

$$(v_1, v_2, v_3, \ldots, v_b) \rightarrow (v_1 + \sum_{i=2}^{b} \alpha_i v_i, v_2, v_3, \ldots, v_b).$$

The quantity of interest changes as follows:

$$
\prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \rightarrow (-1)^{\sum_{i=2}^{b} \alpha_i \langle s_1, v_i \rangle} \cdot \prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle}.
$$

Now take expectation of L.H.S. over the choice of $A = (v_1, \ldots, v_b) \in \mathcal{A} \setminus \mathcal{A}'$ and expectation of R.H.S. over the choice of $A \in \mathcal{A} \setminus \mathcal{A}'$ as well as over a random choice of $\alpha_2, \ldots, \alpha_b$. The two expectations are equal (due to bijectivity) and the expectation of the L.H.S. is what we are interested in. Since $\langle s_1, v_i \rangle \neq 0$ for some $2 \leq i \leq b$, the expectation over the R.H.S. is zero and we are done. \qed

92
**Fact 3.8.2.** Let \( V \) be an \( n \)-dimensional vector space over \( \mathbb{F}_2 \), and \( 1 \leq \ell \leq n - 1 \). Let \( x_1, \ldots, x_\ell \in V \) be chosen randomly and independently. Then \( x_1, \ldots, x_\ell \) are linearly independent with probability \( \geq 1 - 2^{\ell - n} \).

**Proof.** If \( x_1, \ldots, x_\ell \) are linearly dependent, then for some \( 1 \leq i \leq \ell \), \( x_i \) is in the span of \( x_1, \ldots, x_{i-1} \). Hence the probability that these \( \ell \) vectors are linearly dependent can be upper-bounded as

\[
\sum_{i=1}^\ell \Pr_{x_1, \ldots, x_i \in V} [x_i \in \text{Span}\{x_1, \ldots, x_{i-1}\}] \leq \sum_{i=1}^\ell \frac{2^{i-1}}{2^n} \leq 2^{\ell - n}.
\]

\( \square \)

**Fact 3.8.3.** Let \( V \) be an \( n \)-dimensional vector space over \( \mathbb{F}_2 \), \( H \subseteq U \) be a subspace of dimension \( r \), and \( 1 \leq \ell \leq n - \ell \). Let \( x_1, \ldots, x_\ell \in V \) be chosen randomly and independently. Then

\[
\Pr_{x_1, \ldots, x_\ell \in V} [\text{Span}\{x_1, \ldots, x_\ell\} \cap H = \{0\}] \geq 1 - 2^{r + \ell - n}.
\]

**Proof.** If \( \text{Span}\{x_1, \ldots, x_\ell\} \cap H \neq \{0\} \), then for some \( 1 \leq i \leq \ell \), \( x_i \) is in the span of \( H \cup \{x_1, \ldots, x_{i-1}\} \). Hence the probability that \( \text{Span}\{x_1, \ldots, x_\ell\} \cap H \neq \{0\} \) can be upper-bounded as

\[
\sum_{i=1}^\ell \Pr_{x_1, \ldots, x_\ell \in V} [x_i \in H \oplus \text{Span}\{x_1, \ldots, x_{i-1}\}] \leq \sum_{i=1}^\ell \frac{2^{r+i-1}}{2^n} \leq 2^{r + \ell - n}.
\]

\( \square \)

**Fact 3.8.4.** Let \( A, A', B \) be subspaces of a vector space \( V \) over \( \mathbb{F}_2 \) such that \( A \oplus B = A' \oplus B \) and \( (A \oplus A') \cap B = \{0\} \). Then \( A = A' \).

**Proof.** By symmetry, it suffices to show that \( A \subseteq A' \). Let \( a \in A \). Then \( a \in A \oplus B = A' \oplus B \) and so there are \( a' \in A', b \in B \) such that \( a = a' \oplus b \). Hence \( b = a \oplus a' \in A \oplus A' \), and \( b \) must be 0.  

\( \square \)
Chapter 4

Pseudo-random Sets in the Grassmann Graph Have Near-Perfect Expansion

4.1 Introduction

The Grassmann Graph is a key object in the reduction presented in Chapter 3 for the proof of the NP-hardness of 2-to-2 Games (Theorem 1.2.6). In fact, the reduction therein gives hardness for the “linear 2-to-2 Games” described below. Let $\mathbb{F}_2^\ell$ denote the $\ell$-dimensional vector space over the binary field $\mathbb{F}_2$, considered as an additive group with the $\oplus$ operation.

Definition 4.1.1 (Linear 2-to-2 Games). An instance $U_{2\to2}$ of the 2-to-2 Game $[\mathbb{F}_2^\ell]$ problem consists of $n$ variables $x_1, \ldots, x_n$ taking values over the alphabet $\mathbb{F}_2$, considered as an additive group with the $\oplus$ operation.

For constants $0 < \varepsilon < c \leq 1$, let $\text{Gap}_{2\to2}[\mathbb{F}_2^\ell](c, \varepsilon)$ be the gap-version where the instance $U_{2\to2}$ of the 2-to-2 Game $[\mathbb{F}_2^\ell]$ problem is promised to have either $\text{OPT}(U_{2\to2}) \geq c$ or $\text{OPT}(U_{2\to2}) \leq \varepsilon$.

Theorem 4.1.2. For every constant $\varepsilon > 0$, there exists a sufficiently large integer $\ell = \ell(\varepsilon)$ such that $\text{Gap}_{2\to2}[\mathbb{F}_2^\ell](1 - \varepsilon, \varepsilon)$ is NP-hard.

Though weaker than the Unique Games Conjecture, the 2-to-2 Games Conjecture has important applications of its own and also supports the Unique Games Conjecture, please see Section 1.2 for more details.

In the following, we introduce the Grassmann graph and state the main result of this chapter – a structural result on vertex sets whose edge expansion is bounded away from 1. Consider an $n$-vertex $d$-regular graph $G(V, E)$ and a non-empty set of vertices $S \subseteq V$ with $|S| \leq \frac{n}{2}$. Its edge expansion is defined as

$$\Phi(S) = \frac{|E(S, \overline{S})|}{d \cdot |S|},$$

where $E(S, \overline{S})$ denotes the set of edges with one endpoint in $S$ and the other in $\overline{S} = V \setminus S$. Alternately, it is the probability that selecting a uniformly random vertex in $S$ and moving along a uniformly random edge incident on that vertex, one lands outside $S$. It is clear that if $S$ is a randomly selected set of size $o(n)$, then its expansion is $1 - o(1)$ (with high probability over the choice of the set), i.e. small random sets have near-perfect expansion.
Let \( k, \ell \) be integer parameters with \( 1 \ll \ell \ll k \). We will be interested in the Grassmann graph \( \text{Gr}_{k,\ell} \) and subsets \( S \) of its vertices that have expansion strictly bounded away from 1 (say at most \( \frac{31}{32} \)). Such sets will be referred to as “non-perfectly expanding”.

**Definition 4.1.3.** The Grassmann graph \( \text{Gr}_{k,\ell} \) is defined as follows. Its vertex set consists of all \( \ell \)-dimensional subspaces \( L \) of \( \mathbb{F}_2^k \) and \( (L, L') \) is an edge if and only if \( \dim(L \cap L') = \ell - 1 \).

**Definition 4.1.4.** Suppose \( A \subseteq B \subseteq \mathbb{F}_2^k \) are subspaces. Let \( \dim(A) = a, \codim(B) = b \) and think of \( a, b \) as small constants (say \( a = b = 2 \)). Then the subgraph \( \text{Gr}_{k,\ell}[A, B] \) is the induced subgraph of \( \text{Gr}_{k,\ell} \) on the set of vertices \( L \) such that \( A \subseteq L \subseteq B \). It is easy to see that \( \text{Gr}_{k,\ell}[A, B] \) is an isomorphic copy of a lower order Grassmann graph \( \text{Gr}_{k-a-b,\ell-a} \). We call \( a + b \) the co-order of \( \text{Gr}_{k,\ell}[A, B] \) with respect to \( \text{Gr}_{k,\ell} \).

The sets \( \text{Gr}_{k,\ell}[A, B] \) are natural examples of sets in \( \text{Gr}_{k,\ell} \) that have expansion strictly bounded away from 1 (when \( a, b \) are small constants). Indeed, the expansion of the vertex set of \( \text{Gr}_{k,\ell}[A, B] \), considered in of \( \text{Gr}_{k,\ell} \), is \( 1 - 2^{-(a+b)} \) (up to an error \( O(2^{-\ell}) \) which is thought of as negligible and ignored for ease of presentation). The reasoning is as follows. For a vertex \( L \in \text{Gr}_{k,\ell}[A, B] \), its random neighbor \( L' \) is obtained by picking a random subspace \( T \subseteq L, \dim(T) = \ell - 1 \) and a random point \( x \in \mathbb{F}_2^k \setminus L \) and letting \( L' = T \oplus \text{Span}(x) \). Now \( L' \in \text{Gr}_{k,\ell}[A, B] \) if and only if \( A \subseteq T \) and \( x \in B \) and these events happen independently with probabilities \( 2^{-a} \) and \( 2^{-b} \) respectively (up to an error \( O(2^{-\ell}) \)). Thus a random neighbor of a random vertex in \( \text{Gr}_{k,\ell}[A, B] \) is also inside it with probability \( 2^{-(a+b)} \) and hence its expansion is \( 1 - 2^{-(a+b)} \). Furthermore, we observe that if \( S \subseteq \text{Gr}_{k,\ell}[A, B] \subseteq \text{Gr}_{k,\ell} \) is such that

\[
\frac{|S|}{|\text{Gr}_{k,\ell}[A, B]|} = \varepsilon,
\]

then \( \Phi(S) \leq 1 - \varepsilon \cdot 2^{-(a+b)} \). This is because (by an easy application of the Cauchy-Schwarz Inequality) any set of density \( \varepsilon \) inside a Grassmann graph has at least \( \varepsilon^2 \) fraction of the edges inside it (and hence has expansion at most \( 1 - \varepsilon \)). Therefore, a random neighbor of a random vertex in \( S \subseteq \text{Gr}_{k,\ell}[A, B] \) lies inside \( \text{Gr}_{k,\ell}[A, B] \) with probability \( 2^{-(a+b)} \) as seen above and then inside \( S \) with probability at least \( \varepsilon \), justifying the observation. We summarize the overall observation as:

**Fact 4.1.5.** Let \( S \subseteq \text{Gr}_{k,\ell}[A, B] \subseteq \text{Gr}_{k,\ell} \) be such that \( \dim(A) = a, \codim(B) = b \) and the density of \( S \) inside \( \text{Gr}_{k,\ell}[A, B] \) is \( \geq \varepsilon \). Then \( \Phi(S) \leq 1 - \varepsilon \cdot 2^{-(a+b)} \).

The main result of this Chapter is a converse-type statement of the above fact. Informally, it states that any set \( S \) in the Grassmann graph \( \text{Gr}_{k,\ell} \) whose expansion is strictly bounded away from 1, has constant density inside some copy of Grassmann graph of constant co-order. More precisely:

**Theorem 4.1.6 (Rephrasing of Theorem 1.3.4).** For every constant \( 0 < \alpha < 1 \), there exists a constant \( \varepsilon > 0 \) and an integer \( r \geq 0 \) such that for all sufficiently large integers \( \ell \) and (after fixing it) for all sufficiently large integers \( k \), the following holds. Let \( S \subseteq \text{Gr}_{k,\ell} \) be such that \( \Phi(S) \leq \alpha \). Then there exist subspaces \( A \subseteq B \subseteq \mathbb{F}_2^k \) such that \( \dim(A) = a, \codim(B) = b, a + b \leq r \) and

\[
\frac{|S \cap \text{Gr}_{k,\ell}[A, B]|}{|\text{Gr}_{k,\ell}[A, B]|} \geq \varepsilon.
\]

The bulk of this chapter is devoted to the proof of Theorem 4.1.6. To have this thesis a self-contained proof of the 2-to-2 Games Conjecture, we also present an argument due to Barak, Kothari and Stuerer [BKS18] showing that Theorem 4.1.6 implies Theorem 3.2.7.

We end this introductory section with a few remarks.
1. Theorem 4.1.6 was first stated as a hypothesis in [DKK+18b], wherein the authors observe that this statement is necessary to prove Theorem 3.2.7. They prove Theorem 4.1.6 when $\alpha < \frac{1}{7}$, via spectral analysis of the Grassmann graph, introduced therein (the eigenvalues and eigenspaces of the Grassmann graph were known before). Roughly speaking, given a set $S$ with expansion at most $\alpha < 1 - 2^{-1(s+1)}$, it is easily observed that the indicator vector of the set $1_S$ must have a significant projection onto the eigenspace at “level” at most $s$ ($s$ is a constant when $\alpha$ is strictly bounded away from 1). The spectral analysis then attempts to use this projection to deduce the desired structure of $S$. The approach is worked out in [DKK+18b] when $s = 2$, corresponding to $\alpha < \frac{1}{7}$. It already requires rather difficult and lengthy case analysis. In principle, the same approach could be extended to higher levels $s \geq 3$, but the number of cases to handle seems to explode beyond control. We present here a more systematic argument and avoid the explosion in potential case analysis (easier said than done of course).

2. Following [DKK+18b], Barak Kothari and Steurer proved that is suffices to prove Theorem 4.1.6 to establish Theorem 3.2.7. For the sake of completeness, we present their argument in Section 4.2.1.

3. The subspaces $A$ and $B$ are referred to as “zoom-in” and “zoom-out” spaces respectively in [KMS17, DKK+18a, DKK+18b]. This makes sense if one imagines searching for the appropriate subgraph $Gr_{k,\ell}[A, B]$ where the set $S$ happens to have significant density.

4. We note that if $S$ has density $\geq \varepsilon$, then the conclusion of the theorem is vacuously true without any need for a zoom-in or a zoom-out (i.e. $a = b = 0, A = \{0\}, B = \mathbb{F}_2^k$), so the theorem is really about “small” sets.

5. Our proof gives correct dependence of the required zoom-in-out dimension $r$ on the upper bound on expansion $\alpha$. For $\alpha < 1 - 2^{-(s+1)}$, one gets a significant projection onto the eigenspace at level at most $s$ and then in our proof, a (combined) zoom-in-out dimension of at most co-order $r = s$ is sufficient. This is tight (i.e. a lesser zoom-in-out dimension is not sufficient) since we know that subgraphs $Gr_{k,\ell}[A, B]$ have expansion $1 - 2^{-(a+b)}$ and the zoom-in-out dimension $a + b$.

6. In the proof of Theorem 4.1.6, it will be easier to work with the contra-positive statement: a set $S$ that has very small density inside every copy of the Grassmann graph with constant co-order (such a set will be called pseudorandom) has near-perfect expansion (i.e. very near 1).

7. The phenomenon as in Theorem 4.1.6 occurs also in the Johnson graph and has been analyzed in [KMMS18]. In a Johnson graph, the vertices are $\ell$-subsets of a $k$-set and the edges are pairs of $\ell$-subsets with intersection of size $t$ (we are concerned with the case when $t = \alpha \ell$ for constant $\alpha \in (0, 1)$). Therein, only the notion zoom-in is required.

4.2 Preliminaries

4.2.1 The Barak-Kothari-Steurer Reduction

In this section we present an argument of Barak, Kothari and Steurer [BKS18] showing Theorem 4.1.6 implies Theorem 3.2.7.

**Theorem 4.2.1.** Theorem 4.1.6 implies Theorem 3.2.7.

**Proof.** Fix $\delta > 0$, and let $r, \varepsilon$ be as in Theorem 4.1.6 for $\frac{1}{4} \delta$. 

97
Let $F$ be a $\delta$-consistent assignment on $G(V, \ell)$ assigning to every subspace a linear function. For each linear function $f: V \to \mathbb{F}_2$ define the set of $\ell$-dimensional subspace that agree with it

$$A_f = \{ L \mid F[L] = f|_L \}.$$ 

Let $Z[L, f]$ be the indicator random variable of the event that $F[L] = f|_L$. By linearity of expectation,

$$\mathbb{E}_{f} [ |A_f| ] = \mathbb{E}_{f} \left[ \sum_{L} Z[L, f] \right] = \sum_{L} \mathbb{E}_{f} [Z[L, f]] = \sum_{L} 2^{-\ell} = 2^{-\ell} \left[ \frac{k}{\ell} \right].$$

I.e. the expected density of $A_f$ is $2^{-\ell}$. Let $E'$ be the edge set of $G(V, \ell)$, $m$ be its size and $E' \subseteq E$ be the set of edges on which $F$ is consistent. Since $F$ is $\delta$-consistent, we have $|E'| \geq \delta m$ and therefore

$$\mathbb{E}_{f} [ E(A_f, A_f) ] = 2 \mathbb{E}_{f} \left[ \sum_{(L, L') \in E} Z[L, f] Z[L', f] \right] = 2 \sum_{(L, L') \in E} \mathbb{E}_{f} [Z[L, f] Z[L', f]] \geq 2 \sum_{(L, L') \in E'} \mathbb{E}_{f} [Z[L, f] Z[L', f]].$$

The factor 2 in the first equality comes from the double-counting of edges in $E(A_f, A_f)$. Note that for every edge $(L, L')$ consistent according to $F$, we have that the inner expectation is $2^{-(\ell+1)}$. Therefore we get that

$$\mathbb{E}_{f} [ E(A_f, A_f) ] \geq 2 |E'| 2^{-(\ell+1)} \geq \delta m 2^{-\ell}.$$

Combining the two inequalities, we get from linearity of expectation that ($d$ is the degree of $G(V, \ell)$)

$$\mathbb{E}_{f} \left[ E(A_f, A_f) - \frac{\delta}{4} d |A_f| \right] > 0,$$

and in particular there exists $f$ such that $E(A_f, A_f) > \frac{\delta}{4} d |A_f|$, implying

$$\frac{E(A_f, A_f)}{d |A_f|} \geq \frac{\delta}{4}.$$

Equivalently, $\Phi(A_f) \leq 1 - \frac{\delta}{4}$ and by Theorem 4.1.6 there are $Q \subseteq W$ such that $\dim(Q) + \text{codim}(W) \leq r$, and

$$\Pr_{Q \subseteq L \subseteq W} [L \in A_f] \geq \varepsilon,$$

i.e. $\Pr_{Q \subseteq L \subseteq W} [F[L] = f|_L] \geq \varepsilon$, and we are done ($g_{Q,W} = f$).

4.2.2 High Level Proof of Theorem 4.1.6

In this section, we give the high-level plan behind the proof of Theorem 4.1.6. This approach was developed in [DKK+18b], and then completed in [KMS18]. The task boils down to upper-bounding the fourth moment of the “projection of the indicator function $1_S$ onto the Fourier level-$r$”.

- The approach of [DKK+18b] relies on Fourier analysis on $\text{Gr}_{k,\ell}$, and as a result is unfriendly. It is futile to write down the eigenvectors explicitly, and one instead works with the eigenspace spanned by all eigenvectors with a specific eigenvalue, referred to as the “Fourier level".
– The \( r^{th} \) level has eigenvalue very close to \( 2^{-r} \), but there is a error term.

– The direct sum of the eigenspaces at levels \( 0, 1, \ldots, r \), has a clean description: it is spanned by the indicator functions of the subgraphs \( G_{k, \ell}[A, B] \) of co-order at most \( r \). However to get the hands on precisely the \( r^{th} \) eigenspace requires “subtracting” the contribution of the previous levels, which is somewhat messy.

For these and additional reasons, [DKK+18b] worked with approximations to all the quantities and formulas of interest. Extending these approximate formulas to higher Fourier levels \( r \geq 3 \) seems to incur error terms that are unaffordable.

We present here the approach taken in [KMS18], that circumvents both of these issues by working with a closely related graph \( H_{k, \ell} \) (see the definition below). We are able to write down exact recursive formulas relating quantities at successive Fourier levels. The eigenvalues are exactly \( 2^{-\ell} \) providing a hint that things would fall in place, but it still takes significant effort to develop the full Fourier analytic machinery, described in Section 4.3. Furthermore, the recursive (exact) formulas therein are systematic extensions, to higher Fourier levels, of the approximate formulas from [DKK+18b].

We remark that the main difficulty in the proof persists since we still have to consider a large number of elaborate cases. These cases can be analyzed systematically and simultaneously for all Fourier levels \( r \) (i.e. without the number of cases exploding with \( r \)).

We now define the new graph \( H_{k, \ell} \) and show that the task of proving Theorem 4.1.6 reduces to the task of proving an analogous theorem for \( H_{k, \ell} \) (Theorem 4.2.5). Then we describe the very basics of Fourier analysis, just enough to recall the high-level plan (from [DKK+18b]) that reduces the task to showing that the indicator function \( 1_S \) of a pseudorandom set \( S \subseteq H_{k, \ell} \) has low Fourier weight on low Fourier levels (see Section 4.2.6 and Theorem 4.2.13). This task is in turn reduced to that of upper-bounding the fourth moment of the Fourier level-\( r \) component of the indicator function \( 1_S \) (see Sections 4.2.7, 4.2.8), which finally is reduced to the main technical result of this chapter, Theorem 4.2.15, about upper-bounding closely related “4-wise correlations”.

### 4.2.3 Switching to the Graph \( H_{k, \ell} \)

**Definition 4.2.2.** Let \( 2 \leq \ell \leq k \) be integers. The vertices of the graph \( H_{k, \ell} \) are given by \((\{0, 1\}^k)^\ell\). The edges of the graph are best described by describing how to sample a uniformly random neighbor \( z \) of an arbitrary vertex \( x \). Fix a vertex \( x \in (\{0, 1\}^k)^\ell \) and write \( x = (x_1, \ldots, x_\ell) \) where \( x_1, \ldots, x_\ell \in \{0, 1\}^k \).

Sample \( y \leftarrow \{0, 1\}^k, b_1, \ldots, b_\ell \leftarrow \{0, 1\} \) independently and uniformly at random. Let the neighbor of \( x \) be \( z = (x_1 + b_1 \cdot y, x_2 + b_2 \cdot y, \ldots, x_\ell + b_\ell \cdot y) \).

The two graphs \( H_{k, \ell} \) and \( Gr_{k, \ell} \) are closely related as follows:

- The vertices of \( H_{k, \ell} \) are \( \ell \)-tuples of vectors in \( \mathbb{F}_2^k \). The vertices of \( Gr_{k, \ell} \) are \( \ell \)-dimensional subspaces of \( \mathbb{F}_2^k \), or equivalently, \( \ell \)-tuples of vectors in \( \mathbb{F}_2^k \) that are linearly independent and two tuples are considered the same if their vectors have the same linear span.

- When a random vertex \( x = (x_1, \ldots, x_\ell) \) in \( H_{k, \ell} \) is sampled and then a random edge incident on it is sampled by sampling \( y \leftarrow \{0, 1\}^k \) and \( b_1, \ldots, b_\ell \in \{0, 1\} \), with probability \( \approx 2^{-k} + 2^{-\ell} \), either \( y = 0 \) or \( b = (b_1, \ldots, b_\ell) = 0 \), and the edge is a self-loop. Otherwise \( y \neq 0, b \neq 0 \) and the other endpoint is \( z = (x_1 + b_1 \cdot y, \ldots, x_\ell + b_\ell \cdot y) \). Provided that both \( x, z \) have full \( \ell \)-dimensional linear span (which happens with probability \( \approx 1 - 2^{\ell-k} \), and we think of \( \ell \ll k \)), the edge \( (x, z) \) corresponds to a uniformly random edge of the Grassmann graph.
4.2.4 It Suffices to Work with $H_{k,\ell}$

We show that Theorem 4.1.6 for $Gr_{k,\ell}$ follows easily from the corresponding theorem for $H_{k,\ell}$ (Theorem 4.2.5), and then we work with the graph $H_{k,\ell}$ for the rest of the chapter. It will be convenient to restate Theorem 4.1.6 in the contra-positive and in terms of “pseudorandom sets”.

**Definition 4.2.3.** A subset of vertices $S \subseteq Gr_{k,\ell}$ is called $(r, \varepsilon)$-pseudorandom if for any subspaces $A \subseteq B \subseteq \mathbb{F}_2^k$ such that $\dim(A) = a, \text{codim}(B) = b, a + b \leq r$, we have

$$
\mu_{\text{in}(A), \text{out}(B)}(S) \overset{\text{def}}{=} \frac{|S \cap Gr_{k,\ell}[A, B]|}{|Gr_{k,\ell}[A, B]|} \leq \varepsilon.
$$

**Theorem 4.1.6 (Restated).** For every constant $\zeta > 0$, there exists a constant $\varepsilon > 0$ and an integer $r \geq 0$ such that for all sufficiently large integer $\ell$, and sufficiently large integer $k$, the following holds. If $S \subseteq Gr_{k,\ell}$ is $(r, \varepsilon)$-pseudorandom, then $\Phi(S) \geq 1 - \zeta$.

Now we show how to reduce this theorem to Theorem 4.2.5 below. The reasoning is straightforward. We will show that for every $S \subseteq Gr_{k,\ell}$, there is a natural corresponding set $S^* \subseteq H_{k,\ell}$ such that:

- **Lemma 4.2.7** below: If $S$ is $(r, \varepsilon)$-pseudorandom, then $S^*$ is $(r, \varepsilon)$-pseudorandom (for a similar notion of being pseudorandom in $H_{k,\ell}$ and up to a negligible additive loss in the parameter $\varepsilon$).

- **Theorem 4.2.5** below: If $S^*$ is $(r, \varepsilon)$-pseudorandom, then $\Phi(S^*) \geq 1 - \zeta$.

- **Lemma 4.2.6** below: $\Phi(S) = \Phi(S^*)$ (up to a negligible additive difference) and hence $\Phi(S) \geq 1 - \zeta$ as desired.

We elaborate on each of the three items. For a set $S \subseteq Gr_{k,\ell}$, the corresponding set $S^* \subseteq H_{k,\ell}$ is defined naturally as

$$
S^* \overset{\text{def}}{=} \{(x_1, \ldots, x_\ell) \mid \dim(\text{Span}(x_1, \ldots, x_\ell)) = \ell, \text{Span}(x_1, \ldots, x_\ell) \subseteq S\}.
$$

We note that $S^*$ is invariant under change of basis, i.e., if $\text{Span}(x_1, \ldots, x_\ell) = \text{Span}(y_1, \ldots, y_\ell)$, then $(x_1, \ldots, x_\ell) \subseteq S^*$ if and only if $(y_1, \ldots, y_\ell) \subseteq S^*$. We call such subsets of $H_{k,\ell}$ basis-invariant. We will only concern ourselves with basis-invariant subsets of $H_{k,\ell}$. We note moreover that tuples in $S^*$ are linearly independent (this is a minor issue; the only place where this is used is in the proof of Lemmas 4.2.6 and 4.2.7).

The notion of pseudorandom sets in $H_{k,\ell}$ is defined in a similar manner:

**Definition 4.2.4.** A basis-invariant subset of vertices $S^* \subseteq H_{k,\ell}$ is called $(r, \varepsilon)$-pseudorandom if for any sequence $Q = (x_1, \ldots, x_q)$ of points in $\mathbb{F}_2^k$ and a subspace $W \subseteq \mathbb{F}_2^k$ and $q + \text{codim}(W) \leq r$, we have

$$
\mu_{\text{in}(Q), \text{out}(W)}(S^*) \overset{\text{def}}{=} \Pr_{z_{q+1}, \ldots, z_\ell \in W} [(x_1, \ldots, x_q, z_{q+1}, \ldots, z_\ell) \in S^*] \leq \varepsilon.
$$

There is a slight difference between Definitions 4.2.3 and 4.2.4. In the latter, we allow $Q$ to be a sequence of points (so there can be linear dependencies among them) and we do not necessarily require that $Q \subseteq W$. This difference however has no significance and is to be ignored. The following is the main result in the chapter. As noted, together with Lemmas 4.2.6 and 4.2.7 below, it implies Theorem 4.1.6.

**Theorem 4.2.5.** For every constant $\zeta > 0$, there exists a constant $\varepsilon > 0$ and an integer $r \geq 0$ such that for all sufficiently large integers $\ell$ and (after fixing it) for all sufficiently large integers $k$, the following holds.

If $S^* \subseteq H_{k,\ell}$ is a basis-invariant $(r, \varepsilon)$-pseudorandom set, then $\Phi(S) \geq 1 - \zeta$. 

100
Lemma 4.2.6. 

\[ |\Phi(S^*) - \Phi(S)| \leq 2^{-\ell} + 2 \cdot 2^{\ell-k}. \]

**Proof.** Let \( x = (x_1, \ldots, x_\ell) \in S^* \), and denote a randomly chosen neighbor of it by \( z = (x_1 + b_1 \cdot y, \ldots, x_\ell + b_\ell \cdot y) \) where \( b = (b_1, \ldots, b_\ell) \). Then

\[ \Phi(S^*) = \Pr_{x \in S^*} \left[ z \not\in S^* \right] = 2^{-\ell} \cdot 0 + (1 - 2^{-\ell}) \Pr_{x \in S^*} \left[ z \not\in S^* \right]. \] (4.1)

Note that for any \( x \in S^* \), the vectors \( x_1, \ldots, x_\ell \) are linearly independent. Let \( L = \text{Span}(x_1, \ldots, x_\ell) \) and \( L' = \text{Span}(x_1 + b_1 \cdot y, \ldots, x_\ell + b_\ell \cdot y) \). Conditioned on \( y \not\in L, L' \) is \( \ell \)-dimensional subspace. Moreover, for \( b \neq 0 \), its distribution is uniform over all \( \ell \)-dimensional subspaces that intersect \( L \) in dimension \( \ell - 1 \). Therefore

\[ \Pr_{x \in S^*} \left[ z \in S^* \right] = \Pr_{x \in S^*} \left[ y \not\in L \right] \cdot \Pr_{x \in S^*} \left[ z \not\in S^* \right] + \Pr_{x \in S^*} \left[ y \in L \right] \cdot \Pr_{x \in S^*} \left[ z \not\in S^* \right]. \]

Since choosing \( x \in S^* \) uniformly at random corresponds to choosing \( L \in S \) uniformly at random and picking a random basis, we have that the first summand equals \((1 - 2^{\ell-k})\Phi(S)\). The second summand is at most \( \Pr_{x \in S^*} \left[ y \in L \right] \leq 2^{\ell-k} \). Combining everything completes the proof of the lemma.

**Lemma 4.2.7.** If \( S \) is \((r, \varepsilon)\)-pseudorandom, then \( S^* \) is \((r, \varepsilon + 2^{\ell+r-k})\)-pseudorandom.

**Proof.** Consider any sequence \( Q = (x_1, \ldots, x_q) \) of vectors from \( \mathbb{F}_2^k \) and a subspace \( W \subseteq \mathbb{F}_2^k \) such that \( q + \text{codim}(W) \leq r \). If \( Q \) is a linearly dependent set, then \( \mu_{\text{in}(Q), \text{out}(W)}(S^*) = 0 \) and there is nothing to prove. So assume that \( Q \) is linearly independent.

\[ \mu_{\text{in}(Q), \text{out}(W)}(S^*) = \Pr_{z \in S^*} \left[ (x_1, \ldots, x_q, z_{q+1}, \ldots, z_\ell) \in S^* \right]. \]

Denoting by \( \mathcal{E} \) the event that \( \{x_1, \ldots, x_q, z_{q+1}, \ldots, z_\ell\} \) are linearly independent, we have

\[ \mu_{\text{in}(Q), \text{out}(W)}(S^*) \leq \Pr[\mathcal{E}] \cdot \Pr_{z_{q+1}, \ldots, z_\ell \in W} \left[ (x_1, \ldots, x_q, z_{q+1}, \ldots, z_\ell) \in S^* \right] + \Pr \left[ \overline{\mathcal{E}} \right] \]

\[ = \Pr[\mathcal{E}] \cdot \mu_{\text{in}(\text{Span}(Q)), \text{out}(Q \oplus W)}(S) + \Pr \left[ \overline{\mathcal{E}} \right]. \]

The last equality follows from the fact that conditioned on \( \mathcal{E} \), \( L = \text{Span}(x_1, \ldots, x_q, z_{q+1}, \ldots, z_\ell) \) is distributed uniformly among all \( \ell \)-dimensional subspaces containing \( Q \) and contained in \( Q \oplus W \), and \( L \in S \) if and only if \( (x_1, \ldots, x_q, z_{q+1}, \ldots, z_\ell) \in S^* \). We conclude that \( \mu_{\text{in}(Q), \text{out}(W)}(S^*) \leq \varepsilon + 2^{\ell+r-k} \) by noting that \( S \) is \((r, \varepsilon)\)-pseudorandom and hence \( \mu_{\text{in}(\text{Span}(Q)), \text{out}(Q \oplus W)}(S) \leq \varepsilon \) and that

\[ \Pr_{z_{q+1}, \ldots, z_\ell \in W} [\mathcal{E}] \geq 1 - \sum_{i=q}^{\ell-1} 2^{i-(k-r)} \geq 1 - 2^{\ell+r-k}. \]

We note moreover that \( q + \text{codim}(Q \oplus W) \leq q + \text{codim}(W) \leq r \), so we may appeal to the \((r, \varepsilon)\)-pseudorandomness of \( S \).

**4.2.5 The Eigenvectors and Eigenvalues of \( H_{k,\ell} \) and Fourier Levels**

One advantage of working with the graph \( H_{k,\ell} \) is that its vertex set is the Boolean hypercube \( (\{0, 1\}^k)^\ell \), it is a Cayley graph, and determining its eigenvectors and eigenvalues is straightforward.
**Definition 4.2.8.** For $T_1, \ldots, T_k \in \{0,1\}^k$, define $\chi_{T_1,\ldots,T_k} : (\{0,1\}^k)^\ell \to \{-1,1\}$ by
\[
\chi_{T_1,\ldots,T_k}(x_1, \ldots, x_\ell) = \prod_{i=1}^\ell \chi_{T_i}(x_i),
\]
where $\chi_{T_i}(x_i) = (-1)^{T_i \cdot x_i}$ is the standard Fourier character (here $\cdot$ is the inner product over $\mathbb{F}_2$).

We denote by $H_{k,\ell}$ also the normalized transition matrix of the graph $H_{k,\ell}$ (i.e. its entry $(x, z)$ equals the probability that a random neighbor of $x$ equals $z$). We will be interested in the eigenvectors and eigenvalues of $H_{k,\ell}$. Since $H_{k,\ell}$ is a Cayley graph on the Boolean hypercube, its eigenvectors are precisely the characters $\chi_{T_1,\ldots,T_\ell}$.

**Lemma 4.2.9.** If $\dim(\text{Span}(T_1, \ldots, T_\ell)) = r$, then $\chi_{T_1,\ldots,T_\ell}$ is a eigenvector of $H_{k,\ell}$ with eigenvalue $2^{-r}$, i.e.
\[
H_{k,\ell} \cdot \chi_{T_1,\ldots,T_\ell} = 2^{-r} \cdot \chi_{T_1,\ldots,T_\ell}.
\]

**Proof.** Considering a random choice of $y \in \{0,1\}^k$ and $b = (b_1, \ldots, b_\ell) \in \{0,1\}^\ell$,
\[
H_{k,\ell} \cdot \chi_{T_1,\ldots,T_\ell}(x) = \mathbb{E}_{y,b} \left[ \chi_{T_1,\ldots,T_\ell}(x_1 + b_1 y_1, \ldots, x_\ell + b_\ell(y)) \right] = \chi_{T_1,\ldots,T_\ell}(x) \cdot \mathbb{E}_{y,b} \left[ \chi_{T_1,\ldots,T_\ell}(y) \right].
\]
The expectation over $y$ vanishes if $\oplus_{i=1}^\ell b_i T_i \neq 0$ and equals 1 otherwise. Since $\oplus_{i=1}^\ell b_i T_i$ is a uniformly random vector in $\text{Span}(T_1, \ldots, T_\ell)$, the probability over the choice of $b$ that $\oplus_{i=1}^\ell b_i T_i = 0$ is precisely $2^{-r}$.

**Definition 4.2.10.** (Clearly) any function $F: (\{0,1\}^k)^\ell \to \mathbb{R}$ can be written as
\[
F[x_1, \ldots, x_\ell] = \sum_{T_1,\ldots,T_\ell \in \{0,1\}^k} \hat{F}(T_1, \ldots, T_\ell) \cdot \chi_{T_1,\ldots,T_\ell}(x_1, \ldots, x_\ell).
\]
Its $i^{th}$ level component is defined as its projection onto the eigenspace with eigenvalue $2^{-i}$, i.e.
\[
F_{=i}[x_1, \ldots, x_\ell] = \sum_{\text{dim} \left( \text{Span}(T_1, \ldots, T_\ell) \right) = i} \hat{F}(T_1, \ldots, T_\ell) \cdot \chi_{T_1,\ldots,T_\ell}(x_1, \ldots, x_\ell).
\]
The decomposition $F = \sum_{i=0}^\ell F_{=i}$ into “Fourier levels” satisfies Parseval’s identity: $\|F\|_2^2 = \sum_{i=0}^\ell \|F_{=i}\|_2^2$.

**Definition 4.2.11.** A function $F: (\{0,1\}^k)^\ell \to \mathbb{R}$ is called basis-invariant if for every $x_1, \ldots, x_\ell \in \{0,1\}^k$ and an invertible $\ell \times \ell$ matrix $M$ over $\mathbb{F}_2$, we have that
\[
F[x_1, \ldots, x_\ell] = F[M(x_1, \ldots, x_\ell)].
\]
Here $M(x_1, \ldots, x_\ell) = (y_1, \ldots, y_\ell)$ such that $y_i = \sum_{j=1}^\ell M_{ij} x_j$.

In words, a function is basis-invariant if its value is preserved under invertible linear transformation of its arguments. All functions that we deal with in this chapter are basis invariant and in particular the indicators of sets $S \subseteq H_{k,\ell}$ that “arise” from corresponding sets in $\text{Gr}_{k,\ell}$.

The following definition generalizes the notion of pseudorandomness to functions. We note that for Boolean functions, under the natural identification between a set and its indicator function, the definitions coincide.

102
Definition 4.2.12. Let $F : \{\{0, 1\}^k\}^\ell \rightarrow \mathbb{R}$ be a basis invariant function. For a set of points $Q = \{x_1, \ldots, x_q\} \subseteq \{0, 1\}^k$ and subspace $W \subseteq \{0, 1\}^k$, define

$$\mu_{\text{in}}(Q, \text{out}(W)) = \mathbb{E}_{y_{q+1}, \ldots, y_r \in \{0, 1\}^k} \left[ F[x_1, \ldots, x_q, y_{q+1}, \ldots, y_r] \right].$$

The function $F$ is called $(r, \varepsilon)$ pseudo-random if for every $Q \subseteq \{0, 1\}^k$ and subspace $W \subseteq \{0, 1\}^k$ such that $|Q| + \text{codim}(W) \leq r$, it holds that $\mu_{\text{in}}(Q, \text{out}(W)) \leq \varepsilon$.

### 4.2.6 Pseudorandomness \(\Rightarrow\) Low Weight at Low Levels \(\Rightarrow\) Near-Perfect Expansion

Fix a basis-invariant set $S \subseteq \mathbb{H}_{k, \ell} = \{(0, 1)^k\}^\ell$. Let $F : \{(0, 1)^k\}^\ell \rightarrow \{0, 1\}$ be its indicator function. Let $\delta = \mu(F) = \|F\|_2^2$ denote its density and let $\|F_i\|_2^2$ be its “weight at the $i$th Fourier level”. We note that the weight at the $0$th level is $\delta^2$ and the sum of the weights at all Fourier levels equals $\delta$. Theorem 4.2.5 requires us to show that if $S$ is pseudorandom, then it has near prefect expansion. At a high-level, this is accomplished in two steps:

- One shows that a pseudorandom set must have low (say $\leq \zeta \delta$) weight at all lower (say up to $r$) levels.
- One shows that if there is low weight at all lower levels, then the set must have near perfect expansion ($\geq 1 - \zeta (r + 1) - 2^{-(r+1)}$).

We include below a quick proof of the second step (the main task remains thereafter to prove the first step). Assume therefore that $F$ has weight at most $\zeta \delta$ at each level up to $r$. Below a random neighbor of $x$ is denoted as $z \sim x$ and the inner product is $\langle F_1, F_2 \rangle = \mathbb{E}_z [F_1(x)F_2(x)]$. We have

$$1 - \Phi(S) = \Pr_{x \in S, z \sim x} [z \in S] = (1/\delta) \cdot \Pr_{x, z \sim x} [x \in S \land z \in S] = (1/\delta) \cdot \langle F, \mathbb{H}_{k, \ell} F \rangle.$$

Using the decomposition $F = \sum_{i=0}^\ell F_i$ into mutually orthogonal eigenspaces $F_i$ of eigenvalues $2^{-i}$, and that $\delta = \sum_{i=0}^\ell \|F_i\|_2^2$, we get that

$$\delta(1 - \Phi(S)) = \sum_{i=0}^\ell 2^{-i}\|F_i\|_2^2 \leq \sum_{i=0}^r \|F_i\|_2^2 + 2^{-(r+1)} \sum_{i=r+1}^\ell \|F_i\|_2^2 \leq \zeta \delta (r + 1) + \delta 2^{-(r+1)}.$$

Dividing by $\delta$ gives us $\Phi(S) \geq 1 - \zeta (r + 1) - 2^{-(r+1)}$ as claimed. To summarize, to prove Theorem 4.2.5, it suffices to prove:

**Theorem 4.2.13.** Let $S$ be a basis-invariant set of vertices in $\mathbb{H}_{k, \ell}$ that has density $\delta$ and is $(r, \varepsilon)$ pseudorandom. Let $F : \mathbb{H}_{k, \ell} = \{(0, 1)^k\}^\ell \rightarrow \{0, 1\}$ be the indicator function of $S$. Then for any $i = 0, 1, \ldots, r$,

$$\eta = \|F_i\|_2^2 \leq 2^{9r^2} \varepsilon^{1/\delta}.$$

We now summarize the high-level plan to prove Theorem 4.2.13. The idea is to consider the fourth moment of $F_i$ and prove both a lower bound and an upper bound on it. Specifically, let $S$ be a set that has density $\delta$ and is $(r, \varepsilon)$ pseudorandom as in the statement of the theorem. Let $0 \leq i \leq r$ and let $\eta = \|F_i\|_2^2$. The theorem follows by showing that (the expectation is over $x \in \{\{0, 1\}^k\}^\ell$; one cancels $\eta$ from both sides, moves $\delta^3$ on the right and then takes a cubic root)

$$\frac{\eta^4}{\delta^3} \leq \mathbb{E} [F_i^4] \leq 2^{20r^2} \eta \varepsilon.$$  

(4.2)
4.2.7 Lower-bounding the Fourth Moment of $F_{=i}$

**Lemma 4.2.14.** Under the condition and notation of Theorem 4.2.13, $\mathbb{E} \left[ F_{=i}^4 \right] \geq \frac{\eta^4}{\delta^3}$.

**Proof.** We note the decomposition $F = F_{=i} + (F - F_{=i})$ into mutually orthogonal components and that $\|F_{=i}\|^2 = \eta$. Hence, by Hölder’s Inequality

$$\eta = \langle F_{=i}, F_{=i} \rangle = \langle F_{=i}, F \rangle \leq \|F_{=i}\| \|F\|_{4/3}.$$

Since $\|F\|_{4/3} = \delta^{3/4}$, dividing by it and taking fourth power finishes the proof. \qed

4.2.8 Upper-bounding the Fourth Moment of $F_{=i}$

To summarize, the task of proving Theorem 4.2.13 is now reduced to proving the upper bound in Equation (4.2), i.e. under the condition and notation of Theorem 4.2.13, to prove that, for $0 \leq i \leq r$,

$$\mathbb{E} \left[ F_{=i}^4 \right] \leq 2^{25r^3} \eta \varepsilon,$$

where $\eta \overset{\text{def}}{=} \mathbb{E} \left[ F_{=i}^2 \right]$.

We next describe the first step of the proof. In Section 4.3 we take lengthy detour to develop the required analytic machinery; then in Section 4.5 return to the main line of the proof.

In Section 4.3, Lemma 4.3.13, establishes the following alternative characterization $F_{=i}$ (in addition to that of Definition 4.2.10): there exists a (unique) function $f_{=i} : \mathbb{R} \rightarrow \{0, 1\}^k$ such that for all $x = (x_1, \ldots, x_\ell) \in \{0, 1\}^\ell$,

$$F_{=i}[x] = \frac{1}{\beta_{i,i}} \sum_{M \in M[i, \ell]} f_{=i}(Mx).$$

Let’s explain the notation: here $M[i, \ell]$ is the set of all $i \times \ell$ matrices over $\mathbb{F}_2$ that have full row-rank $i$. For $M \in M[i, \ell], Mx \in \{0, 1\}^i$ is an $i$-tuple where $(Mx)_j = \sum_{t=1}^\ell M_{jt} x_t$. And $\beta_{i,i}$ is a normalizing factor that equals the number of invertible $i \times i$ matrices.

Using the alternative representation, we may compute (or rather upper bound) $\mathbb{E} \left[ F_{=i}^4 \right]$: take the sum to the fourth power, expand, and take the expectation over $x$:

$$\mathbb{E} \left[ F_{=i}^4 \right] = \frac{1}{\beta_{i,i}^4} \sum_{M_1, M_2, M_3, M_4 \in M[i, \ell]} \mathbb{E}_x \left[ f_{=i}(M_1 x) f_{=i}(M_2 x) f_{=i}(M_3 x) f_{=i}(M_4 x) \right].$$

We partition the sum according to the direct sum of row spaces of $M_1, \ldots, M_4$, that is, according to $A = \oplus_{s=1}^4 \text{rowspan}(M_s)$. We note that $A \subseteq \{0, 1\}^\ell$ is a subspace and $i \leq d = \text{dim}(A) \leq 4i$.

$$\mathbb{E} \left[ F_{=i}^4 \right] = \frac{1}{\beta_{i,i}^4} \sum_{d=i}^{4i} \sum_{A: \text{dim}(A) = d} \sum_{M_1, M_2, M_3, M_4 \in M[i, \ell]} \oplus_{s=1}^4 \text{rowspan}(M_s) = A \mathbb{E}_x \left[ f_{=i}(M_1 x) f_{=i}(M_2 x) f_{=i}(M_3 x) f_{=i}(M_4 x) \right].$$

The remaining task is to upper bound each individual expectation above. We note that the number of choices for $A$ is at most $2^{d^4}$ (the number of $d$-dimensional subspaces of an $\ell$-dimensional space). For a fixed $A$, the number of choices for each of $M_1, M_2, M_3, M_4$ is at most $2^{id}$, and $\beta_{i,i} \geq 1$. Hence, to show the desired upper bound of $2^{25r^3} \eta \varepsilon$ on the entire sum, it is sufficient to show an upper bound of $\frac{2^{25r^3} \eta \varepsilon}{2^{4i}}$ on each individual expectation. More precisely:
Theorem 4.2.15 (Main Technical Theorem). Let $F : H_{k,\ell} = (\{0, 1\}^k)^\ell \to \{0, 1\}$ be basis invariant, $(r, \varepsilon)$ pseudorandom function, and $\eta = \|F_{-i}\|^2_2$. Then for any $0 \leq i \leq r$, $i \leq d \leq 4i$, $A \subseteq \{0, 1\}^\ell$ of dimension $d$ and $M_1, \ldots, M_4 \in M[i, \ell]$ such that $\oplus_{s=1}^4 \text{rowspan}(M_s) = A$, we have that
\[
\mathbb{E}_{x \in (\{0, 1\}^k)^\ell} \left[ f_{-i}(M_1 x) f_{-i}(M_2 x) f_{-i}(M_3 x) f_{-i}(M_4 x) \right] \leq 2^{r \eta^3} \frac{\eta^\varepsilon}{2^d r}.
\]
(4.3)

4.3 Analytic Machinery

In this section, we present the Fourier analytic machinery required for the proof of Theorem 4.2.13. Lemmas 4.3.13, 4.3.19, 4.3.20 should be treated as the key ones, at least in the sense that these will be referred to and used directly in the main proof.

In what follows, $F : (\{0, 1\}^k)^\ell \to \mathbb{R}$ is a basis-invariant function in the sense of Definition 4.2.11. Much of what is stated applies to all such functions and not necessarily Boolean functions. However, the latter type of functions are the ones that we are mainly interested in, and the reader may assume that $F$ is of this type. We recall Definition 4.2.10 of the Fourier representation and the decomposition into Fourier levels, $F = \sum_{r=0}^\ell F_{=r}$:

\[
F[x_1, \ldots, x_\ell] = \sum_{T_1, \ldots, T_\ell \in (\{0, 1\}^k)^d} \tilde{F}(T_1, \ldots, T_\ell) \cdot \chi_{T_1, \ldots, T_\ell}(x_1, \ldots, x_\ell).
\]

\[
F_{=r}[x_1, \ldots, x_\ell] = \sum_{T_1, \ldots, T_\ell \in (\{0, 1\}^k)^d \text{dim(Span}(T_1, \ldots, T_\ell))=r} \tilde{F}(T_1, \ldots, T_\ell) \cdot \chi_{T_1, \ldots, T_\ell}(x_1, \ldots, x_\ell).
\]

Lemma 4.3.1. $\tilde{F}(T_1, \ldots, T_\ell)$ depends only on $\text{Span}(T_1, \ldots, T_\ell)$.

Proof. Suppose $\text{dim(Span}(T_1, \ldots, T_\ell)) = r$ and let $A_1, \ldots, A_r$ be a basis for the span. Then there is a $r \times \ell$ matrix of row-rank $r$ such that $(T_1, \ldots, T_\ell) = M^T(A_1, \ldots, A_r)$, where $M^T$ is the $\ell \times r$ transposed matrix. Moreover, in this case, defining vectors $(y_1, \ldots, y_r)$ such that $(y_1, \ldots, y_r) = M(x_1, \ldots, x_\ell)$,

\[
\prod_{j=1}^\ell \chi_{T_j}(x_j) = (-1)^{\ell \cdot r - \sum_{j=1}^\ell T_j \cdot x_j} = (-1)^{\ell \cdot r - \sum_{j=1}^\ell A_j \cdot y_j} = \prod_{s=1}^r \chi_{A_s}(y_s).
\]

We extend $M$ to an $\ell \times \ell$ invertible matrix $M'$ by appropriately appending $\ell - r$ rows. Let $(y_1, \ldots, y_r, y_{r+1}, \ldots, y_\ell) = M'(x_1, \ldots, x_\ell)$. It follows, using basis-invariance of $F$, that

\[
\tilde{F}(T_1, \ldots, T_\ell) = \mathbb{E}_{x_1, \ldots, x_\ell} \left[ F[x_1, \ldots, x_\ell] \prod_{j=1}^\ell \chi_{T_j}(x_j) \right] = \mathbb{E}_{x_1, \ldots, x_\ell} \left[ F[x_1, \ldots, x_\ell] \prod_{s=1}^r \chi_{A_s}(y_s) \right] = \mathbb{E}_{y_1, \ldots, y_\ell} \left[ F[y_1, \ldots, y_\ell] \prod_{s=1}^r \chi_{A_s}(y_s) \right] = \tilde{F}(A_1, \ldots, A_r, 0, \ldots, 0).
\]

Due to this lemma, we may write $\tilde{F}(T_1, \ldots, T_\ell)$ instead of $\tilde{F}(T_1, \ldots, T_\ell)$ if $\text{dim(Span}(T_1, \ldots, T_\ell)) = r$ and the first $r$ characters $T_1, \ldots, T_r$ are linearly independent.

105
4.3.1 High-level Perspective

Recall the goal outlined earlier: to show that for a pseudorandom set \( S \subseteq \mathbb{H}_{k,\ell} \), its indicator function \( F \) has low Fourier weight on low levels. Clearly, there are two notions of interest here:

- The zoom-in-out densities \( \mu_{\text{in}(Q),\text{out}(W)}(S) \).
  
  \( S \) is \( (r, \varepsilon) \)-pseudorandom if, by definition, all zoom-in-out densities, for \( |Q| + \text{codim}(W) \leq r \), are at most \( \varepsilon \).

- The Fourier level functions \( F_{=r} \).
  
  The Fourier weight at level \( r \) is, by definition, \( \|F_{=r}\|^2 \).

And then there is a third notion: as mentioned in Section 4.2.8, \( F_{=r} \) has an alternate characterization: there exists a (unique) function \( f_{=r} : (\{0,1\}^k)^r \to \mathbb{R} \) such that for all \( x = (x_1, \ldots, x_\ell) \in (\{0,1\}^k)^\ell \),

\[
F_{=i}[x] = \frac{1}{\beta_{i,i}} \sum_{M \in M[i,\ell]} f_{=i}(Mx).
\]

The functions \( f_{=r} \) play a crucial role in our analysis, however we refrain from naming them.

A large part of our Fourier analytic machinery is devoted to establishing relations between the three notions, \( \mu_{\text{in}(Q),\text{out}(W)}(S) \), \( F_{=r} \), \( f_{=r} \). Interestingly (and rather bafflingly), we only work with zoom-in densities \( \mu_{\text{in}(Q)}(S) \), and not with the zoom-out densities. The zoom-out densities come up indirectly, as Fourier sums of \( f_{=r} \) (see Lemmas 4.3.19, 4.3.20). The relationship between \( f_{=r} \) and the zoom-in densities is somewhat clearer, see Definition 4.3.5. Especially for \( r = 1 \), the relationship is immediate: \( f_{=1}(x_1) \) is precisely the change in density of the set \( S \) after zooming into point \( x_1 \). For higher levels \( r \), there is an inclusion-exclusion type formula that relates \( f_{=r} \) to densities of the set \( S \) after zooming into subspaces of dimension at most \( r \).

4.3.2 Zoom-out Restriction Lemma

In this section, we prove a recursive formula that relates the Fourier coefficients of \( F \) to those of restrictions of \( F \) to a hyperplane.

**Definition 4.3.2.** Let \( F : (\{0,1\}^k)^\ell \to \mathbb{R} \) be a function. For a subspace \( W \subseteq \{0,1\}^k \), we define the function \( F_W : W^\ell \to \mathbb{R} \) to be the restriction of \( F \) to \( W^\ell \) (referred to as the zoom-out function).

**Definition 4.3.3.** For a character \( T \), the subspace orthogonal to \( T \) is \( W_T = \{ x \in \{0,1\}^k \mid T \cdot x = 0 \} \).

**Lemma 4.3.4.** Let \( A, T_1, \ldots, T_r \) be linearly independent characters. Then

\[
\hat{F}(A, T_1, \ldots, T_r) = \frac{1}{2^\ell - 2^r} \left( \hat{F}_{W_A}(T_1, \ldots, T_r) - \sum_{D \subseteq \text{Span}(A,T_1,\ldots,T_r) \atop \dim(D) = r, A \notin D} \hat{F}(D) \right).
\]

**Proof.** The lemma follows from a direct computation below. In the third step we use the Fourier representation of \( F \) and in the fourth step we use the observation that for a character \( R \), the expectation \( \mathbb{E}_{y \in W_A} \left[ \chi_R(y) \right] \) equals 1 when \( R = 0 \) or when \( R = A \) and vanishes otherwise.
Defining $f_{\ell}$

For integers $\ell$, the remaining subsections are devoted to this derivation. Sometimes we write $f_{\ell}$ as $f = f_{\ell}$. Moreover each subspace of this type is counted exactly once.

\[ \ell \defeq 2^\ell - 2^r \]

The last equality is justified as follows. There are $2^\ell$ terms in the summation that split into two groups:

- In $2^\ell - 2^r$ terms, there is some $j \geq r + 1$ such that $Q_j = A$. In this case, $\text{Span}(Q_1, \ldots, Q_\ell)$ is the same as $\text{Span}(A, T_1, \ldots, T_r)$ and since the Fourier coefficients depend only on this span, $\hat{F}(Q_1, \ldots, Q_\ell) = \hat{F}(A, T_1, \ldots, T_r)$.

- For the remaining $2^r$ terms, for all $j \geq r + 1$, $Q_j = 0$. In this case, $\text{Span}(Q_1, \ldots, Q_\ell)$ is the same as $\text{Span}(Q_1, \ldots, Q_r)$ which is an $r$-dimensional subspace of $\text{Span}(A, T_1, \ldots, T_r)$ that does not contain $A$. Moreover each subspace of this type is counted exactly once.

4.3.3 Defining $f_{\ell}$ and Relating $F_{\ell}$, $f_{\ell}$ and Zoom-in Densities

For a (basis-invariant) function $F : \{0, 1\}^k \to \mathbb{R}$, we have the decomposition $F = \sum_{r=0}^{\ell} F_{\ell}$ where

\[ F_{\ell}[x_1, \ldots, x_\ell] = \sum_{T_1, \ldots, T_\ell \subseteq \{i\}^k} \hat{F}(T_1, \ldots, T_\ell) \cdot \chi_{T_1, \ldots, T_\ell}(x_1, \ldots, x_\ell). \]

As mentioned, we will need an alternate formula for $F_{\ell} : \{0, 1\}^k \to \mathbb{R}$ in terms of related functions $f_{\ell} : \{0, 1\}^r \to \mathbb{R}$. Deriving this formula turns out to be rather cumbersome (but still illuminating at the same time). Next few subsections are devoted to this derivation. Sometimes we write $f_{\ell, F}$ to make the relation to $F$ explicit. We will use the following notations:

- For integers $1 \leq i \leq r$, $\mathcal{M}[i, r]$ denotes the set of $i \times r$ matrices over $\mathbb{F}_2$ with (row)-rank $i$. We have $|\mathcal{M}[i, r]| = \prod_{j=0}^{i-1} (2^r - 2^j) \defeq \beta_{i, r}$.

- For $r \geq 0$, define $\beta_{0, r} = 1$ (considering $\{0\}$ as the only matrix in $\mathcal{M}[0, r]$).

- For $x = (x_1, \ldots, x_r)$ and $M \in \mathcal{M}[i, r]$, $Mx$ denotes the tuple $(y_1, \ldots, y_i)$ where $y_j = \sum_{t=1}^{r} M_{jt} x_t$.

Defining $f_{\ell}$ in terms of Zoom-in Densities

**Definition 4.3.5.** For $0 \leq r \leq \ell$, define $f_{\ell} : \{0, 1\}^r \to \mathbb{R}$ inductively as

\[ f_{\ell}(\{0\}) = \mu(F) \defeq \mathbb{E}_{x_1, \ldots, x_\ell} [F[x_1, \ldots, x_\ell]] , \]

107
\[
    f_{=r}(x_1, \ldots, x_r) = \mu_{in}(\{x_1, \ldots, x_r\})(F) - \sum_{d=0}^{r-1} \frac{1}{\beta_{d,d}} \sum_{M \in \mathcal{M}[d,r]} f_{=d}(Mx).
\]

We note that

- Recall that \( \mu_{in}(\{x_1, \ldots, x_r\})(F) = \mathbb{E}_{z_{r+1}, \ldots, z_\ell} [F[x_1, \ldots, x_r, z_{r+1}, \ldots, z_\ell]] \) is the zoom-in density (as in Definition 4.2.12).

- In the above definition, the term corresponding to \( d = 0 \) equals \( \mu(F) \).

- In the case \( r = 1 \),
  \[
    f_{=1}(x_1) = \mu_{in}(\{x_1\})(F) - \mu(F).
\]

**Lemma 4.3.6.** If \( F: \{0,1\}^{k\ell} \to \mathbb{R} \) is basis-invariant, then for every \( r \geq 0 \) the function \( f_{=r} \) is basis-invariant. I.e. for every \( x = (x_1, \ldots, x_r) \in (\{0,1\}^k)^r \) and \( M \in \mathcal{M}[r,r] \), we have that
  \[
    f_{=r}(x) = f_{=r}(Mx).
\]

**Proof.** By induction on \( r \). For \( r = 0,1 \), this is trivial. Let \( r \geq 2 \) and fix \( x_1, \ldots, x_r \) and an \( r \times r \) invertible matrix \( M \) as in the claim. By definition

\[
    f_{=r}(Mx) = \mu_{in}(Mx)(F) - \sum_{d=0}^{r-1} \frac{1}{\beta_{d,d}} \sum_{M' \in \mathcal{M}[d,r]} f_{=d}(M'Mx).
\]

Observe that for any \( d \geq 0 \), the mapping \( M' \mapsto M'M \) is a bijection on \( \mathcal{M}[d,r] \). Also observe that \( \mu_{in}(Mx)(F) = \mu_{in}(x)(F) \), since \( F \) is basis invariant. Thus the last expression equals

\[
    \mu_{in}(x)(F) - \sum_{d=0}^{r-1} \frac{1}{\beta_{d,d}} \sum_{M' \in \mathcal{M}[d,r]} f_{=d}(M'x) = f_{=r}(x).
\]

\[ \square \]

**Zoom-in Restriction Lemma**

**Definition 4.3.7.** Let \( F: (\{0,1\}^k)^\ell \to \mathbb{R} \) be a function and let \( Q = \{a_1, \ldots, a_j\} \subseteq \{0,1\}^k \) where \( j \leq \ell - 1 \). The function \( F_Q: (\{0,1\}^{k\ell-j} \to \mathbb{R} \) (the zoom-in restriction function) is defined by

\[
    F_Q[x_1, \ldots, x_{\ell-j}] = F[a_1, \ldots, a_j, x_1, \ldots, x_{\ell-j}].
\]

We have the following recursive formula for \( f_{=r} \). Here \( e_1 \) refers to a vector with the first coordinate 1 and all other coordinates zero.

**Lemma 4.3.8.** Let \( F: (\{0,1\}^k)^\ell \to \mathbb{R} \) and \( r \geq 0 \). Let \( D = (a, x_1, \ldots, x_r) = (a, x) \). Then

\[
    f_{=r+1,F}(D) = f_{=r,F(e_1)}(x) - \frac{1}{\beta_{r,r}} \sum_{M' \in \mathcal{M}[r,r+1]} \sum_{e_1 \notin \text{rowspan}(M')} f_{=r,F}(M'D).
\]

108
Proof. The proof is by induction. The case \( r = 0 \) is trivial, both sides being \( \mu_{\text{in}}(\{a\})(F) - \mu(F) \), so assume \( r \geq 1 \). For brevity, we write \( f_{=r} \) instead of \( f_{=r,F} \) (but do write \( f_{=r,F_a} \) when working with the zoom-in function). From Definition 4.3.5,

\[
f_{=r+1}(D) = \mu_D(F) - \mu(F) - \sum_{j=1}^{r} \frac{1}{\beta_{j,j}} \sum_{M^* \in \mathcal{M}[j,r+1]} f_{=j}(M^* D).
\]

Using \( D = (a, x) \), \( \mu_D(F) = \mu_{\text{in}(x)}(F_a) \) and splitting the summation into two parts, we get

\[
f_{=r+1}(D) = \mu_{\text{in}(x)}(F_a) - \mu(F) - \sum_{j=1}^{r} \frac{1}{\beta_{j,j}} \left( \sum_{M^* \in \mathcal{M}[j,r+1]} f_{=j}(M^* D) \right) \left( \sum_{M^* \in \mathcal{M}[j,r+1]} f_{=j}(M^* D) \right)
\]

For fixed \( j \), let’s call the two terms above \( \Gamma_j \) and \( \Delta_j \) respectively. Below, the computation of \( \Gamma_j \) results in an “extra” \( -\Delta_j \) term that cancels with the previous \( \Delta \)-term in a telescoping manner.

\[
\Gamma_j = \frac{1}{\beta_{j,j}} \sum_{M^* \in \mathcal{M}[j,r+1] \atop e_1 \in \text{row-span}(M^*)} f_{=j}(M^* D) = \beta_{j,j+1} \frac{2^{j+1} - 1}{2^{r+1} - 1} \sum_{M^* \in \mathcal{M}[j,r+1] \atop e_1 \in \text{row-span}(M^*)} \frac{\mathbb{E}}{M^* \in \mathcal{M}[j,r+1] \atop e_1 \in \text{row-span}(M^*)} [f_{=j}(M^* D)]
\]

where we replaced summation by expectation (with appropriate normalization factor), and then used the definition of the \( \beta \)-parameters. The normalization factor is justified by observing that there are \( \beta_{j,r+1} \) matrices in \( \mathcal{M}[j, r + 1] \) and a fraction \( \frac{2^{j+1} - 1}{2^{r+1} - 1} \) of them will contain \( e_1 \) in their row-span (all non-zero vectors being symmetric in this regard).

Using Lemma 4.3.9, we see that \( M^* \) can be sampled by picking \( M \in \mathcal{M}[j-1, r] \) randomly, constructing \( M' \) from \( M \), sampling \( M'' \), and then letting \( M^* = M'' \cdot M'. \) Since \( M'' \) is invertible and \( f_{=j} \) is basis invariant,

\[
f_{=j}(M^* D) = f_{=j}(M''M'D) = f_{=j}(M'D) = f_{=j}(M'(a, x)) = f_{=j}(a, M.x).
\]

Hence the expectation in (4.5) is equal to \( \mathbb{E}_M[f_{=j}(a, M.x)] \). Applying the induction hypothesis (note that \( M.x \) is a \( (j-1) \)-tuple):

\[
\mathbb{E}_{M \in \mathcal{M}[j-1, r]} \left[ f_{=j-1,F_a}(M.x) - \frac{1}{\beta_{j-1,j-1}} \sum_{M'' \in \mathcal{M}[j-1,j] \atop e_1 \in \text{row-span}(M'')} f_{=j-1}(M''(a, M.x)) \right]
\]

\[
= \mathbb{E}_{M \in \mathcal{M}[j-1, r]} \left[ f_{=j-1,F_a}(M.x) - \frac{1}{\beta_{j-1,j-1}} \sum_{M'' \in \mathcal{M}[j-1,j] \atop e_1 \in \text{row-span}(M'')} \frac{2^j - 2^{j-1}}{2^j - 1} \mathbb{E}_{M \in \mathcal{M}[j-1, r]} \left[ f_{=j-1}(M''(a, M.x)) \right] \right],
\]

where the normalization factor is justified as before. Using Lemma 4.3.10, the distribution of \( M^* = M''M' \) here (\( M' \) is constructed from \( M \) as in the lemma) is uniform over matrices in \( \mathcal{M}[j-1, r+1] \) whose row-span does not contain \( e_1 \). It is also observed that

\[
f_{=j-1}(M'(a, M.x)) = f_{=j-1}(M''M'(a, x)) = f_{=j-1}(M^* D).
\]
Using the definition of the $\beta$-parameters, the expression can be rewritten as

\[
\mathbb{E}_{M \in \mathcal{M}[j-1, r]} [f_{=j-1, F_a}(Mx)] - 2^{j-1} \mathbb{E}_{M^* \in \mathcal{M}[j-1, r+1]}^{e_1 \notin \text{rowspan}(M^*)} [f_{=j-1}(M^*D)].
\]

Substituting in (4.5), we get

\[
\Gamma_j = \frac{\beta_{j-1,r}}{\beta_{j-1,j-1}} \mathbb{E}_{M \in \mathcal{M}[j-1, r]} [f_{=j-1, F_a}(Mx)] - \frac{\beta_{j-1,r}}{\beta_{j-1,j-1}} \cdot 2^{j-1} \mathbb{E}_{M^* \in \mathcal{M}[j-1, r+1]} [f_{=j-1}(M^*D)]
\]

\[
= \frac{1}{\beta_{j-1,j-1}} \sum_{M \in \mathcal{M}[j-1, r]} f_{=j-1, F_a}(Mx) - \frac{1}{\beta_{j-1,j-1}} \sum_{e_1 \notin \text{rowspan}(M^*)} f_{=j-1}(M^*D)
\]

\[
= \frac{1}{\beta_{j-1,j-1}} \sum_{M \in \mathcal{M}[j-1, r]} f_{=j-1, F_a}(Mx) - \Delta_{j-1}.
\]

Substituting in (4.4), telescoping, and noting that $\Delta_0 = \mu(F)$ (one can think of $\Delta_0$ as the “extra” term while calculating $\Gamma_1$ as above), we get

\[
f_{=r+1}(D) = \left( \mu_{\text{in}(x)}(F_a) - \mu(F) + \Delta_0 - \sum_{j=0}^{r-1} \frac{1}{\beta_{j,j}} \sum_{M \in \mathcal{M}[j, r]} f_{=j, F_a}(Mx) \right) - \Delta_r
\]

\[
= f_{=r, F_a}(x) - \Delta_r,
\]

completing the proof. \qed

**Some Auxiliary Lemmas**

**Lemma 4.3.9.** A uniformly random matrix $M^*$ in $\mathcal{M}[j, r + 1]$ whose row-span contains the vector $e_1$ can be sampled as:

- Pick a uniformly random matrix $M \in \mathcal{M}[j - 1, r]$.
- Augment $M$ to a matrix $M' \in \mathcal{M}[j, r + 1]$ so that $M'$ has top-left corner entry 1; the rest of the entries in the leftmost column and the top row are 0. The remaining entries of $M'$ are $M$'s entries (i.e., deleting the leftmost column and top row of $M'$ yields $M$).
- Pick a uniformly random matrix $M'' \in \mathcal{M}[j, j]$ and output $M^* = M'' \cdot M'$.

**Proof.** Let $W$ be the $r$-dimensional subspace of $\mathbb{F}_2^{r+1}$ consisting of vectors whose first coordinate is 0. Clearly, a random $j$-dimensional subspace $L' \subseteq \mathbb{F}_2^{j+1}$ that contains $e_1$ is obtained by picking a random $(j - 1)$-dimensional subspace $L \subseteq W$ and letting $L' = \text{Span}(e_1) \oplus L$. Writing a random basis of $L$ as rows of a matrix yields $M$. Writing $e_1$ followed by a random basis of $L$ as rows of a matrix yields $M'$ and its row-span equals $L'$. It follows that the row-span of $M'$ is a random $j$-dimensional subspace of $\mathbb{F}_2^{r+1}$ containing $e_1$.

Multiplying $M'$ by a random invertible matrix $M''$ yields the matrix $M^*$ whose rows now form a random basis of a random $j$-dimensional subspace of $\mathbb{F}_2^{r+1}$ containing $e_1$ as claimed. \qed

**Lemma 4.3.10.** A uniformly random matrix $M^*$ in $\mathcal{M}[j - 1, r + 1]$ whose row-span does not contain the vector $e_1$ can be sampled as (the two incarnations of $e_1$ in the statement of this lemma are different):
Proof. Let \( W, L, L' \) be as in the proof of the previous lemma. As therein, \( L' \) is a random \( j \)-dimensional subspace of \( \mathbb{F}_2^{r+1} \) that contains \( e_1 \) and the row-span of \( M' \) equals \( L' \) and its rows are \( v_1 = e_1 \) followed by a basis \( v_2, \ldots, v_j \) of \( L \). From Lemma 4.3.11 below, the rows of \( M^* = M'' \cdot M' \) then form a random basis of a random \( (j-1) \)-dimensional subspace of \( L' \) that does not contain \( v_1 = e_1 \). □

**Lemma 4.3.11.** Let \( v_1, \ldots, v_j \) be vectors that are linearly independent (over \( \mathbb{F}_2 \)). Let \( M'' \) be a uniformly random matrix in \( \mathcal{M}[j-1,j] \) whose row-span does not contain the vector \( e_1 \). Let \[
(w_1, \ldots, w_{j-1}) = M'' \cdot (v_1, \ldots, v_j).
\]
Then \( (w_1, \ldots, w_{j-1}) \) is a random basis of a random \((j-1)\)-dimensional subspace of \( \text{Span}(v_1, \ldots, v_j) \) that does not contain \( v_1 \).

**Proof.** It is clear that

- Since the rows of \( M'' \) are linearly independent, so are \( w_1, \ldots, w_{j-1} \).
- The matrix \( M'' \) and the tuple \( (w_1, \ldots, w_{j-1}) \) determine each other.
- \( e_1 \not\in \text{rowspan}(M'') \) if and only if \( v_1 \not\in \text{Span}(w_1, \ldots, w_{j-1}) \).

Thus there is a one-to-one correspondence between matrices \( M'' \) in \( \mathcal{M}[j-1,j] \) whose row-span does not contain the vector \( e_1 \) and tuples \( (w_1, \ldots, w_{j-1}) \) that span a \((j-1)\)-dimensional subspace of \( \text{Span}(v_1, \ldots, v_j) \) that does not contain the vector \( v_1 \). □

### Relating \( F_{=r} \) to Zoom-in Densities

**Lemma 4.3.12.** For every \( 0 \leq r \leq \ell \) and \( x_1, \ldots, x_\ell \in \{0,1\}^k \),
\[
\left(\sum_{i=0}^{r} \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,\ell}} F_{=i}\right) [x_1, \ldots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} \sum_{y \in \mathcal{M}_x} \mu_{in(y_1, \ldots, y_r)}(F).
\]

**Proof.** By definition,
\[
F_{=i}[x_1, \ldots, x_\ell] = \sum_{\text{dim}(T_1, \ldots, T_\ell) = i} \hat{F}(T_1, \ldots, T_\ell) \cdot \chi_{T_1, \ldots, T_\ell}(x_1, \ldots, x_\ell)
\]
\[
= \sum_{\text{dim}(D) = i} \hat{F}(Q_1, \ldots, Q_i, Q_{i+1}, \ldots, Q_\ell) \sum_{\text{Span}(Q_1, \ldots, Q_\ell) = D} \prod_{j=1}^{\ell} \chi_{T_j}(x_j),
\]
where the outer summation is over all \( i \)-dimensional subspaces \( D \) and for given \( D, (Q_1, \ldots, Q_i) \) is a specific ordered basis for it, and \( Q_{i+1} = \cdots = Q_\ell = 0 \). We used the fact that the Fourier coefficient depends only
on the span of its arguments. For given $D$, consider the inner sum. It is not difficult to see that all $\ell$-tuples $(T_1, \ldots, T_\ell)$ such that $\text{Span}(T_1, \ldots, T_\ell) = D$ are obtained precisely as
\[(T_1, \ldots, T_\ell) = M^{T r}(Q_1, \ldots, Q_i)\]
where $M^{T r}$ is a $\ell \times i$ matrix that is a transpose of a matrix $M \in \mathcal{M}[i, \ell]$. Moreover, in that case, defining vectors $(y_1, \ldots, y_i)$ such that $(y_1, \ldots, y_i) = M(x_1, \ldots, x_\ell)$ (which we abbreviate as $y = Mx$)
\[
\prod_{j=1}^\ell \chi_{T_j}(x_j) = (-1)^{\otimes_{j=1}^\ell T_j \cdot x_j} = (-1)^{\otimes_{s=1}^i Q_s \cdot y_s} = \prod_{s=1}^i \chi_{Q_s}(y_s).
\]
Thus we can write
\[
F_{=i}[x_1, \ldots, x_\ell] = \sum_{\dim(D) = i} \sum_{M \in \mathcal{M}[i, \ell]} \sum_{y = Mx} \prod_{j=1}^i \chi_{Q_j}(y_j).
\]
Using the definition of $\hat{F}(Q_1, \ldots, Q_i, Q_{i+1} = 0, \ldots, Q_\ell = 0)$ and interchanging the order of summation,
\[
F_{=i}[x_1, \ldots, x_\ell] = \sum_{\dim(D) = i} \sum_{M \in \mathcal{M}[i, \ell]} \sum_{y \in Mx} \prod_{j=1}^i \chi_{Q_j}(y_j)
\]
\[
= \sum_{\dim(D) = i} \sum_{M \in \mathcal{M}[i, \ell]} \sum_{y \in Mx} \prod_{j=1}^i \chi_{Q_j}(y_j)
\]
\[
= \frac{\beta_{i, \ell}}{\beta_{r, \ell}} \sum_{M \in \mathcal{M}[r, \ell]} \sum_{y \in Mx} \prod_{j=1}^i \chi_{Q_j}(y_j)
\]
Note that in the last line, the summation is over $\ell \times r$ matrices instead of $i \times \ell$ matrices and out of the vectors $(y_1, \ldots, y_r)$, only the first $i$ vectors are “used”. The normalization factor takes into account the number of matrices of the two different sizes. For a randomly chosen $\ell \times r$ invertible matrix $M'$, consider the change of basis $y = M'y'$, $(z_1', \ldots, z_r') = M'(z_1, \ldots, z_r)$ and $(A_1, \ldots, A_r) = M'^{T r}(Q_1, \ldots, Q_i, Q_{i+1} = 0, \ldots, Q_r = 0)$. By similar reasoning as before
\[
\prod_{j=1}^i \chi_{Q_j}(y_j \oplus z_j) = \prod_{j=1}^r \chi_{Q_j}(y_j \oplus z_j) = \prod_{j=1}^r \chi_{A_j}(y_j' \oplus z_j').
\]
Since $F$ is basis-invariant and the distribution of $y$ and $y'$ is the same, we may as well write the above equation as
\[
F_{=1}[x_1, \ldots, x_\ell] = \frac{\beta_{i, \ell}}{\beta_{r, \ell}} \sum_{M \in \mathcal{M}[r, \ell]} \sum_{y \in Mx} \prod_{j=1}^r \chi_{A_j}(y_j' \oplus z_j').
\]
In the above expression, first an $i$-dimensional subspace $D$ is chosen along with a fixed ordered basis $Q_1, \ldots, Q_i$ and then $(A_1, \ldots, A_r) = M'^{T r}(Q_1, \ldots, Q_i, Q_{i+1} = 0, \ldots, Q_r = 0)$ for a random $\ell \times r$ invertible matrix $M'$. Up to a factor of $\frac{1}{\beta_{r, \ell}}$, one can instead consider a summation over all $M'$, and then every
tuple \((A_1, \ldots, A_r)\) such that \(\dim(A_1, \ldots, A_r) = i\) occurs exactly \(\frac{\beta_{r,i}}{\beta_{r,r}}\) times. Hence the above equation can be written as

\[
F_{=i}[x_1, \ldots, x_\ell] = \frac{\beta_i,\ell}{\beta_r,\ell} \frac{1}{\beta_{r,i}} \sum_{(M \in \mathcal{M}[r,\ell]) \atop y = Mx} \mathbb{E} \left[ F(z_1, \ldots, z_\ell) \sum_{\dim(A_1, \ldots, A_r) = i} \prod_{j=1}^r \chi_{A_j}(y_j \oplus z_j) \right].
\]

Moving the \(\beta\)-factors to the left hand side and summing over \(i = 0, 1, \ldots, r\) counts every \(r\)-tuple \((A_1, \ldots, A_r)\) exactly once (irrespective of its dimension). Hence

\[
\left( \sum_{i=0}^r \beta_i,\ell \frac{\beta_{r,i}}{\beta_i,\ell} F_{=i} \right) [x_1, \ldots, x_\ell] = \sum_{(M \in \mathcal{M}[r,\ell]) \atop y = Mx} \mathbb{E} \left[ F(z_1, \ldots, z_\ell) \sum_{A_1, \ldots, A_r; j=1}^r \chi_{A_j}(y_j \oplus z_j) \right].
\]

We observe finally that the inner sum equals \(2^{kr}\) if \(z_j = y_j\) for \(1 \leq j \leq r\) and vanishes otherwise. In the former case, we can “fix” \(z_j = y_j\) for \(1 \leq j \leq r\) and drop the \(2^{kr}\) factor (since \(2^{kr}\) is the probability that randomly chosen \(z_j\) happens to equal \(y_j\) for \(1 \leq j \leq r\)). This yields

\[
\left( \sum_{i=0}^r \beta_i,\ell \frac{\beta_{r,i}}{\beta_i,\ell} F_{=i} \right) [x_1, \ldots, x_\ell] = \sum_{(M \in \mathcal{M}[r,\ell]) \atop y = Mx} \mathbb{E} \left[ F(y_1, \ldots, y_r, z_{r+1}, z_\ell) \right].
\]

The proof of Lemma 4.3.12 is completed by dividing both sides by \(\beta_{r,r}\) and noting that the expectation is precisely \(\mu_{in(y=Mx)}(F)\).

**Relating \(F_{=r}\) and \(f_{=r}\)**

The following lemma establishes an important alternative definition of \(F_{=r}\) as discussed earlier. We consider it as one of the key lemmas in this section (at least in the sense that it is used numerous times in the proof of our main result in this chapter).

**Lemma 4.3.13.** For every \(0 \leq r \leq \ell\) and \(x_1, \ldots, x_\ell \in \{0, 1\}^k\),

\[
F_{=r}[x_1, \ldots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} f_{=r}(Mx).
\]

**Proof.** The proof is by induction on \(r\). The case \(r = 0\) is trivial (both sides equal \(\mu(F)\)). Otherwise, using Lemma 4.3.12 we have

\[
\left( \sum_{i=0}^r \beta_i,\ell \frac{\beta_{r,i}}{\beta_i,\ell} F_{=i} \right) [x_1, \ldots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} \mu_{in(Mx)}(F).
\]

We note that the coefficient of \(F_{=r}\) on the left side is 1. Therefore we get

\[
F_{=r}[x_1, \ldots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} \mu_{in(Mx)}(F) - \left( \sum_{i=0}^{r-1} \frac{\beta_i,\ell \beta_{r,i}}{\beta_{r,r} \beta_i,\ell} F_{=i} \right) [x_1, \ldots, x_\ell].
\]
Using the induction hypothesis, we get that

\[
\left( \sum_{i=0}^{r-1} \beta_{i,r} \beta_{r,\ell} F_{i} \right) \left[ x_1, \ldots, x_\ell \right] = \sum_{i=0}^{r-1} \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,i}} \sum_{M \in \mathcal{M}[i,\ell]} f_{i}(M x)
\]

\[
= \frac{1}{\beta_{r,r}} \sum_{i=0}^{r-1} \beta_{i,i} \sum_{Q \in \mathcal{M}[i,r]} f_{i}(Q x). \tag{4.7}
\]

The last equality follows by observing that a full row-rank \( i \times \ell \) matrix \( M \) can be obtained as a product of full row-rank \( i \times r \) and \( r \times \ell \) matrices \( Q \) and \( A \) respectively (in uniform manner). In both summations, each \( M \) is counted exactly \( \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r}} \) times. Substituting (4.7) into (4.6) yields

\[
F_{r} \left[ x_1, \ldots, x_\ell \right] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} \mu_\text{in}(M x)) \left( F \right) - \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} \sum_{i=0}^{r-1} \frac{1}{\beta_{i,i}} \sum_{Q \in \mathcal{M}[i,r]} f_{i}(Q M x)
\]

\[
= \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} f_{r}(M x),
\]

where in the last equality we combined the two sums over \( M \), and used Definition 4.3.5 of \( f_{r} \), thereby finishing the inductive step. \( \square \)

### 4.3.4 Relating \( F_{=r} \) and \( f_{=r} \) in the Fourier Domain

**Lemma 4.3.14.** Let \( 0 \leq j \leq r - 1 \) and let \( a_1, \ldots, a_j \in \{0, 1\}^k \). Then

\[
\mathbb{E}_{y_{j+1}, \ldots, y_r \in \{0, 1\}^k} \left[ f_{=r}(a_1, \ldots, a_j, y_{j+1}, \ldots, y_r) \right] = 0.
\]

**Proof.** The proof is by double induction, first by induction on \( r \) with \( j = 0 \), and then by induction on \( j \) (as long as \( j \leq r - 1 \)). So assume first that \( j = 0 \). In the case \( r = 1 \),

\[
\mathbb{E}_{y_1} \left[ f_{=1}(y_1) \right] = \mathbb{E}_{y_1} \left[ \mu_\text{in}(y_1)) \left( F \right) - \mu(F) \right] = 0.
\]

Now assume \( j = 0 \) and \( r \geq 2 \).

\[
\mathbb{E}_{y_1, \ldots, y_r} \left[ f_{=r}(y_1, \ldots, y_r) \right] = \mathbb{E}_{y_1, \ldots, y_r} \left[ \mu_\text{in}(y_1, \ldots, y_r)) \left( F \right) - \sum_{i=0}^{r-1} \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i,r]} f_{i}(M y) \right]
\]

\[
= -\sum_{i=1}^{r-1} \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i,r]} \mathbb{E}_{y_1, \ldots, y_r} \left[ f_{=i}(M y) \right],
\]

where we used the fact that in the summation, the term with index \( i = 0 \) is \( \mu(F) \) and

\[
\mathbb{E}_{y_1, \ldots, y_r} \left[ \mu_\text{in}(y_1, \ldots, y_r)) \left( F \right) \right] = \mu(F)
\]

as well. We note that for any \( 1 \leq i \leq r - 1 \) and \( M \in \mathcal{M}[i,r] \), the distribution of \( M y \) is uniform over \((\{0, 1\}^k)^i\) and hence by the induction hypothesis

\[
\mathbb{E}_{y_1, \ldots, y_r} \left[ f_{=i}(M y) \right] = 0.
\]

114
proving the inductive step. We now know that the claim is true for \( j = 0 \) and all \( r \geq 1 \). In the following, we consider the case \( 1 \leq j \leq r - 1 \) and “reduce” it to the case \( j = 1 \). Using Lemma 4.3.8 we see

\[
\mathbb{E}_{y_{j+1} \ldots y_r} \left[ f_{=r}(a_1, \ldots, a_j, y_{j+1}, \ldots, y_r) \right] = \mathbb{E}_{y_{j+1} \ldots y_r} \left[ f_{=r-1}(a_2, \ldots, a_j, y_{j+1}, \ldots, y_r) - \frac{1}{\beta_{r-1, r-1}} \sum_{M \in \mathcal{M}[r-1, r]} \sum_{e_1 \notin \text{rowspan}(M)} f_{=r-1}(M(a, y)) \right].
\]

(4.8)

The expectation of the first term vanishes by induction hypothesis. For the second term, fix any matrix \( M \) therein. Since \( f_{=r-1} \) is basis-invariant, we can rewrite the rows of \( M \) as long as the row-span is preserved. By Lemma 4.3.15 below, we may assume that \( M \) is semi-diagonal. Hence

\[
M(a, y) = M(a_1, \ldots, a_j, y_{j+1}, \ldots, y_r) = (a'_1, \ldots, a'_j, y'_{j+1}, \ldots, y'_r),
\]

where \( a'_i = a_i \) or \( a_i + a_1 \) (hence fixed) and similarly, \( y'_i = y_i \) or \( y_i + a_1 \) (hence distributed same as \( y_i \)). By induction hypothesis

\[
\mathbb{E}_{y_{j+1} \ldots y_r} \left[ f_{=r-1}(M(a, y)) \right] = \mathbb{E}_{y_{j+1} \ldots y_r} \left[ f_{=r-1}(a'_1, \ldots, a'_j, y'_{j+1}, \ldots, y'_r) \right] = 0,
\]

completing the proof.

**Lemma 4.3.15.** An \((r - 1) \times r\) matrix is called semi-diagonal if deleting the first column gives a square matrix that is diagonal and has 1’s on the diagonal. Then for any matrix \( M \in \mathcal{M}[r - 1, r] \) such that \( e_1 \notin \text{rowspan}(M) \), there is a semi-diagonal matrix \( M' \in \mathcal{M}[r - 1, r] \) with the same row-span.

**Proof.** Let \( D \) be the row-span of \( M \) and for \( 2 \leq s \leq r \), let \( W_s \subseteq \{0, 1\}^r \) be the subspace of vectors with the last \( r - s \) coordinates zero. Since \( \dim(D) = r - 1 \), \( \dim(W_s) = s \), \( e_1 \in W_s \setminus D \), it must be the case that \( \dim(D \cap W_s) = s - 1 \). Thus \( \{D \cap W_s\}_{s=2}^r \) is an “increasing” sequence of subspaces that finally equals \( D \). Hence a basis for \( D \) can be chosen so that its successive members are in \( W_2 \setminus \{e_1\}, W_3 \setminus W_2, \ldots, W_r \setminus W_{r-1} \) respectively. Moreover, in this process, when we choose a vector \( v_s \in W_s \setminus W_{s-1} \), the \( s^{th} \) coordinate of \( v \) is 1, and we can zero-out its coordinates \( 2, \ldots, s - 1 \) by adding to it \( v_2, \ldots, v_{s-1} \) if necessary.

We now consider the Fourier representation of \( f_{=r} : \{0, 1\}^k \to \mathbb{R} \):

\[
f_{=r}(x_1, \ldots, x_r) = \sum_{T_1, \ldots, T_r} \hat{f}_{=r}(T_1, \ldots, T_r) \chi_{T_1}(x_1) \cdots \chi_{T_r}(x_r).
\]

**Lemma 4.3.16.** \( \hat{f}_{=r}(T_1, \ldots, T_r) \) depends only on \( \text{Span}(T_1, \ldots, T_r) \).

**Proof.** Immediately implied by the basis-invariance of \( f_{=r} \) (Lemma 4.3.6), and a proof identical to that of Lemma 4.3.1.

**Lemma 4.3.17.** Suppose \( T_1, \ldots, T_r \) are linearly dependent. Then \( \hat{f}_{=r}(T_1, \ldots, T_r) = 0 \).

**Proof.** Assume w.l.o.g. that \( T_r \) depends on \( T_1, \ldots, T_{r-1} \). By Lemma 4.3.16,

\[
\hat{f}_{=r}(T_1, \ldots, T_r) = \hat{f}_{=r}(T_1, \ldots, T_{r-1}, 0)
\]

\[
= \mathbb{E}_{y_1 \ldots y_r} \left[ f_{=r}(y_1, \ldots, y_r) \prod_{j=1}^{r-1} \chi_{T_j}(y_j) \right]
\]

\[
= \mathbb{E}_{y_1 \ldots y_{r-1}} \left[ \mathbb{E}_{y_r} \left[ f_{=r}(y_1, \ldots, y_{r-1}, y_r) \right] \prod_{j=1}^{r-1} \chi_{T_j}(y_j) \right],
\]

115
and the inner expectation vanishes according to Lemma 4.3.14.

Therefore, we may write

\[ f_r(y_1, \ldots, y_r) = \sum_{\substack{T_1, \ldots, T_r \in \mathcal{M}[r] \cap \mathbb{Z}^r \setminus \{0\}, \dim(T_1, \ldots, T_r) = r}} T_1(y_1) \cdots T_r(y_r). \quad (4.9) \]

**Lemma 4.3.18.** Let \(0 \leq r \leq \ell\) and let \(T_1, \ldots, T_r\) be characters such that \(\dim(T_1, \ldots, T_r) = r\). Then

\[ \hat{f}_r(T_1, \ldots, T_r) = \hat{F}(T_1, \ldots, T_r). \]

**Proof.** For any \(x_1, \ldots, x_\ell \in \{0, 1\}^k\), by Lemma 4.3.13,

\[
F_{=r}[x_1, \ldots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} f_r(Mx) \\
= \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r,\ell]} \sum_{\substack{T_1, \ldots, T_r \in \mathcal{M}[r,\ell], \dim(T_1, \ldots, T_r) = r}} \hat{f}_r(T_1, \ldots, T_r) \prod_{j=1}^r \chi_{T_j}((Mx)_j) \\
= \frac{1}{\beta_{r,r}} \sum_{\substack{T_1, \ldots, T_r \in \mathcal{M}[r,\ell], \dim(T_1, \ldots, T_r) = r}} \hat{f}_r(T_1, \ldots, T_r) \sum_{M \in \mathcal{M}[r,\ell]} \prod_{j=1}^r \chi_{T_j}((Mx)_j).
\]

As we have done previously, if \(T = (T_1, \ldots, T_r)\) and \(M^T_r\) is the transposed matrix, we have

\[
\prod_{j=1}^r \chi_{T_j}((Mx)_j) = \prod_{i=1}^\ell \chi_{(M^T)_i}(x_i).
\]

Hence,

\[
F_{=r}[x_1, \ldots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{\substack{T_1, \ldots, T_r \in \mathcal{M}[r,\ell], \dim(T_1, \ldots, T_r) = r}} \hat{f}_r(T_1, \ldots, T_r) \sum_{M \in \mathcal{M}[r,\ell]} \prod_{i=1}^\ell \chi_{(M^T)_i}(x_i) \\
= \sum_{\substack{P_1, \ldots, P_\ell \in \mathcal{M}[r,\ell] \setminus \{0\}, \dim(P_1, \ldots, P_\ell) = r}} \hat{f}_r((P_1, \ldots, P_\ell)) \prod_{i=1}^\ell \chi_{P_i}(x_i),
\]

where we used the (easy to verify) fact that each tuple \((P_1, \ldots, P_\ell)\) whose span is of dimension \(\leq r\), is counted exactly once. On the other hand, by definition

\[
F_{=r}[x_1, \ldots, x_\ell] = \sum_{\substack{P_1, \ldots, P_\ell \in \mathcal{M}[r,\ell] \setminus \{0\}, \dim(P_1, \ldots, P_\ell) = r}} \hat{F}((P_1, \ldots, P_\ell)) \prod_{i=1}^\ell \chi_{P_i}(x_i).
\]

By uniqueness of the Fourier representation, we conclude the assertion of the lemma.
4.3.5 Bounding Restricted Fourier Sums of $f_{=r}$

**Lemma 4.3.19.** Let $F : \{0, 1\}^\ell \to \{0, 1\}$ be basis invariant, $(r, \varepsilon)$ pseudorandom function, and let $0 \leq j \leq \frac{r}{2}$. Then for any characters $A_1, \ldots, A_j$,

$$
\sum_{T_j+1, \ldots, T_r} \hat{F}_{=r}^2(A_1, \ldots, A_j, T_{j+1}, \ldots T_r) \leq \frac{2^{4r^2}}{2^{(r-j)\ell}} \varepsilon.
$$

**Proof.** We will prove an upper bound of $\frac{C_r}{2^{(r+j)\ell}} \varepsilon$ with $C_r = 2^{4r^2}$. First, note that an $(r, \varepsilon)$-pseudorandom function automatically has average at most $\varepsilon$, and hence $\|F\|_2^2 \leq \varepsilon$. For $j = r = 0$, the upper bound clearly holds with $C_0 = 1$, so we assume $r \geq 1$. The proof is by induction on $j$. When $j = 0$, we have (note that non-zero Fourier coefficients of $f_{=r}$ have linearly independent arguments)

$$
\sum_{\dim(T_1, \ldots, T_r) = r} \hat{F}_{=r}^2(T_1, \ldots T_r) = \sum_{\dim(T_1, \ldots, T_r) = r} \hat{F}^2(T_1, \ldots T_r)
$$

where we used the fact that for a fixed $r$-dimensional span of the arguments, there are $\beta_{r,r}$ terms $(T_1, \ldots, T_r)$ in the first summation and $\beta_{r,\ell}$ terms $(Q_1, \ldots, Q_\ell)$ in the second summation. The expression is then upper bounded by ($C_r = 2^{4r^2}$),

$$
\frac{\beta_{r,r}}{\beta_{r,\ell}} \|F_{=r}\|_2^2 \leq \frac{2^{4r^2}}{2^{2r\ell}} \|F\|_2^2 \leq \frac{2^{4r^2+1}}{2^{2r\ell}} \varepsilon \leq \frac{C_r}{2^{r\ell}} \varepsilon.
$$

Now, let $j \geq 1$. By Lemma 4.3.18 and 4.3.4, for any $\dim(A_1, \ldots, A_j, T_{j+1}, \ldots, T_r) = r$,

$$
\hat{F}_{=r}^2(A_1, \ldots, A_j, T_{j+1}, \ldots T_r) = \hat{F}^2(A_1, \ldots, A_j, T_{j+1}, \ldots T_r)
$$

$$
= \left( \frac{1}{(2^\ell - 2^{r-1})^2} \right) \left( \hat{F}_{W_{A_1}}(A_2, \ldots, A_j, T_{j+1}, \ldots, T_r) - \sum_{D \subseteq \text{Span}(A_1, \ldots, A_j, T_{j+1}, \ldots, T_r), \dim(D) = r-1, A_i \notin D} \hat{F}(D) \right)^2
$$

$$
\leq \frac{4 \cdot 2^r}{2^{2\ell}} \left( \hat{F}_{W_{A_1}}^2(A_2, \ldots, A_j, T_{j+1}, \ldots, T_r) + \sum_{D \subseteq \text{Span}(A_1, \ldots, A_j, T_{j+1}, \ldots, T_r), \dim(D) = r-1, A_i \notin D} \hat{F}^2(D) \right), \quad (4.10)
$$

where the last inequality holds by Cauchy-Schwarz (there are $2^{r-1}$ choices for $D$). Summing over $T_{j+1}, \ldots, T_r$, the first term sums up to at most

$$
\sum_{\dim(A_2, \ldots, A_j, T_{j+1}, \ldots, T_r) = r-1} \hat{F}_{W_{A_1}}^2(A_2, \ldots, A_j, T_{j+1}, \ldots, T_r)
$$

$$
= \sum_{\dim(A_2, \ldots, A_j, T_{j+1}, \ldots, T_r) = r-1} \hat{F}_{=r-1, W_{A_1}}^2(A_2, \ldots, A_j, T_{j+1}, \ldots, T_r)
$$

$$
\leq \frac{C_{r-1}}{2^{(r-1+j-1)\ell}} \varepsilon,
$$

117
using the induction hypothesis and the \((r - 1, \varepsilon)\) pseudorandomness of \(F_{W_1}\) (that is holds by the \((r, \varepsilon)\) pseudorandomness of \(F\)). For the second term, consider any \(D \subseteq \text{Span}(A_1, \ldots, A_j, T_{j+1}, \ldots, T_r)\) of dimension \(r - 1\) that does not contain \(A_1\). By Lemma 4.3.15, we may assume that \(D\) has basis
\[
D = \text{Span}(A_2', \ldots, A_j', T_{j+1}', \ldots, T_r')
\]
where \(A_i' = A_i + b_i \cdot A_1\) and \(T_i' = T_i + b_i \cdot A_1\) for some \(b = (b_2, \ldots, b_r)\). In the following calculation, \(b\) is thought of as fixed. Summing over all \(T_{j+1}, \ldots, T_r\),
\[
\sum_{T_{j+1}, \ldots, T_r, \dim(A_1, \ldots, A_j, T_{j+1}, \ldots, T_r) = r} \hat{F}^2(D) \leq \sum_{T_{j+1}, \ldots, T_r, \dim(A_1', \ldots, A_j', T_{j+1}', \ldots, T_r') = r-1} \hat{F}^2(A_2', \ldots, A_j', T_{j+1}', \ldots, T_r')
\]
\[
= \sum_{T_{j+1}, \ldots, T_r, \dim(A_1', \ldots, A_j', T_{j+1}', \ldots, T_r') = r-1} \hat{f}_{r-1}^2(A_2', \ldots, A_j', T_{j+1}', \ldots, T_r')
\]
\[
\leq \frac{C_{r-1}}{2^{(r-1+j-1)\ell}} \varepsilon,
\]
using the induction hypothesis. We note that there are \(2^{r-1}\) choices for \(b\) (or equivalently \(D\)). Combining both upper bounds, we get an upper bound of
\[
\left(\frac{\varepsilon}{2^{(r+j)\ell}}\right) \cdot (4 \cdot 2^r) \cdot (1 + 2^{r-1}) \cdot C_{r-1}.
\]
This is upper bounded by \(\frac{C_r}{2^{(r+j)\ell}} \varepsilon\) provided \(C_r \geq 2^{2r+2}C_{r-1}\) (and \(C_0 = 1\)). Letting \(C_r = 2^{4r^2}\) proves the lemma.

\[\square\]

**Lemma 4.3.20.** Let \(F : \{(0, 1)\}^k \to \{0, 1\}\) be basis invariant, \((r, \varepsilon)\) pseudorandom function, and let \(s, t, p, q\) be non-negative integers such that \(s + t + p + q = r \leq \frac{k}{2}\). Let \(a_1, \ldots, a_s, x, y, z\) be characters. Define the restriction
\[
g_{a_1, \ldots, a_s, x, y, z, t}(y_1, \ldots, y_p, z_1, \ldots, z_q) = f_{=r}(a_1, \ldots, a_s, x_1, \ldots, x_t, y_1, \ldots, y_p, z_1, \ldots, z_q).
\]
Then
\[
\mathbb{E}_{x, y, z, t \in \{0, 1\}^k} \left[ \sum_{T_1, \ldots, T_q} \hat{g}_{a_1, \ldots, a_s, x, y, z, t}(A_1, \ldots, A_p, T_1, \ldots, T_q) \right] \leq \frac{2^{2sr^2+4r^2}}{2^{(t+p+q)\ell}} \varepsilon \leq \frac{2^{6r^2}}{2^{(t+p+q)\ell}} \varepsilon.
\]

**Proof.** We will prove an upper bound of \(\frac{C_{s,r}}{2^{(t+p+q+2\ell)}} \varepsilon\) where \(C_{s,r} = 2^{2sr^2+4r^2}\). The proof is by induction on \(s\). First let us consider the case \(s = 0\). The expectation on the left hand side of the lemma equals (we denote by \(x, y, z\) the respective tuples of variables)
\[
\mathbb{E}_x \left[ \sum_{T_1, \ldots, T_q} \left( \mathbb{E}_{y, z} \left[ f_{=r}(x, y, z) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] \right)^2 \right].
\]
(4.11)
Consider the inner expectation. Using the Fourier decomposition of \(f_{=r}\), it equals
\[
\sum_{Q_1, \ldots, Q_t} \mathbb{E}_{y, z} \left[ \hat{f}_{r}(Q, R, S) \prod_{i=1}^t \chi_{Q_i}(x_i) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i \in S_i}(z_i) \right]
\]
\[
= \sum_{Q_1, \ldots, Q_t} \hat{f}_{r}(Q, A, T) \prod_{i=1}^t \chi_{Q_i}(x_i).
\]
118
Squaring this, taking the expectation over $x$, and then summing over $T_1, \ldots, T_q$ shows that (4.11) equals,

$$\sum_{Q_1, \ldots, Q_t, T_1, \ldots, T_q} \tilde{f}_{=r}^2(Q_1, \ldots, Q_t, A_1, \ldots, A_p, T_1, \ldots, T_q),$$

which is upper bounded by $\frac{2^{4r^2}}{2^{(t+2p+q)r}} \varepsilon$ by Lemma 4.3.19. Now consider the case $s \geq 1$. As before, the expectation on the left hand side of the lemma equals (with an additional argument $a = (a_1, \ldots, a_s)$)

$$\mathbb{E}_x \left[ \sum_{T_1, \ldots, T_q} \left( \mathbb{E}_{y,z} \left[ f_{=r}(a, x, y, z) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] \right)^2 \right]. \quad (4.12)$$

Applying Lemma 4.3.8 we get

$$f_{=r}(a_1, \ldots, a_s, x, y, z) = f_{=r-1,F_{a_1}}(a_2, \ldots, a_s, x, y, z) + \frac{1}{\beta_{r-1,r-1}} \sum_{M \in M[r-1,r]} \sum_{e_1 \not\in \text{rowspan}(M)} f_{=r-1,F_{a_1}}(M(a, x, y, z)).$$

We take expectation over $y, z$. For the first term, we have

$$\mathbb{E}_{y,z} \left[ f_{=r-1,F_{a_1}}(a_2, \ldots, a_s, x, y, z) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] = \tilde{h}_{a_2,\ldots,a_s,x}(A_1, \ldots, A_p, T_1, \ldots, T_q), \quad (4.13)$$

where $h_{a_2,\ldots,a_s,x}$ is the restriction of the function $f_{=r-1,F_{a_1}}$ in a manner similar to $g$ is the restriction of $f_{=r}$. For the second term, let $M \in M[r-1,r]$ whose row-span does not contain $e_1$. By Lemma 4.3.15, we may assume that $M$ is semi-diagonal and then

$$M(a, x, y, z) = (a' = (a_2', \ldots, a_s'), x', y', z'),$$

where each new coordinate is the same as earlier, except possibly adding $a_1$. Hence

$$\mathbb{E}_{y,z} \left[ f_{=r-1}(M(a, x, y, z)) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] = \text{sign} \cdot \tilde{h}_{a',x'}(A_1, \ldots, A_p, T_1, \ldots, T_q) \quad (4.14)$$

for a sign $\in \{-1, 1\}$ (which takes into account the possible additions of $a_1$ to get the new coordinates $y', z'$). To upper-bound (4.12), we sum up the absolute values of (4.13) and (4.14) (the latter summed over $\beta_{r-1,r}$ matrices along with the leading coefficient $\frac{1}{\beta_{r-1,r-1}}$), square the sum, upper bound it by Cauchy-Schwartz, and finally take the outer summation over $T = (T_1, \ldots, T_q)$ and expectation over $x$. We end up with an overall upper bound

$$\left( 1 + \frac{\beta_{r-1,r}}{\beta_{r-1,r-1}^2} \right) \left( \mathbb{E}_x \left[ \sum_T \tilde{h}_{a_2,\ldots,a_s,x}(A, T) \right] + \sum_{M \in M[r-1,r]} \mathbb{E}_x \left[ \sum_T \tilde{h}_{a',x'}^2(A, T) \right] \right).$$

We may now apply the induction hypothesis since sequences $(a_2, \ldots, a_s)$ and $a'$ have length $s-1$, $x'$ is distributed the same as $x$, and furthermore, $F_{a_1}$ is $(r-1, \varepsilon)$ pseudorandom. Thus we get an upper bound of

$$\frac{\varepsilon}{2(t+2p+q)\varepsilon} \cdot C_{s-1,r-1} \cdot \left( 1 + \frac{\beta_{r-1,r}}{\beta_{r-1,r-1}^2} \right) (1 + \beta_{r-1,r}) \right).$$

Using very crude estimates $\beta_{r-1,r-1} \geq 1$ and $\beta_{r-1,r} \leq 2^{r^2} - 1$, we upper bound by $\frac{\varepsilon}{2(t+2p+q)\varepsilon} \cdot C_{s,r}$. It suffices to have $C_{s,r} \geq 2^{r^2} C_{s-1,r-1}$ and $C_{0,r} = 2^{4r^2}$, i.e. $C_{s,r} = 2^{2sr^2 + 4r^2}$. □

119
4.4 Pair-wise and Three-wise Correlations of $f_{=i}$

The rest of the chapter is devoted to the proof of our main technical result, Theorem 4.2.15, that upper bounds the four-wise correlations of $f_{=i}$. It is natural and instructive to first understand pair-wise and three-wise correlations of $f_{=i}$.

4.4.1 Pairwise Correlations

Studying pairwise correlations is simple. There are two cases depending on whether $\text{rowspan}(M_1), \text{rowspan}(M_2)$ are equal or not. In the former case, using the basis-invariance of $f_{=i}$, we may assume that the two matrices are the same.

**Lemma 4.4.1.** Let $M_1, M_2 \in \mathcal{M}[i, \ell]$. If $\text{rowspan}(M_1) \neq \text{rowspan}(M_2)$, then

$$\mathbb{E}_{x_1, \ldots, x_\ell \in \{0,1\}^k} [f_{=i}(M_1^T x) f_{=i}(M_2^T x)] = 0.$$  

**Proof.** It is clearly possible to choose linearly independent vectors $v_1, \ldots, v_s, u_1, \ldots, u_{i-s}, w_1, \ldots, w_{i-s}$ in $\{0,1\}^\ell$ such that

- $(v_1, \ldots, v_s)$ is a basis for $\text{rowspan}(M_1) \cap \text{rowspan}(M_2)$,
- $(v_1, \ldots, v_s, u_1, \ldots, u_{i-s})$ is a basis for $\text{rowspan}(M_1)$, and
- $(v_1, \ldots, v_s, w_1, \ldots, w_{i-s})$ is a basis for $\text{rowspan}(M_2)$.

By the assumption $i - s \geq 1$. By the basis-invariance of $f_{=i}$, we can assume that in the last two items, the respective sets are in fact the rows of the two matrices. For a row-vector $a \in \{0,1\}^\ell$ and $x = (x_1, \ldots, x_\ell)$, let us denote $a' = \langle a, x \rangle = \sum_{j=1}^\ell a_j x_j$. Denote

$$v'_j = \langle v_j, x \rangle, \quad u'_j = \langle u_j, x \rangle, \quad w'_j = \langle w_j, x \rangle.$$  

Thus $\{v'_j, u'_j, w'_j\}$ are uniformly and independently distributed over $\{0,1\}^k$. Moreover

$$M_1 x = (v'_1, \ldots, v'_s, u'_1, \ldots, u'_{i-s}), \quad M_2 x = (v'_1, \ldots, v'_s, w'_1, \ldots, w'_{i-s}).$$  

It follows using Lemma 4.3.14 that

$$\mathbb{E}_{x} [f_{=i}(M_1^T x) f_{=i}(M_2^T x)] = \mathbb{E}_{v'_j, u'_j, w'_j} \left[ f_{=i}(v'_1, \ldots, v'_s, u'_1, \ldots, u'_{i-s}) f_{=i}(v'_1, \ldots, v'_s, w'_1, \ldots, w'_{i-s}) \right]$$

$$= \mathbb{E}_{v'_j, u'_j} \left[ \mathbb{E}_{u'_1, \ldots, u'_{i-s}} \left[ f_{=i}(v'_1, \ldots, v'_s, u'_1, \ldots, u'_{i-s}) f_{=i}(v'_1, \ldots, v'_s, w'_1, \ldots, w'_{i-s}) \right] \right]$$

$$= 0.$$  

□

**Lemma 4.4.2.** Let $M \in \mathcal{M}[i, \ell]$. Then

$$\mathbb{E}_{x_1, \ldots, x_\ell \in \{0,1\}^k} [f_{=i}^2(M^T x)] = \frac{\beta_{i,i}}{\beta_{i,\ell}} \| F_{=i} \|^2.$$  

120
Proof. Denote the expectation on the left hand side of the lemma by $\Gamma$.

Denoting $(y_1, \ldots, y_t) = Mx$, we note that $y_1, \ldots, y_t$ are uniformly and independently distributed in $\{0, 1\}^k$ and hence $\Gamma$ does not depend on the choice of $M \in M[i, \ell]$. Also, due to basis invariance, the expectation is the same as $\Gamma = \mathbb{E}_x [f_{=i}(M_1 x) f_{=i}(M_2 x)]$ as long as $\text{rowspan}(M_1) = \text{rowspan}(M_2)$. Using Lemma 4.3.13, squaring, and taking expectation over $x$,

$$F_{=i}[x_1, \ldots, x_\ell] = \frac{1}{\beta_{i,1}} \sum_{M \in M[i, \ell]} f_{=i}(M x).$$

$$\beta_{i,i}^2 \cdot \|F_{=i}\|_2^2 = \sum_{M_1, M_2 \in M[i, \ell]} \sum_{\text{rowspan}(M_1) = \text{rowspan}(M_2)} \Gamma + \sum_{M_1, M_2 \in M[i, \ell]} \sum_{\text{rowspan}(M_1) \neq \text{rowspan}(M_2)} \mathbb{E}_x [f_{=i}(M_1 x) f_{=i}(M_2 x)].$$

The lemma follows by noting that there are $\beta_{i,\ell} \beta_{i,i}$ pairs $M_1, M_2$ with the same row-span and by previous Lemma 4.4.1, the expectation vanishes when the row-spans are distinct. \hfill \Box

The left hand side in the statement of this Lemma equals $\| f_{=i} \|_2^2$ and we have $\beta_{i,i} \leq 2^i - 1$, $\beta_{i,\ell} \geq \frac{1}{2} \cdot 2^i$. We record this very useful fact for future:

**Lemma 4.4.3.** $\| f_{=i} \|_2^2 \leq \frac{2^i}{2^r} \| F_{=i} \|_2^2$.

### 4.4.2 Three-wise Correlations

Understanding three-wise correlations is more difficult. Here, we will need to use the Fourier analytic machinery developed in Section 4.3. Our formal result is:

**Theorem 4.4.4.** Let $F: (\{0, 1\}^k)^\ell \to \{0, 1\}$ be basis invariant, $(r, \varepsilon)$ pseudorandom function, and $\eta = \| F_{=i} \|_2^2$. Then for any $0 \leq i \leq r$, $i \leq d \leq 3i$, $A \subseteq \{0, 1\}^d$ of dimension $d$ and $M_1, M_2, M_3 \in M[i, \ell]$ such that $A \subseteq \{0, 1\}^d$ of dimension $d$, and $M_1, M_2, M_3 \in M[i, \ell]$ whose direct sum of row spaces is $A$. Since $f_{=i}$ is basis invariant, we are free to rewrite the rows of each matrix as long as the row-span is preserved. We will spend some effort into bringing the matrices into a convenient form. We begin with the following simple observation.

**Lemma 4.4.5.** Suppose $\text{rowspan}(M_3) \not\subseteq \text{rowspan}(M_1) \oplus \text{rowspan}(M_2)$ (or the other two symmetric cases). Then

$$\mathbb{E}_{x \in (\{0, 1\}^k)^\ell} [f_{=i}(M_1 x) f_{=i}(M_2 x) f_{=i}(M_3 x)] = 0.$$
We denote \( v_j' = \langle v_j, x \rangle \) and \( w_j' = \langle w_j, x \rangle \) so that \( \{v_j', w_j'\} \) are uniformly and independently distributed in \( \{0, 1\}^k \). Clearly, \( M_1 x, M_2 x \) depend only on \( v_j' \) and \( M_3 x = (w_1', \ldots, w_t', y_1, \ldots, y_{i-t}) \) where \( y_1, \ldots, y_{i-t} \) depend only on \( v_j' \). Fixing an arbitrary choice of \( v_j' \) fixes \( a^* = M_1 x, b^* = M_2 x \) and \( c^* = (y_1, \ldots, y_{i-t}) \). The expectation is then

\[
E_{g, w, y, z} \left[ f_{=i}(a^*) f_{=i}(b^*) \mathbb{E}_{w_1', \ldots, w_t'} \left[ f_{=i}(w_1', \ldots, w_t', c^*) \right] \right],
\]

which vanishes according to Lemma 4.3.14.

Thanks to Lemma 4.4.5, we assume henceforth that the row-span of each matrix is contained in the direct sum of the other two. Let \( H = \cap_{j=1}^3 \text{rowspan}(M_j) \), \( \dim(H) = s \), and let \( g_1, \ldots, g_s \) be a basis for it. We may assume w.l.o.g. that the first \( s \) rows of each matrix are \( g_1, \ldots, g_s \). Let \( M_1', M_2', M_3' \) be the matrices \( M_1, M_2, M_3 \) after removing these first \( s \) rows. By our assumptions, we have \( \cap_{j=1}^3 \text{rowspan}(M_j') = \{0\} \) and moreover the row-span of each is contained in the direct sum of the other two. We show that one can assume a strong structure on the row-spans of \( M_1', M_2', M_3' \) as below. We recommend reading the proof as similar tricks are used hereafter.

**Lemma 4.4.6.** The row spans of \( M_1', M_2', M_3' \) have the form

\[
\text{Span}(w_1, \ldots, w_t, y_1, \ldots, y_n, y_{n+1}, \ldots, y_{i-s-t})
\]

\[
\text{Span}(w_1, \ldots, w_t, z_1, \ldots, z_n, z_{n+1}, \ldots, z_{i-s-t})
\]

\[
\text{Span}(y_1 + z_1, \ldots, y_n + z_n, y_{n+1}, \ldots, y_{i-s-t}, z_{n+1}, \ldots, z_{i-s-t})
\]

where the vectors \( w_1, \ldots, w_t, y_1, \ldots, y_{i-s-t}, z_1, \ldots, z_{i-s-t} \) are linearly independent.

**Proof.** Let \( \{w_1, \ldots, w_t\} \) be the basis for \( \text{rowspan}(M_1') \cap \text{rowspan}(M_2') \). Let

\[ A = \text{rowspan}(M_1'), B = \text{rowspan}(M_2'), C = \text{rowspan}(M_3') \]

so that \( A \cap B \cap C = \{0\} \) and each is contained in the direct sum of the remaining two. If we pretend that \( w_1 = \ldots = w_t = 0 \) (which really amounts to working with a quotient space, but we find this informal description clearer), we can pretend that \( A \cap B = \{0\} \). Apply Lemma 4.8.1 to get the desired form. One subtlety to note however is that each variable \( y_{j} \) above (the same goes for \( z_{j} \)) is really \( y_{j} + \sigma(w) \) where \( \sigma(w) \) denotes some arbitrary linear combination of \( w_1, \ldots, w_t \) (not necessarily the same for different \( y_j, z_j \)), a side effect of “pulling back” from the quotient space. Nevertheless, this can be fixed by simply redefining \( y_j \leftarrow y_j + \sigma(w) \).

We now turn back to the task of upper bounding the expectation in (4.15). Apply a change of variables as follows: \( g_j' = \langle y_j, x \rangle \) for \( j = 1, \ldots, s \), \( w_j' = \langle w_j, x \rangle \) for \( j = 1, \ldots, t \), and \( y_j' = \langle y_j, x \rangle \) and \( z_j' = \langle z_j, x \rangle \) for \( j = 1, \ldots, i-s-t \). Since these vectors are linearly independent, we have that our \( g', w', y', z' \) variables are independent and uniform over \( \{0, 1\}^k \). For notational simplicity, we will just drop the primes in the superscripts, and relabel these variables as \( g, w, y, z \). Thus the expectation in (4.15) equals

\[
E_{g, w, y, z} \left[ f_{=i}(g_1, \ldots, g_s, w_1, \ldots, w_t, y_1, \ldots, y_n, y_{n+1}, \ldots, y_{i-s-t})
\right]
\]

\[
E_{g, w, y, z} \left[ f_{=i}(g_1, \ldots, g_s, w_1, \ldots, w_t, z_1, \ldots, z_n, z_{n+1}, \ldots, z_{i-s-t})
\right]
\]

\[
E_{g, w, y, z} \left[ f_{=i}(g_1, \ldots, g_s, y_1 + z_1, \ldots, y_n + z_n, y_{n+1}, \ldots, y_{i-s-t}, z_{n+1}, \ldots, z_{i-s-t})
\right].
\]
Denote \( h_{g_1,...,g_s}(a_1,\ldots,a_{i-s}) = f_{=i}(g_1,\ldots,g_s,a_1,\ldots,a_{i-s}) \). To reduce cumbersome notation, we drop the subscript from \( h \) for now and remember that it is \( g_1,\ldots,g_s \) throughout. Then our expectation is

\[
\mathbb{E}_{g,w,y,z} \left[ h(w_1,\ldots,w_t, y_1,\ldots,y_n, y_{n+1},\ldots,y_{i-s-t}) \right. \\
\left. h(w_1,\ldots,w_t, z_1,\ldots,z_n, z_{n+1},\ldots,z_{i-s-t}) \\
\right. \\
\left. h(y_1 + z_1,\ldots,y_n + z_n, y_{n+1},\ldots,y_{i-s-t}, z_{n+1},\ldots,z_{i-s-t}) \right].
\]

For a tuple \((b_1, b_2, \ldots, b_n)\) and \( m_1 \leq m_2 \), we will denote by \( b_{[m_1:m_2]} \) the sub-tuple \((b_{m_1}, b_{m_1+1}, \ldots, b_{m_2})\). Applying Fourier transform on \( h \) and using the expectation over \( w, y, z \), we see that the expectation equals

\[
\mathbb{E}_g \sum_{T_1,\ldots,T_n} \sum_{W_1,\ldots,W_t} \sum_{P_{n+1},\ldots,P_{n+i-s-t}} \sum_{Q_{n+1},\ldots,Q_{i-s-t}} \hat{h}(T_{[n:1]}, P_{[n+1:i-s-t]}, Q_{[n+1:i-s-t]}) \cdot \hat{h}(W_{[1:t]}, T_{[1:n]}, P_{[n+1:i-s-t]}) \cdot \hat{h}(W_{[1:t]}, T_{[1:n]}, Q_{[n+1:i-s-t]})
\]

For ease of notation, we will denote by \( T \) the tuple \((T_1, \ldots, T_n)\) and similarly for \( W, P, Q \). Thus the expression above can be written as

\[
\mathbb{E}_g \left[ \sum_{T,W,P,Q} \hat{h}(W, T, P) \cdot \hat{h}(W, T, Q) \cdot \hat{h}(T, P, Q) \right].
\]

For a fixed \( g \) and \( T \), the sum over \( W, P, Q \) can be upper bounded in absolute value by repeated Cauchy-Schwartz as (this technique will be very useful; it is summarized as Lemma 4.8.4):

\[
\sum_{W,P} \left| \hat{h}(W, T, P) \cdot \hat{h}(W, T, Q) \cdot \hat{h}(T, P, Q) \right| \\
\leq \sum_{W,P} \left| \hat{h}(W, T, P) \right| \left( \sum_{Q} \hat{h}^2(W, T, Q) \right) \left( \sum_{Q} \hat{h}^2(T, P, Q) \right) \\
= \sum_{W} \left( \sum_{Q} \hat{h}^2(W, T, Q) \right) \left( \sum_{P} \hat{h}^2(W, T, P) \right) \left( \sum_{Q} \hat{h}^2(T, P, Q) \right) \\
\leq \sum_{W} \sqrt{\sum_{Q} \hat{h}^2(W, T, Q)} \sqrt{\sum_{P} \hat{h}^2(W, T, P)} \sqrt{\sum_{Q} \hat{h}^2(T, P, Q)} \\
= \sqrt{\sum_{P,Q} \hat{h}^2(T, P, Q)} \sqrt{\sum_{W} \hat{h}^2(W, T, Q)} \sqrt{\sum_{W,P} \hat{h}^2(W, T, P)} \\
= \sqrt{A_1(T)} \sqrt{A_2(T)} \sqrt{A_2(T)},
\]

where we denoted the three expressions inside the square roots as \( A_1(T), A_2(T), A_2(T) \) respectively (noting that the second and the third are really the same). Considering the expectation over \( g \), and further upper
bounding
\[ E \left[ \sum_T \sqrt{A_1(T) A_2(T)} \right] \leq \sqrt{\max_{g,T} A_1(T)} \cdot E \left[ \sum_T A_2(T) \right]. \]

By Parseval and by Lemma 4.4.3,
\[ E \left[ \sum_T A_2(T) \right] = E \left[ \sum_{T, W, P} \hat{h}^2(T, W, P) \right] = E \left[ \|h\|_2^2 \right] = \|f_{=i}\|_2^2 \leq \frac{2^r}{2\ell} \eta \leq \frac{2^r}{2\ell} \eta. \quad (4.16) \]

By Lemma 4.3.20, we have
\[ \max_{g,T} A_1(T) = \max_{g,T} \sum_{P, Q} \hat{h}^2(T, P, Q) \leq \frac{2^6 \eta^3}{2^{2n+(2-(s-t-n))\ell}} \varepsilon \leq \frac{2^6 \eta^3}{2^{2(s-t)\ell}} \varepsilon. \quad (4.17) \]

Combining both upper bounds (4.16) and (4.17), and noting that \( d = 2i - s - t \), we conclude the required upper bound
\[ \left( \frac{2^6 \eta^3}{2^{2(s-t)\ell}} \varepsilon \right)^\frac{1}{2} \frac{2^r}{2\ell} \eta \leq 2^{4r^2} \frac{\eta \sqrt{\varepsilon}}{2(2i-s-t)\ell}. \]

### 4.5 Four-wise Correlations: Transforming the Matrices into Form

We now begin the proof of our main technical Theorem 4.2.15. In this section we bring the matrices \( M_1, M_2, M_3, M_4 \) into a convenient form, and the actual analysis is presented in subsequent sections. We emphasize again that we can rewrite rows of the four matrices as long as each row-span is preserved. Thanks to the lemma below, we assume henceforth that the row-span of each matrix is contained in the direct sum of the remaining three.

**Lemma 4.5.1.** Suppose \( \text{rowspan}(M_4) \not\subseteq \bigoplus_{j=1}^3 \text{rowspan}(M_j) \) (or the other three symmetric cases). Then
\[ E_{x \in \{0,1\}^k} \left[ f_{=i}(M_1x)f_{=i}(M_2x)f_{=i}(M_3x)f_{=i}(M_4x) \right] = 0. \]

**Proof.** Essentially the same as that of Lemma 4.4.5. \( \square \)

#### 4.5.1 Removing 4-wise and 3-wise Intersections of Rowspaces

Consider the subspace \( \cap_{j=1}^4 \text{rowspan}(M_j) \). Let \( H_4 \) be a basis for it and \( h_4 = |H_4| \) be its dimension. We may assume w.l.o.g. that the first \( h_4 \) rows of each matrix are precisely \( H_4 \) and the rest of their rows are linear combinations of vectors \( v_1, \ldots, v_r \) that are linearly independent of \( H_4 \). The rows \( H_4 \) are removed now from each matrix; they will only come into play at the very end of the analysis. For notational convenience, we refer to the matrices with these rows removed also as \( M_1, M_2, M_3, M_4 \) respectively. We assume henceforth that \( \cap_{j=1}^4 \text{rowspan}(M_j) = \{0\} \) and that the row-span of each matrix is contained in the direct sum of the remaining three.

We handle 3-wise intersections of the row-spaces in the same manner. Suppose there is a non-zero vector \( w \in \cap_{j=1}^3 \text{rowspan}(M_j) \). Since we assumed that the 4-wise intersection of the row-spaces is trivial, \( w \not\in \text{rowspan}(M_4) \). We may assume w.l.o.g. that \( w \) is the first row of \( M_1, M_2, M_3 \) and the rest of their rows, as well as the rows of \( M_4 \), are linear combinations of vectors \( v_1, \ldots, v_r \) (not necessarily the same as in the previous paragraph) that are linearly independent of \( w \). The row \( w \) is removed now from \( M_1, M_2, M_3 \); it
will only come into play at the very end of the analysis. For ease of notation, we refer to the matrices with this row removed also as $M_1, M_2, M_3$ respectively (and $M_4$ is unaffected). This process is repeated as long as there is a non-trivial 3-wise intersection of the row-spaces. At the end of this process, let $H_3$ denote the set of all row-vectors thus removed, $h_3 = |H_3|$, and $s_1, s_2, s_3, s_4$ be the number of remaining rows of the respective matrices. Since the original number of rows was $i$ and $h_4$ were removed in the earlier step, the number of rows removed from the $j^{th}$ matrix in the current step is $i - h_4 - s_j$.

We assume henceforth that the matrices $M_1, M_2, M_3, M_4$ do not have non-trivial 3-wise intersection of their row-spaces, that the row-space of each is contained in the direct sum of the remaining three, and that their number of rows is $s_1, s_2, s_3, s_4$ respectively.

### 4.5.2 Getting $M_1, M_2, M_3$ into Form

We first write $M_1, M_2, M_3$ in a convenient form. Letting $A = \text{rowspan}(M_1)$, $B = \text{rowspan}(M_2)$, and

$$C = \text{rowspan}(M_3) \cap (\text{rowspan}(M_1) \oplus \text{rowspan}(M_2)),$$

and applying an argument similar to Lemma 4.4.6 and Lemma 4.8.1, we can write $A, B, C$ as

$$\text{Span}(v_1, \ldots, v_t, p_1, \ldots, p_n, u, y)$$
$$\text{Span}(v_1, \ldots, v_t, q_1, \ldots, q_n, w, z)$$
$$\text{Span}(p_1 + q_1, \ldots, p_n + q_n, u, w)$$

where $u, y, w, z$ denote tuples of vectors (we do not wish to use an index/subscript to denote their length) and the vectors $(v_j, p_j, q_j, u_j, y_j, w_j, z_j)$ are linearly independent. Now we complete the basis for $C$ to that of $\text{rowspan}(M_3)$ by adding linearly independent vectors $a = (a_1, \ldots, a_h)$ from $\text{rowspan}(M_3) \setminus C$. Hence the row-spans of $M_1, M_2, M_3$ can be assumed to be of the form $(p, q, a)$ have the same length $n$:

$$\text{Span}(v, p, u, y)$$
$$\text{Span}(v, q, w, z)$$
$$\text{Span}(a, p_1 + q_1, \ldots, p_n + q_n, u, w).$$

(4.18)

### 4.5.3 Getting $M_4$ into Form: Part I

Now we begin the rather tedious process of getting $M_4$ into a convenient form given the form (4.18) for the first three matrices.

We pretend first that $v = p = u = q = w = 0$ (formally, taking a quotient). The first three row-spaces now amount to $Y = \text{Span}(y), Z = \text{Span}(z)$, and $A = \text{Span}(a)$. Denoting $W = \text{rowspan}(M_4)$ (its quotient to be precise), we have that $W \subseteq A \oplus Y \oplus Z$. Using Lemma 4.8.2, there is a basis for $W$ of the following form $\bigcup_{s=1}^{7} A_s$ where

$$A_1 = \{ a_i + y_j + z_k \mid i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \}$$
$$A_2 = \{ a_i + y_j \mid i \in \Sigma_2, j \in \Phi_2 \}$$
$$A_3 = \{ a_i + z_k \mid i \in \Sigma_3, k \in \Psi_3 \}$$
$$A_4 = \{ a_i + \sigma(y_{\Phi_1}, \Phi_5) \mid i \in \Sigma_4 \}$$
$$A_5 = \{ y_j + z_k \mid j \in \Phi_5, k \in \Psi_5 \}$$
$$A_6 = \{ y_j \mid j \in \Phi_6 \}$$
$$A_7 = \{ z_k \mid k \in \Psi_7 \}.$$
Here $\sigma(y_{\Phi_1, \Phi_5})$ are arbitrary linear forms in $\{y_j | j \in \Phi_1 \cup \Phi_5\}$ (distinct appearances of $\sigma(\cdot)$ may stand for different linear combinations; this issue will be essentially irrelevant for us). We emphasize that the notation (and similar ones) $\{a_i + y_j + z_k | i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1\}$ is imprecise, but chosen for ease of notation. Here $|\Sigma_1| = |\Phi_1| = |\Psi_1|$ and there are exactly $|\Sigma_1|$ vectors in this set, forming a kind of a perfect matching. We further emphasize the following observation.

Informally speaking, if all forms $\sigma(y_{\Phi_1, \Phi_5})$ are ignored, then each $a, y, z$ variable appears in the above representation exactly once. Formally,

- $\Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4$ (disjointly) cover all $a$-variables.
- $\Phi_1 \cup \Phi_2 \cup \Phi_5 \cup \Phi_6$ (disjointly) cover all $y$-variables.
- $\Psi_1 \cup \Psi_3 \cup \Psi_5 \cup \Psi_7$ (disjointly) cover all $z$-variables.

These three statements are simply consequences of $A \subseteq Y \oplus Z \oplus W, Y \subseteq A \oplus Z \oplus W, Z \subseteq A \oplus Y \oplus W$ respectively. Now we “pull back” from the quotient space. This has the effect of adding a term to each vector that denotes an arbitrary linear form in those variables (and since these forms would be essentially irrelevant, we use the same notation for all). This yields a partial basis for $\text{rowspan}(M_4)$ summarized below.

**Lemma 4.5.2.** $\text{rowspan}(M_4)$ has a partial basis of the following form $\bigcup_{s=1}^7 A_s$ where

$$
A_1 = \{ a_i + y_j + z_k + \sigma(v, p, u, q, w) \mid i \in \Sigma_1, \ j \in \Phi_1, \ k \in \Psi_1 \} \\
A_2 = \{ a_i + y_j + \sigma(v, p, u, q, w) \mid i \in \Sigma_2, \ j \in \Phi_2 \} \\
A_3 = \{ a_i + z_k + \sigma(v, p, u, q, w) \mid i \in \Sigma_3, \ k \in \Psi_3 \} \\
A_4 = \{ a_i + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5} v) \mid i \in \Sigma_4 \} \\
A_5 = \{ y_j + z_k + \sigma(v, p, u, q, w) \mid j \in \Phi_5, \ k \in \Psi_5 \} \\
A_6 = \{ y_j + \sigma(v, p, u, q, w) \mid j \in \Phi_6 \} \\
A_7 = \{ z_k + \sigma(v, p, u, q, w) \mid k \in \Psi_7 \}.
$$

Moreover, if all forms $\sigma(y_{\Phi_1, \Phi_5})$ are ignored, then each $a, y, z$ variable appears in the above representation exactly once.

### 4.5.4 Getting $M_4$ into Form: Part II

In the previous subsection, we obtained a partial basis for $\text{rowspan}(M_4)$ by pretending that $v = p = u = q = w = 0$ (but did add $\sigma(v, p, u, q, w)$ terms back to account for this). This basis can now be extended to a basis for $\text{rowspan}(M_4)$ by adding in a basis for

$$W = \text{rowspan}(M_4) \cap \text{Span}(v, p, u, q, w).$$

We do this in two steps. First, we pretend that $v = u = w = 0$. Let $P = \text{Span}(p)$ and $Q = \text{Span}(q)$ (note that $n = \dim(P) = \dim(Q)$). Since $W \subseteq P \oplus Q$, by Lemma 4.8.3, for a partition of the index set $\{1, \ldots, n\} = \Delta_0 \cup \Delta_1 \cup \Delta_2 \cup \ldots \cup \Delta_m \cup \Omega_1 \cup \Omega_2$, we may assume that $W$ has a basis

$$A_8 = \{ \ p_i + \sigma(q) \mid i \in \Delta_1 \}$$

$$A_9 = B_2 \cup \ldots \cup B_m$$

$$A_{10} = \{ \ p_i + \sigma(q_{\Omega_1}) \mid i \in \Omega_1 \cup \Omega_2 \}$$

$$A_{11} = \{ \ q_j \mid j \in \Omega_2 \}.$$  

Here $B_s = \{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) \mid j \in \Delta_s \}$. We recall that $\Delta_{[s+1:m]} = \Delta_{s+1} \cup \ldots \cup \Delta_m \cup \Omega_1 \cup \Omega_2$. As usual $\sigma(\cdot)$ are linear forms in its inputs that we do not really care about. We “pull back” from the quotient space by adding $\sigma(v, u, w)$ to every vector yielding:
Lemma 4.5.3. A partial basis for \( \text{rowspan}(M_4) \) from Lemma 4.5.2 can be further extended as \( A_8 \cup A_9 \cup A_{10} \cup A_{11} \) where

\[
A_8 = \{ p_i + \sigma(q) + \sigma(v, u, w) \mid i \in \Delta_1 \} \\
A_9 = B_2 \cup \ldots \cup B_m \\
A_{10} = \{ p_i + \sigma(q_{\Omega_1}) + \sigma(v, u, w) \mid i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} = \{ q_j + \sigma(v, u, w) \mid j \in \Omega_2 \}.
\]

Here \( B_s = \{ q_j + \sigma(p_{\Delta_{[s+1,m]}}) + \sigma(v, u, w) \mid j \in \Delta_s \} \). Finally, we can complete a basis for \( \text{rowspan}(M_4) \) by adding a basis for \( W = \text{rowspan}(M_4) \cap \text{Span}(v, u, w) \). This can clearly be done by first pretending \( v = w = 0 \), writing the basis for \( W \subseteq \text{Span}(u) \), pulling it back by adding forms \( \sigma(v, w) \), and then finally completing the basis by adding a basis for \( \text{rowspan}(M_4) \cap \text{Span}(v, w) \). We summarize this as:

Lemma 4.5.4. A partial basis for \( \text{rowspan}(M_4) \) from Lemmas 4.5.2 and 4.5.3 can be completed by adding \( A_{12} \cup A_{13} \) where (for some index set \( \Gamma \))

\[
A_{12} = \{ u_i + \sigma(v, w) \mid i \in \Gamma \} \\
A_{13} = \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)).
\]

To summarize, a basis for \( \text{rowspan}(M_4) \) can be assumed to be \( \cup_{s=1}^{13} A_s \) where

\[
A_1 = \{ a_i + y_j + z_k + \sigma(v, p, u, q, w) \mid i \in \Sigma_1, \ j \in \Phi_1, \ k \in \Psi_1 \} \\
A_2 = \{ a_i + y_j + \sigma(v, p, u, q, w) \mid i \in \Sigma_2, \ j \in \Phi_2 \} \\
A_3 = \{ a_i + z_k + \sigma(v, p, u, q, w) \mid i \in \Sigma_3, \ k \in \Psi_3 \} \\
A_4 = \{ a_i + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_2}) \mid i \in \Sigma_4 \} \\
A_5 = \{ y_j + z_k + \sigma(v, p, u, q, w) \mid j \in \Phi_5, \ k \in \Psi_5 \} \\
A_6 = \{ y_j + \sigma(v, p, u, q, w) \mid j \in \Phi_6 \} \\
A_7 = \{ z_k + \sigma(v, p, u, q, w) \mid k \in \Psi_7 \} \\
A_8 = \{ p_i + \sigma(q) + \sigma(v, u, w) \mid i \in \Delta_1 \} \\
A_9 = B_2 \cup \ldots \cup B_m \\
A_{10} = \{ p_i + \sigma(q_{\Omega_1}) + \sigma(v, u, w) \mid i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} = \{ q_j + \sigma(v, u, w) \mid j \in \Omega_2 \}.
\]

Here \( B_s = \{ q_j + \sigma(p_{\Delta_{[s+1,m]}}) + \sigma(v, u, w) \mid j \in \Delta_s \} \). The variables \( \{ a_i, y_j, z_k \} \) appearing in \( A_1, \ldots, A_7 \), the variables \( \{ p_i, q_j \} \) appearing in \( A_8, \ldots, A_{11} \) and the variables \( \{ u_i \} \) appearing in \( A_{12} \) will be called pivots (the reader should ignore the \( \sigma(\cdot) \) forms to clearly understand which variables we are referring to as pivots).

4.5.5 Getting \( M_4 \) into Form: Part III

In this section, we make further changes to the basis for \( \text{rowspan}(M_4) \) that are needed towards our final proof. We recommend however that the reader skips this section and jumps to the next section where we present a proof in a special but instructive case.
Step 1

We start with the basis in (4.19). We observe that:

- \((v, p, u)\) variables can be “absorbed into” the pivot \(y\)-variables,
- \((v, q, w)\) variables can be absorbed into the pivot \(z\)-variables,
- \(u\)-variables can be absorbed into the pivot \(p\)-variables, and
- \(w\)-variables can be absorbed into the pivot \(q\)-variables.

Therefore, if we have a \(y\)-variable as a pivot, there is no need to include \((v, p, u)\)-variables in the corresponding \(\sigma(\cdot)\) form (and similarly for the \(z, p, q\) pivots). This leads to the simplified \(\sigma(\cdot)\) forms as shown below.

\[
\begin{align*}
A_1 &= \{ a_i + y_j + z_k \} & i \in \Sigma_1, \ j \in \Phi_1, \ k \in \Psi_1 \\
A_2 &= \{ a_i + y_j + \sigma(q, w) \} & i \in \Sigma_2, \ j \in \Phi_2 \\
A_3 &= \{ a_i + z_k + \sigma(p, u) \} & i \in \Sigma_3, \ k \in \Psi_3 \\
A_4 &= \{ a_i + \sigma(v, p, u, q, w) + \sigma(yq_1, q_2) \} & i \in \Sigma_4 \\
A_5 &= \{ y_j + z_k \} \quad j \in \Phi_5, \ k \in \Psi_5 \\
A_6 &= \{ y_j + \sigma(q, w) \} \quad j \in \Phi_6 \\
A_7 &= \{ z_k + \sigma(p, u) \} \quad k \in \Psi_7 \\
A_8 &= \{ p_i + \sigma(q) + \sigma(v, w) \} \quad i \in \Delta_1 \\
A_9 &= B_2 \cup \ldots \cup B_m \\
A_{10} &= \{ p_i + \sigma(q\Omega_1) + \sigma(v, w) \} \quad i \in \Omega_1 \cup \Omega_2 \\
A_{11} &= \{ q_j + \sigma(v, u) \} \quad j \in \Omega_2 \\
A_{12} &= \{ u_i + \sigma(v, w) \} \quad i \in \Gamma \\
A_{13} &= \text{basis(rowspan}(M_4) \cap \text{Span}(v, w))
\end{align*}
\]

Here \(B_8 = \{ q_j + \sigma(p_{\Delta_{[i+1:m]}}) + \sigma(v, u) \mid j \in \Delta_8 \}.\)

Step 2

Consider \(A_7\) and its vectors \(\{ z_k + \sigma(p, u) \mid k \in \Psi_7 \}.\) By adding vectors from \(A_8\) if necessary, we can assume that the form \(\sigma(\cdot)\) does not depend on \(p_{\Delta_1}\) (we may need re-absorption of \(v, q, w\) into \(z_k\)). We wish to make the dependence on \(p_{\Delta_0}\) more restrictive. So our concern is with their \(z_k + \sigma(p_{\Delta_0})\) component. We can change the matched basis for \(p_{\Delta_0}\) so that for a partition \(\Delta_0 = \Delta'_0 \cup \Delta''_0, \Psi_7 = \Psi_7a \cup \Psi_7b,\) these components turn into

\[
\{ z_k + p_s \mid k \in \Psi_7a, s \in \Delta'_0 \} \cup \{ z_k + \sigma(p_{\Delta''_0}) \mid k \in \Psi_7b \}
\]

Further, adding the former to the latter as necessary (which amounts to a change of basis for \(z_{\Psi_7}\)), the latter components can be made independent of \(p_{\Delta_0}\) altogether. Additionally, for the former we may absorb \(u\) and \(p_{\Delta_{[i+1:m]}}\) into \(p_s\). Thus we may split \(A_7\) into \(A_{7a}\) and \(A_{7b}\) as shown below. We emphasize that \(|\Delta'_0| = |\Psi_7a|\).

With these changes, the basis for rowspan \((M_4)\) can be written as:

128
Finally, we consider Step 3. Out of these two component spaces, we retain only the latter as new words, we can assume that the form \( \sigma(\cdot) \) depends only on the remaining variables of \( u \) denoted as \( u_\Gamma \). By a change of basis on \( u_\Gamma \) variables, we can write, for some \( \Gamma_0 \subseteq \Gamma \),

\[
\text{Span}(A_{11}) = \text{Span}((\{ u_i + \sigma(q_{02}, v) | i \in \Gamma_0 \}) + \text{Span}(A_{11}) \cap \text{Span}(q_{02}, v)).
\]

Out of these two component spaces, we retain only the latter as new \( A_{11} \) and merge the former with \( A_{12} \) (redefining new \( \Gamma \) as \( \Gamma \cup \Gamma_0 \)). Thus we reach our final form:

\[
A_{11} = \begin{cases} 
\{ a_i + y_j + z_k | i \in \Sigma_1, \ j \in \Phi_1, \ k \in \Psi_1 \} \\
\{ a_i + y_j + z_k + \sigma(q, w) | i \in \Sigma_2, \ j \in \Phi_2 \} \\
\{ a_i + z_k + \sigma(p, u) | i \in \Sigma_3, \ k \in \Psi_3 \} \\
\{ a_i + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Psi_5}) | i \in \Sigma_4 \} \\
\{ y_j + z_k | j \in \Phi_5, \ k \in \Psi_5 \} \\
\{ y_j + \sigma(q, w) | j \in \Phi_6 \} \\
\{ z_k + p_s | k \in \Psi_{7a}, s \in \Delta'_0 \} \\
\{ z_k + \sigma(p_{\Delta_{[2:m]}}, u) | k \in \Psi_{7b} \}
\end{cases}
\]

\[
A_{12} = \begin{cases} 
\{ u_i + \sigma(v, w) | i \in \Gamma \}
\end{cases}
\]

\[
A_{13} = \text{basis(rowspan}(M_4) \cap \text{Span}(v, w)).
\]

(4.20)
4.6 Four-wise Correlations: A (somewhat) Simplified Case

We now begin the proof of our main technical Theorem 4.2.15, i.e. to upper bound the expectation

$$
E_{x \in \{0,1\}^k} [f_{i=1}(M_1 x) f_{i=2}(M_2 x) f_{i=3}(M_3 x) f_{i=4}(M_4 x)].
$$

(4.21)

For the benefit of the reader, this section presents the proof in the special case where in the basis for $\text{rowspan}(M_4)$ given in (4.19) (ignore Section 4.5.5 and modifications therein for now):

- All the linear forms $\sigma(\cdot)$ are zero.
- $A_{10} = A_{11} = A_{13} = \phi$. $A_9$ consists of just $B_2$.
- There is no further partition of $A_7$ into $A_{7a}$ and $A_{7b}$.

Given matrices $M_1, M_2, M_3, M_4$, we note:

- Let $H_4 = g = \{g_1, \ldots, g_{h_4}\}$ be the rows that appeared in all four matrices (and were removed).
- Let $H_3 = r = \{r_1, \ldots, r_{h_3}\}$ be the rows that appeared in (exactly) three matrices (and were removed).

Let $r(1), r(2), r(3), r(4) \subseteq H_3$ be the sets of rows that appeared in the four matrices respectively, so that $|r(1)| + |r(2)| + |r(3)| + |r(4)| = 3 \cdot h_3$.

- When we take expectation over $x \in \{0,1\}^k$, if $w$ is a row of a matrix, we make the change of basis $w' = \langle w, x \rangle$ where $w'$ is uniformly distributed over $\{0,1\}^k$ and moreover independently for rows that are linearly independent. For ease of notation, we drop the prime from the superscript and call the new variable $w$ as well.

Thus we assume that:

$$
M_1 x = g, r(1), v, p, u, y
$$

$$
M_2 x = g, r(2), v, q, w, z
$$

$$
M_3 x = g, r(3), a, p_1 + q_1, \ldots, p_n + q_n, u, w
$$

$$
M_4 x = g, r(4), A_1, \ldots, A_9, A_{12}.
$$

(4.22)

We recall that (in the present special case):

$$
A_1 = \{ a_i + y_j + z_k \mid i \in \Sigma_1, \ j \in \Phi_1, \ k \in \Psi_1 \}
$$

$$
A_2 = \{ a_i + y_j \mid i \in \Sigma_2, \ j \in \Phi_2 \}
$$

$$
A_3 = \{ a_i + z_k \mid i \in \Sigma_3, \ k \in \Psi_3 \}
$$

$$
A_4 = \{ a_i \mid i \in \Sigma_4 \}
$$

$$
A_5 = \{ y_j + z_k \mid j \in \Phi_5, \ k \in \Psi_5 \}
$$

$$
A_6 = \{ y_j \mid j \in \Phi_6 \}
$$

$$
A_7 = \{ z_k \mid k \in \Psi_7 \}
$$

$$
A_8 = \{ p_i \mid i \in \Delta_1 \}
$$

$$
A_9 = \{ q_j \mid j \in \Delta_2 \}
$$

$$
A_{12} = \{ u_i \mid i \in \Gamma \}.
$$
Lemma 4.6.1. The dimension $d$ of $\oplus_{j=1}^4 (\text{rowspan}(M_j))$ is:

$$d = |g| + |r| + |v| + (|p| + |q|) + (|u| + |w| + (|a| + |y| + |z|)

$$

$$= |g| + |r| + |v| + (2|\Delta_0| + 2|\Delta_1| + 2|\Delta_2|) + (|\Gamma| + |\bar{\Gamma}| + |w| +

(3|\Sigma_1| + 2|\Sigma_2| + 2|\Sigma_3| + |\Sigma_4| + 2|\Phi_3| + |\Phi_6| + |\Psi_7|).$$

Proof. The number of all the variables appearing above are added up. It is noted that the $p$ and $q$ variables both equal in number to $|\Delta_0| + |\Delta_1| + |\Delta_2|$. Also, $|\Sigma_1| = |\Phi_1| = |\Psi_1|$ and similar equalities. We note that the input $u$ is partitioned as $(u_\Gamma, u_\Psi)$.

We split inputs $M_1x, \ldots, M_4x$ in (4.22) into three parts: Fourier analysis will be applied on the third part, Cauchy-Schwartz on the second part, and the first part will be thought of as a “restriction”. The splits are as below. To clarify the notation, $p_{\Delta_0}$ denotes, as before, the variables $\{p_i \mid i \in \Delta_0\}$, $(p + q)_{\Delta_0}$ denotes the variables $\{p_j + q_j \mid j \in \Delta_0\}$, and for ease of notation, $(a + y + z)_1$ denotes the triples $\{a_i + y_j + z_k \in A_1\}$ (and similarly).

$$L \quad J \quad K$$

$$M_1x =: \{g, r(1), u_\Gamma\} \quad \{v, p_{\Delta_1 \cup \Delta_2}, u_\Psi, y_{\Phi_0}\} \quad \{p_{\Delta_0}, y_{\Phi_1}, y_{\Phi_2}, y_{\Phi_3}\}$$

$$M_2x =: \{g, r(2)\} \quad \{v, q_{\Delta_1 \cup \Delta_2}, w, z_{\Psi_1}\} \quad \{q_{\Delta_0}, z_{\Psi_1}, z_{\Psi_2}, z_{\Psi_3}\}$$

$$M_3x =: \{g, r(3), u_\Gamma\} \quad \{a_{\Sigma_1}, (p + q)_{\Delta_1 \cup \Delta_2}, u_\Psi, w\} \quad \{a_{\Sigma_1}, a_{\Sigma_2}, a_{\Sigma_3}, (p + q)_{\Delta_0}\}$$

$$M_4x =: \{g, r(4), u_\Gamma\} \quad \{a_{\Sigma_1}, y_{\Phi_0}, z_{\Psi_1}, p_{\Delta_1}, q_{\Delta_2}\} \quad \{(a + y + z)_1, (a + y)_2, (a + z)_3, (y + z)_5\}$$

Denoting the parts in the splits as $(L_1, J_1, K_1), \ldots, (L_4, J_4, K_4)$ respectively, consider the restrictions:

$$\lambda_{1,L_1,J_1}(K_1) = f_{=i}(L_1, J_1, K_1), \ldots, \lambda_{4,L_4,J_4}(K_4) = f_{=i}(L_4, J_4, K_4).$$

Dropping the subscripts (but keeping in mind that they are always there), the goal is to upper bound

$$\mathbb{E}[\lambda_1(K_1)\lambda_2(K_2)\lambda_3(K_3)\lambda_4(K_4)],$$

where for notational ease, we did not write the long list of variables that the expectation is taken over. We do note that $L_s, J_s, K_s$ all depend on the inputs. Writing the $K_s$ explicitly:

$$\lambda_1(\quad p_{\Delta_0}, \quad y_{\Phi_1}, \quad y_{\Phi_2}, \quad y_{\Phi_3}, \quad )$$

$$\lambda_2(\quad q_{\Delta_0}, \quad z_{\Psi_1}, \quad z_{\Psi_2}, \quad z_{\Psi_3}, \quad )$$

$$\lambda_3(\quad a_{\Sigma_1}, \quad a_{\Sigma_2}, \quad a_{\Sigma_3}, \quad (p + q)_{\Delta_0} \quad )$$

$$\lambda_4(\quad a_{\Sigma_1} + y_{\Phi_1}, \quad a_{\Sigma_2} + y_{\Phi_2}, \quad a_{\Sigma_3} + z_{\Psi_1}, \quad y_{\Phi_3} + z_{\Psi_5} \quad ).$$

The notation (and similar ones) $a_{\Sigma_1} + y_{\Phi_1} + z_{\Psi_1}$ is imprecise, but we use it for the ease. It really refers to $\{a_i + y_j + z_k \in A_1 \mid i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1\}$. Now writing the $\lambda_s$ in the Fourier representation and taking
expectation over its inputs, we see that the expectation in (4.23) equals (there is a product of four terms that are written one below the other for visual ease)

$$
\mathbb{E} \left[ \sum_{W,Y,Z,S,B} \begin{bmatrix}
\hat{\lambda}_1(S, W, Y, B) \\
\hat{\lambda}_2(S, W, Z, B) \\
\hat{\lambda}_3(W, Y, Z, S) \\
\hat{\lambda}_4(W, Y, Z, B).
\end{bmatrix} \right]
$$

(4.24)

To explain the reasoning, we note that the Fourier expansion will have a term (as part of a larger product term)

$$
\cdots \cdots \chi_W(a_{\Sigma_1} + y_{\phi_1} + z_{\phi_1}) \chi_W'(y_{\phi_1}) \chi_{W''}(z_{\phi_1}) \chi_{W'''}(a_{\Sigma_1}) \cdots
$$

and taking expectation over $a_{\Sigma_1}, y_{\phi_1}, z_{\phi_1}$, the term vanishes unless $W = W' = W'' = W'''$. Similar reasoning is applied above to “Fourier tuples” $Y, Z, S, B$.

For fixed $L_1, \ldots, L_4, W, Y, Z, S, B$, we consider the expectation over $J_1, \ldots, J_4$ (or rather inputs in those sets). The point here is that all inputs in $J_1, \ldots, J_4$ appear twice:

- Exactly twice, these being $\{v, u_T, y_{\phi_0}, w, z_{\phi_7}, a_{\Sigma_4}\}$.

- Or “effectively” exactly twice, these being $p_{\Delta_1}, p_{\Delta_2}, q_{\Delta_1}, q_{\Delta_2}$. What we mean here is that for indices in $\Delta_1$ (and similarly in $\Delta_2$), we have inputs $p_{\Delta_1}, q_{\Delta_1}, (p + q)_{\Delta_1}, p_{\Delta_1}$ appearing in $J_1, J_2, J_3, J_4$ respectively. These can be paired as $(p_{\Delta_1}, q_{\Delta_1})$ and $(p + q)_{\Delta_1}, p_{\Delta_1})$. The latter pair is distributed the same as the former and this is what matters for applying Cauchy-Schwarz.

Replacing the Fourier coefficients by their absolute values and using repeated Cauchy-Schwartz (see Lemma 4.8.4), we see that (4.24) is upper bounded by

$$
\mathbb{E}_{g,r,u_T} \left[ \sum_{W,Y,Z,S,B} \sqrt{\mathbb{E}_{J_1} \left[ \hat{\lambda}^2_{1,J_1}(S, W, Y, B) \right]} \sqrt{\mathbb{E}_{J_2} \left[ \hat{\lambda}^2_{2,J_2}(S, W, Z, B) \right]} \right. \\
\left. \sqrt{\mathbb{E}_{J_3} \left[ \hat{\lambda}^2_{3,J_3}(W, Y, Z, S) \right]} \sqrt{\mathbb{E}_{J_4} \left[ \hat{\lambda}^2_{4,J_4}(W, Y, Z, B) \right]} \right].
$$

Again applying Cauchy-Schwartz (note that the pairing is first-third and fourth-second factors) we get an upper bound $\sqrt{\text{Term}_1 \cdot \text{Term}_2}$ where

$$
\text{Term}_1 = \mathbb{E}_{g,r,u_T} \left[ \sum_{W,Y,Z,S,B} \mathbb{E}_{J_1} \left[ \hat{\lambda}^2_{1,J_1}(S, W, Y, B) \right] \mathbb{E}_{J_3} \left[ \hat{\lambda}^2_{3,J_3}(W, Y, Z, S) \right] \right]
$$

$$
\text{Term}_2 = \mathbb{E}_{g,r,u_T} \left[ \sum_{W,Y,Z,S,B} \mathbb{E}_{J_2} \left[ \hat{\lambda}^2_{2,J_2}(S, W, Z, B) \right] \mathbb{E}_{J_4} \left[ \hat{\lambda}^2_{4,J_4}(W, Y, Z, B) \right] \right].
$$

We consider $\text{Term}_1$. Noting that $W, Y, S$ appear in both $\hat{\lambda}_1(\cdot), \hat{\lambda}_3(\cdot), B$ appears only in $\hat{\lambda}_1(\cdot)$, $Z$ appears only in $\hat{\lambda}_3(\cdot)$, and that $\lambda_3(\cdot)$ does not depend on $r(1) \setminus r(3)$ (so expectation over it can be pushed inside), we can rewrite $\text{Term}_1$ as:

$$
\text{Term}_1 = \mathbb{E}_{g,r,u_T} \left[ \sum_{W,Y,S} \left( \mathbb{E}_{r(1) \setminus r(3),J_1} \left[ \sum_{B} \hat{\lambda}^2_{1,J_1}(S, W, Y, B) \right] \right) \left( \mathbb{E}_{J_3} \left[ \sum_{Z} \hat{\lambda}^2_{3,J_3}(W, Y, Z, S) \right] \right) \right].
$$
Finally, using Lemma 4.8.5, we have the upper bound:

\[
\text{Term}_1 \leq \left( \max_{W,Y,S} \mathbb{E}_{g,r(1)\cup r(3),u_r, r(1)\cup r(3),J_1} \left[ \sum_B \tilde{\lambda}_{1,j_1}^2 (S,W,Y,B) \right] \right) \left( \mathbb{E}_{g_r,u_r,J_3} \left[ \sum_{W,Y,S,Z} \tilde{\lambda}_{3,j_3}^2 (W,Y,Z,S) \right] \right).
\]

The second factor is \( \mathbb{E}_{L_3,J_3,K_3} [||\lambda_{3,L_3,J_3}(K_3)||_2^2] = \|f_i\|_2^2 \leq \frac{2^2}{2\pi} \eta \). The first factor is bounded by, using Lemma 4.3.20, \( 2^{\frac{6\epsilon}{2\pi}} \) where

\[
d_1 = (|J_1| + |r(1) \setminus r(3)|) + 2(|W| + |Y| + |S|) + |B|
\]

\[
= |v| + |\Delta_1 + |\Delta_2| + |\Gamma| + |\Phi_6| + |r(1) \setminus r(3)| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0| + |\Phi_5|.
\]

We similarly rewrite \( \text{Term}_2 \) as:

\[
\text{Term}_2 = \mathbb{E}_{g_r,u_r} \left[ \sum_{W,Z,B} \left( \mathbb{E}_{J_4,j_4,J_5} \left[ \sum_Y \tilde{\lambda}_{4,j_4}^2 (W,Y,Z,B) \right] \right) \left( \mathbb{E}_{J_2} \sum_S \tilde{\lambda}_{2,j_2}^2 (S,W,Z,B) \right) \right].
\]

Here \( \lambda_2(\cdot) \) does not depend on \( u_r \) and \( r(4) \setminus r(2) \), so both are pushed inside. As before,

\[
\text{Term}_2 \leq \left( \max_{g,r(4)\cup r(2),W,Z,B} \mathbb{E}_{J_4,j_4,J_5} \left[ \sum_Y \tilde{\lambda}_{4,j_4}^2 (W,Y,Z,B) \right] \right) \left( \mathbb{E}_{g_r,u_r,J_2} \sum_{W,Z,B,S} \tilde{\lambda}_{2,j_2}^2 (S,W,Z,B) \right).
\]

The second factor is \( \mathbb{E}_{L_2,J_2,K_2} [||\lambda_{2,L_2,J_2}(K_2)||_2^2] = \|f_i\|_2^2 \leq \frac{2^2}{2\pi} \eta \). The first factor is bounded by, using Lemma 4.3.20, \( 2^{\frac{6\epsilon}{2\pi}} \) where

\[
d_2 = (|J_4| + |r(4) \setminus r(2)| + |\Gamma|) + 2(|W| + |Z| + |B|) + |Y|
\]

\[
= |\Sigma_1| + |\Phi_6| + |\Psi_7| + |\Delta_1| + |\Delta_2| + |r(4) \setminus r(2)| + |\Gamma| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Delta_0| + |\Phi_5| + |\Sigma_2|.
\]

The proof of Theorem 4.2.15 (in the special case) is complete by recalling that we have an upper bound of \( \sqrt{\text{Term}_1} \sqrt{\text{Term}_2} \) and that \( i \leq r \) and \( \frac{1}{2}((d_1 + i) + (d_2 + i)) = d \) as below. Combining this, we get an upper bound of \( \frac{2^2}{2\pi} \eta \varepsilon \) (as required in Theorem 4.2.15).

**Lemma 4.6.2.** \( d_1 + i + d_2 + i = 2d \).

**Proof.** We write down expressions for \( d_1, d_2 \) as above followed by expressions for \( i = |L_3 \cup J_3 \cup K_3| \) and \( i = |L_2 \cup J_2 \cup K_2| \):

\[
d_1 = |v| + |\Delta_1| + |\Delta_2| + |\Gamma| + |\Phi_6| + |r(1) \setminus r(3)| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0| + |\Phi_5|.
\]

\[
d_2 = |\Sigma_4| + |\Phi_6| + |\Sigma_6| + |\Delta_1| + |\Delta_2| + |r(4) \setminus r(2)| + |\Gamma| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Delta_0| + |\Phi_5| + |\Sigma_2|.
\]

\[
i = |g| + |r(3)| + |\Gamma| + |\Sigma_4| + |\Delta_1| + |\Delta_2| + |\Gamma| + |w| + |\Sigma_1| + |\Delta_2| + |\Delta_0| + |\Psi_7| + |\Phi_6|.
\]

\[
i = |g| + |r(4)| + |v| + |\Delta_1| + |\Delta_2| + |w| + |\Psi_7| + |\Phi_6| + |\Sigma_4| + |\Delta_0| + |\Phi_5|.
\]

It can be verified that the overall sum is exactly \( 2d \) where as in Lemma 4.6.1,

\[
d = |g| + |r| + |v| + 2|\Delta_0| + 2|\Delta_1| + 2|\Delta_2| + |\Gamma| + |\w| + 3|\Sigma_1| + 2|\Sigma_2| + 2|\Sigma_3| + |\Sigma_4| + 2|\Phi_5| + |\Phi_6| + |\Psi_7|.
\]

One notes that since every element of \( r = r(1) \cup r(2) \cup r(3) \cup r(4) \) is contained in precisely three of these sets, \( |r| = |r(3)| + |r(1) \setminus r(3)| + |r(2)| + |r(4) \setminus r(2)| \). \( \square \)
4.7 Four-wise Correlations: the General Case

We now begin the full proof of our main technical Theorem 4.2.15, i.e. to upper bound the expectation

\[ \mathbb{E}_{x \in \{0,1\}^k} [f_{-i}(M_1x)f_{-i}(M_2x)f_{-i}(M_3x)f_{-i}(M_4x)]. \]  

(4.25)

Given matrices \( M_1, M_2, M_3, M_4 \), we recall:

- Let \( H_4 = g = \{ g_1, \ldots, g_{h_4} \} \) be the rows that appeared in all four matrices (and were removed).
- Let \( H_3 = r = \{ r_1, \ldots, r_{h_3} \} \) be the rows that appeared in (exactly) three matrices (and were removed). Let \( r(1), r(2), r(3), r(4) \subseteq H_3 \) be the sets of rows that appeared in the four matrices respectively, so that \(|r(1)| + |r(2)| + |r(3)| + |r(4)| = 3 \cdot h_3.\)
- When we take expectation over \( x \in \{0,1\}^k \), if \( w \) is a row of a matrix, we make the change of basis \( w' = \langle w, x \rangle \) where \( w' \) is uniformly distributed over \( \{0,1\}^k \) and moreover independently for rows that are linearly independent. For ease of notation, the prime is dropped from the superscript, the new variable is called \( w \) as well.

Thus we assume that (given the basis for rowspan(\( M_4 \)) by (4.20), repeated next):

\[
\begin{align*}
M_1x &= g, r(1), v, p, u, y \\
M_2x &= g, r(2), v, q, w, z \\
M_3x &= g, r(3), a, p_1 + q_1, \ldots, p_n + q_n, u, w \\
M_4x &= g, r(4), A_1, \ldots, A_6, A_{7a}, A_{7b}, A_8, \ldots, A_{13}. 
\end{align*}
\]

(4.26)

**Lemma 4.7.1.** The dimension \( d \) of \( \oplus_{j=1}^4 (\text{rowspan}(M_j)) \) is:

\[
d = |g| + |r| + |v| + (|p| + |q|) + (|u|) + |w| + (|a| + |y| + |z|)
\]

\[
= |g| + |r| + |v| + (2 |\Delta_0'| + 2 |\Delta''_0| + 2 |\Delta_1| + 2 |\Delta_2| + \ldots + 2 |\Delta_m| + 2 |\Omega_1| + 2 |\Omega_2|) + (|\Gamma| + |T|) + |w| + (3 |\Sigma_1| + 2 |\Sigma_2| + 2 |\Sigma_3| + |\Sigma_4| + 2 |\Phi_5| + |\Phi_6| + |\Psi_{7a}| + |\Psi_{7b}|).
\]

**Proof.** The number of all the variables appearing above are added up. It is noted that the \( p \) and \( q \) variables both equal in number to \(|\Delta_0'| + |\Delta''_0| + |\Delta_1| + |\Delta_2| + \ldots + |\Delta_m| + |\Omega_1| + |\Omega_2| \) and \( \Delta_0 = \Delta_0' \cup \Delta''_0 \). We have \(|\Sigma_1| = |\Phi_1| = |\Psi_1| \) and similar equalities. We emphasize that \(|\Delta_0'| = |\Psi_{7a}|. \)

Recall that:

\[
\begin{align*}
A_1 &= \{ a_i + y_j + z_k, & i \in \Sigma_1, & j \in \Phi_1, & k \in \Psi_1 \} \\
A_2 &= \{ a_i + y_j + \sigma( q, w), & i \in \Sigma_2, & j \in \Phi_2 \} \\
A_3 &= \{ a_i + z_k, & i \in \Sigma_3, & k \in \Psi_3 \} \\
A_4 &= \{ a_i + \sigma( p, u, q, w) + \sigma( y_{\Phi_1}, y_{\Phi_5}), & i \in \Sigma_4 \} \\
A_5 &= \{ y_j + z_k, & j \in \Phi_5, & k \in \Psi_5 \} \\
A_6 &= \{ y_j + \sigma( q, w), & j \in \Phi_6 \} \\
A_{7a} &= \{ z_k + p_s, & k \in \Psi_{7a}, & s \in \Delta_0' \} \\
A_{7b} &= \{ z_k + \sigma( p_{\Delta_2}, u), & k \in \Psi_{7b} \} \\
A_8 &= \{ p_i + \sigma(q), & i \in \Delta_1 \} \\
A_9 &= B_2 \cup \ldots \cup B_m \\
A_{10} &= \{ p_i + \sigma(q_{\Omega_1}), & i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} &\subseteq \text{Span}(q_{\Omega_2}, v).
\end{align*}
\]
Here \( B_s = \left\{ q_j + \sigma(p_{\Delta_{i+1,m_i}}) + \sigma(v, u) \mid j \in \Delta_s \right\} \).

\[
A_{12} = \{ u_i + \sigma(q_{\theta_{2i}}, v, w) \mid i \in \Gamma \}
\]
\[
A_{13} = \text{basis(rowspan}(M_4) \cap \text{Span}(v, w)).
\]

We split each input in (4.26) into two parts. The mix of Fourier analysis and Cauchy-Schwartz will not be very clean. The splits are as below.

\[
\begin{align*}
J & \\
M_1x & =: \{ g, r(1), v, p_{\Delta_{[2,m]}}, u_1, u_1' \} \\
M_2x & =: \{ g, r(2), v, q_{\Delta_{[2,m]}}, w, z_{\psi_{i_1}} \} \\
M_3x & =: \{ g, r(3), (p + q)_{\Delta_{[2,m]}}, u_1, u_1', w \} \\
M_4x & =: \{ g, r(4), z_{\psi_{i_2}}, A_9, A_{10}, A_{11}, A_{12}, A_{13} \}
\end{align*}
\]

\[
\begin{align*}
K & \\
& \{ p_{\Delta_{0} \cup \Delta_{1}}, y_{\Phi_{1}}, y_{\Phi_{2}}, y_{\Phi_{5}}, y_{\Phi_{6}} \} \\
& \{ q_{\Delta_{0} \cup \Delta_{1}}, z_{\psi_{1}}, z_{\psi_{3}}, z_{\psi_{5}}, z_{\psi_{7a}} \} \\
& \{ a, (p + q)_{\Delta_{0} \cup \Delta_{1}} \} \\
& \{(a + y + z)_1, (a + y)_2, (a + z)_3, a_{\Sigma_4}, (y + z)_5, y_{\Phi_6}, z_{\psi_{7a}} + p_{\Delta_{0}', p_{\Delta_{1}}} \}
\end{align*}
\]

We are using an imprecise notation: inputs for \( M_4x \) (except for \( g, r(4), (a+y+z)_1, (y+z)_5 \)) have the additional \( \sigma(\cdot) \) terms that are omitted from the notation. Denoting the parts in the splits as \( (J_1, K_1), \ldots, (J_4, K_4) \) respectively, consider the restrictions:

\[
\lambda_{1,J_1}(K_1) = f_{=i}(J_1, K_1), \ldots, \lambda_{4,J_4}(K_4) = f_{=i}(J_4, K_4).
\]

Dropping the subscripts (but keeping in mind that they are always there), the goal is to upper bound

\[
\mathbb{E} [\lambda_1(K_1)\lambda_2(K_2)\lambda_3(K_3)\lambda_4(K_4)],
\]

(4.27)

where for ease of notation, we did not write the long list of variables that the expectation is taken over. We do note that \( J_s, K_s \) all depend on the inputs. Writing the \( K_s \) explicitly:

\[
\begin{align*}
\lambda_1 & ( \ p_{\Delta_{0}'}, \ p_{\Delta_{0}'}, \ p_{\Delta_{1}}, \ y_{\Phi_{1}}, \ y_{\Phi_{2}}, \ y_{\Phi_{5}}, \ y_{\Phi_{6}} ) \\
\lambda_2 & ( \ q_{\Delta_{0}'}, \ q_{\Delta_{0}'}, \ q_{\Delta_{1}}, \ z_{\psi_{1}}, \ z_{\psi_{3}}, \ z_{\psi_{5}}, \ z_{\psi_{7a}} ) \\
\lambda_3 & ( \ a_{\Sigma_1}, \ a_{\Sigma_2}, \ a_{\Sigma_3}, \ a_{\Sigma_4}, \ (p + q)_{\Delta_{0}'}, \ (p + q)_{\Delta_{0}'}, \ (p + q)_{\Delta_{1}} ) \\
\lambda_4 & ( \ a_{\Sigma_1} + y_{\Phi_{1}} + z_{\psi_{1}}, \ a_{\Sigma_2} + y_{\Phi_{2}}, \ a_{\Sigma_3} + z_{\psi_{3}}, \ a_{\Sigma_4}, \ y_{\Phi_{5}} + z_{\psi_{5}}, \ y_{\Phi_{6}}, \ z_{\psi_{7a}} + p_{\Delta_{0}'}, \ p_{\Delta_{1}} ).
\end{align*}
\]

Now writing the \( \lambda_s \) in the Fourier representation and taking expectation over their inputs, we see that the expectation in (4.27) equals, up to a sublety to be fixed shortly, (there is a product of four terms that are written one below the other for visual ease)
Denoting the Fourier coefficients as $\lambda_i$ with

$$
\sum_{W,Y,Z,B,P,T,Q,N,S,D,X} \text{sign} \cdot \\
\lambda_1( S + Q, D, X + N, W, Y, P, T ) \\
\lambda_2( S, D, X, W, Z, P, Q ) \\
\lambda_3( W, Y, Z, B, S, D, X ) \\
\lambda_4( W, Y, Z, B, P, T, Q, N )
$$

(4.28)

A remark: there are $\sigma(\cdot)$ terms that were omitted from the notation. They have a two-fold effect. Firstly, there is a sign $\in \{-1, 1\}$ that depends on $(Y, Z, B, T, N, S, D, X; v, u, w, p, q)$. We will take absolute values immediately next, so this sign does not really matter. Secondly, there are additional $\sigma(\cdot)$ terms now in the Fourier domain, and the form of the Fourier coefficients is not quite as in (4.28), but actually as below:

\[
\begin{align*}
\hat{\lambda}_1( & S + Q, D, X + N, W, Y, P, T ) \\
\hat{\lambda}_2( & S, D, X, W, Z, P, Q ) \\
\hat{\lambda}_3( & W, Y, Z, B, S, D, X ) \\
\hat{\lambda}_4( & W, Y, Z, B, P, T, Q, N ).
\end{align*}
\]

Denoting the Fourier coefficients as $\hat{\lambda}_1(V_1), \hat{\lambda}_2(V_2), \hat{\lambda}_3(V_3), \hat{\lambda}_4(V_4)$, an upper bound on the desired expectation is:

\[
\mathbb{E}_{g,r,v,w,u,\tau,w} \mathbb{E}_{\Delta[2,m] \cdot \Psi_{\tau}} \sum_{W,Y,Z,B,P,T,Q,N,S,D,X} \left[ |\hat{\lambda}_1(V_1)| \cdot |\hat{\lambda}_2(V_2)| \cdot |\hat{\lambda}_3(V_3)| \cdot |\hat{\lambda}_4(V_4)| \right]
\]

We note that $\Psi_{\tau}$ appears only in $J_2, J_4$. Using Cauchy-Schwartz, we get an upper bound

\[
\mathbb{E}_{g,r,v,w,u,\tau,w} \mathbb{E}_{\Delta[2,m] \cdot \Psi_{\tau}} \sum_{W,Y,Z,B,P,T,Q,N,S,D,X} \left[ |\hat{\lambda}_1(V_1)| \cdot |\hat{\lambda}_2(V_2)| \cdot |\hat{\lambda}_3(V_3)| \cdot |\hat{\lambda}_4(V_4)| \right].
\]

A point to note here is as follows: in $\lambda_4$, the variables $\Psi_{\tau}$ actually appear along with additional $\sigma(\Delta[2,m], u)$ terms. However the expectation over these additional variables is still not considered and is still at the “outer” level. Hence the Cauchy-Schwartz over $\Psi_{\tau}$ can be safely applied. Moreover, once Cauchy-Schwartz, i.e., expectation over $\Psi_{\tau}$, is applied, we can ignore these $\sigma(\cdot)$ terms henceforth.\(^{20}\) We will use this trick repeatedly.

\(^{20}\) Formally, if one wishes to, by change of variables $\Psi_{\tau} \sim \Psi_{\tau} + \sigma(\Delta[2,m], u)$.
Next, we consider the variables \((p_{\Delta_2}, q_{\Delta_2}), \ldots, (p_{\Delta_m}, q_{\Delta_m})\), one pair at a time. Let’s consider \((p_{\Delta_2}, q_{\Delta_2})\) as an illustration. We note that \(p_{\Delta_2}\) appears in \(J_1\), \(q_{\Delta_2}\) appears in \(J_2\), \((p+q)_{\Delta_2}\) appears in \(J_3\) and \(q_{\Delta_2}\) appears in \(J_4\). We note two points. In \(J_3\), the distribution of \((p+q)_{\Delta_2}\) is the same as that of \(p_{\Delta_2}\). In \(J_4\), there are additional \(\sigma(p_{\Delta_{[3,m]}}, v, u)\) terms but the expectation over these variables is still at the outer level. Thus we may safely apply Cauchy-Schwarz over \((p_{\Delta_2}, q_{\Delta_2})\), pairing the first-second and third-fourth factors, ignore the \(\sigma(\cdot)\) terms henceforth, and get the upper bound

\[
\mathbb{E}_{g,r,v,u,\tau \mid w \mid p_{\Delta_{[3,m]}}, q_{\Delta_{[3,m]}}} \sum_{W,Y,Z,B,P, T,Q,N,S,D,X} \left( \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_1^2(V_1) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_2^2(V_2) \right] \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_3^2(V_3) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_4^2(V_4) \right] \right). 
\]

We apply the same argument iteratively to get an upper bound

\[
\mathbb{E}_{g,r,v,u,\tau \mid \alpha \mid \beta} \sum_{W,Y,Z,B,P, T,Q,N,S,D,X} \left( \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_1^2(V_1) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_2^2(V_2) \right] \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_3^2(V_3) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_4^2(V_4) \right] \right). 
\]

Next, we Cauchy-Schwarz over \(\alpha\). This is possible since it appears explicitly only in \(J_1, J_3\). It appears in \(J_4\) implicitly as part of several \(\sigma(\cdot)\) terms, but all those terms got “ignored” or “eliminated” in prior steps! Hence we get an upper bound

\[
\mathbb{E}_{g,r,v,u,\tau \mid w \mid p_{\Delta_{[3,m]}}, q_{\Delta_{[3,m]}}} \sum_{W,Y,Z,B,P, T,Q,N,S,D,X} \left( \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_1^2(V_1) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_2^2(V_2) \right] \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_3^2(V_3) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_4^2(V_4) \right] \right). 
\]

Finally, we apply Cauchy-Schwarz twice (the pairing is first-third and fourth-second factors) to get an upper bound \(\sqrt{\text{Term}_1 \cdot \sqrt{\text{Term}_2}}\) where

\[
\text{Term}_1 = \mathbb{E}_{g,r,v,u,\tau \mid w \mid p_{\Delta_{[3,m]}}, q_{\Delta_{[3,m]}}} \sum_{W,Y,Z,B,P, T,Q,N,S,D,X} \left( \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_1^2(V_1) \right] \mathbb{E}_{p_{\Delta_2}} \left[ \hat{\lambda}_3^2(V_3) \right] \right)
\]

\[
\text{Term}_2 = \mathbb{E}_{g,r,v,u,\tau \mid w \mid p_{\Delta_{[3,m]}}, q_{\Delta_{[3,m]}}} \sum_{W,Y,Z,B,P, T,Q,N,S,D,X} \left( \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_2^2(V_2) \right] \mathbb{E}_{q_{\Delta_2}} \left[ \hat{\lambda}_4^2(V_4) \right] \right). \quad (4.30)
\]

**Lemma 4.7.2.** We have the upper bound \(\text{Term}_1 \leq 2^{7T^3 \frac{n^2}{2(4^1 + \tau)}}\) where

\[
d_1 = |r(1)| + |r(3)| + |\Delta_2\cdot m| + |\Omega_1| + |\Omega_2| + |v| + |\tau| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0''| + |\Psi_7a| + |\Delta_1| + |\Phi_5| + |\Phi_6|.
\]

**Proof.** Let us recall the the definitions of \(V_1, V_3:\)

\[
V_1 = (S + Q + \sigma(B, Z), \ D + \sigma(B, Z), \ X + N + \sigma(B, Z), \ W + \sigma(B), \ P + \sigma(B), \ Y, \ T),
\]

\[
V_3 = (W, Y, Z, B, S, D, X).
\]

Since \(P, Q, N\) do not appear in \(V_3\) and we will only be concerned about summing over all possibilities, we might as well take \(V_1\) as

\[
V_1 = (Q, \ D + \sigma(B, Z), \ N, \ W + \sigma(B), \ P, Y, T).
\]

137
Further in $V_1$, we may replace $D + \sigma(B, Z)$ by $D$ and $W + \sigma(B)$ by $W$. This will induce a change in $V_3$, but since $B, Z$ are present therein and the Fourier coefficients are basis invariant, the $\sigma(B, Z), \sigma(Z)$ terms there can be cleared. To summarize, we may assume that $V_1$ and $V_3$ are:

$$V_1 = (Q, D, N, W, P, Y, T), \quad V_3 = (W, Y, Z, B, S, D, X).$$

Noting that $W, Y, D$ are common to $V_1$ and $V_3$, we may thus write

$$\text{Term}_1 = \mathbb{E}_{q, u} \sum_{r(3)} \left[ \mathbb{E}_{v, r(1) \setminus r(3), p_{\Omega_1 \cup \Omega_2}, \Delta_{2, ..., m}, \Delta_{0}} \left[ \sum_{Q, N, P, T} \hat{\lambda}^2_1(W, Y, D, Q, N, P, T) \right] \right] \left( \mathbb{E}_{[W, Y, D]} \sum_{Z, S, X} \hat{\lambda}^2_3(W, Y, D, B, Z, S, X) \right).$$

The second factor equals (as usual) $\|f_{\eta}\|_2^2 \leq 2^{v^2} \eta$. The first factor is bounded, using Lemma 4.8.5, Term 1 is bounded by

$$\left( \max_{\Delta_{2, ..., m}} \mathbb{E}_{[W, Y, D]} \sum_{Q, N, P, T} \hat{\lambda}^2_1(W, Y, D, Q, N, P, T) \right) \left( \mathbb{E}_{[W, Y, D]} \sum_{Z, S, X} \hat{\lambda}^2_3(W, Y, D, B, Z, S, X) \right).$$

The second factor equals (as usual) $\|f_{\eta}\|_2^2 \leq 2^{v^2} \eta$. The first factor is bounded, using Lemma 4.3.20, by $2^{13} \frac{\epsilon}{2^{n^2}}$ where

$$d_1 = |r(1) \setminus r(3)| + (|\Delta_{2, ..., m}| + |\Omega_1| + |\Omega_2|) + |v| + |\Gamma| + (2|W| + 2|Y| + 2|D|) + |Q| + |N| + |P| + |T|$$

$$= |r(1) \setminus r(3)| + |\Delta_{2, ..., m}| + |\Omega_1| + |\Omega_2| + |v| + |\Gamma| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0| + |\Psi_7| + |\Delta_1| + |\Phi_5| + |\Phi_6|.$$ 

**Lemma 4.7.3.** We have the upper bound $\text{Term}_2 \leq 2^{7v^3} \frac{n e}{2^{n^2}}$ where

$$d_2 = |r(4) \setminus r(2)| + |\Delta_{2, ..., m}| + |\Omega_1| + |\Omega_2| + |\Gamma| + |\Psi_7| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Phi_5| + 2|\Psi_7| + |\Sigma_2| + |\Sigma_4| + |\Phi_6| + |\Delta_1|.$$

**Proof.** Let us recall the definitions of $V_2, V_4$.

$$V_2 = (S + \sigma(Y, B, T, N), D + \sigma(Y, B, T, N), X + \sigma(Y, B, T, N), W, Z, P, Q), \quad V_4 = (W, Y, Z, B, P, T, Q, N).$$

Since $S, D, X$ do not appear in $V_1$, we might as well write $V_2 = (S, D, X, W, Z, P, Q)$. Noting that $W, Z, P, Q$ are common to $V_2$ and $V_4$, we may thus write

$$\text{Term}_2 = \mathbb{E}_{q, v, r(2), u} \sum_{W, Z, P, Q} \left[ \mathbb{E}_{[W, Z, P, Q]} \left[ \sum_{Y, B, T, N} \hat{\lambda}^2_3(W, Z, P, Q, Y, B, T, N) \right] \right] \left( \mathbb{E}_{x} \sum_{S, D, X} \hat{\lambda}^2_3(W, Z, P, S, D, X) \right).$$

138
We have pushed the expectation over $r(4) \setminus r(2), u_{r}, p_{\Delta_{1} \cup \Delta_{2}}$ inside as $\lambda_{2}$ does not depend on them. Using Lemma 4.8.5, we upper bound $\text{Term}_{2}$ by

$$
\max_{g,v,r(4) \setminus r(2), w \in \mathcal{Q}_{1} \cup \mathcal{Q}_{2}, W,Z,P,Q} E_{r(4) \setminus r(2), u_{r}, p_{\Delta_{1} \cup \Delta_{2}}} \left[ \sum_{Y,B, T,N} \sum_{S,D,X} \lambda_{1}^{2} (W, Z, P, Q, Y, B, T, N) \right] \left( \mathbb{E} \left[ \sum_{W,Z,P,Q} \lambda_{2}^{2} (W, Z, P, Q, S, D, X) \right] \right).
$$

As before, the second factor equals $\|f_{\xi} - h\|_{2}^{2} \leq \frac{9\eta}{4\tau}$. The first factor is bounded, using Lemma 4.3.20, by $2^{6i} \frac{\epsilon}{2\tau^2}$ where

$$
d_{2} = |r(4) \setminus r(2)| + |\Delta_{2,...,m}| + |\Omega_{1}| + |\Omega_{2}| + |\Gamma| + |\Psi_{7b}| + (2|W| + 2|Z| + 2|P| + 2|Q|) + |Y| + |B| + |T| + |N|
$$

$$
= |r(4) \setminus r(2)| + |\Delta_{2,...,m}| + |\Omega_{1}| + |\Omega_{2}| + |\Gamma| + |\Psi_{7b}| + 2|\Sigma_{1}| + 2|\Sigma_{2}| + 2|\Phi_{5}| + 2|\Psi_{7a}| + |\Sigma_{2}| + |\Sigma_{4}| + |\Phi_{6}| + |\Delta_{1}|.
$$

We get the overall upper bound $\sqrt{\text{Term}_{1}} \sqrt{\text{Term}_{2}}$, which is at most $2^{7r^{3} \frac{\eta}{2\tau^{4}}}$ (noting $i \leq r$) provided $\frac{1}{2}((d_{1} + i) + (d_{2} + i)) = d$. This is proved below completing the proof of Theorem 4.2.15.

**Lemma 4.7.4.** $d_{1} + i + d_{2} + i = 2d$.

**Proof.** We write down expressions for $d_{1}, d_{2}$ as above followed by expressions for $i$ (from $M_{3}$) and $i$ (from $M_{2}$):

$$
d_{1} = |r(1) \setminus r(3)| + |\Delta_{2,...,m}| + |\Omega_{1}| + |\Omega_{2}| + |v| + |\Gamma| + 2|\Sigma_{1}| + 2|\Sigma_{2}| + 2|\Delta_{0}^{''}| + |\Psi_{7a}| + |\Delta_{1}| + |\Phi_{5}| + |\Phi_{6}|.
$$

$$
d_{2} = |r(4) \setminus r(2)| + |\Delta_{2,...,m}| + |\Omega_{1}| + |\Omega_{2}| + |\Gamma| + |\Psi_{7b}| + 2|\Sigma_{1}| + 2|\Sigma_{2}| + 2|\Sigma_{3}| + 2|\Phi_{5}| + 2|\Psi_{7a}| + |\Sigma_{2}| + |\Sigma_{4}| + |\Phi_{6}| + |\Delta_{1}|.
$$

$$
i = |g| + |r(3)| + |\Sigma_{1}| + |\Sigma_{2}| + |\Sigma_{3}| + |\Sigma_{4}| + |\Delta_{0}^{''}| + |\Delta_{1}| + |\Delta_{2,...,m}| + |\Omega_{1}| + |\Omega_{2}| + |\Gamma| + |\Gamma| + |\Delta_{1}|.
$$

$$
i = |g| + |r(2)| + |v| + |\Delta_{0}^{'}| + |\Delta_{1}| + |\Delta_{2,...,m}| + |\Omega_{1}| + |\Omega_{2}| + |w| + |\Sigma_{1}| + |\Sigma_{3}| + |\Phi_{5}| + |\Psi_{7a}| + |\Psi_{7b}|.
$$

It can be verified that the overall sum is exactly $2d$ where as in Lemma 4.7.1,

$$
d = |g| + |r| + |v| + (2|\Delta_{0}^{''}| + 2|\Delta_{1}| + 2|\Delta_{2,...,m}| + 2|\Omega_{1}| + 2|\Omega_{2}|) + (|\Gamma| + |\Gamma|) + |w| + (3|\Sigma_{1}| + 2|\Sigma_{2}| + 2|\Sigma_{3}| + |\Sigma_{4}| + 2|\Phi_{5}| + |\Phi_{6}| + |\Psi_{7a}| + |\Psi_{7b}|).
$$

One notes that since every element of $r = r(1) \cup r(2) \cup r(3) \cup r(4)$ is contained in precisely three of these sets, $|r| = |r(3)| + |r(1) \setminus r(3)| = |r(2)| + |r(4) \setminus r(2)|$. Also as emphasized before $|\Psi_{7a}| = |\Delta_{0}^{'}|$. \qed
4.8 Appendix: Auxiliary Lemmas

Lemma 4.8.1. Suppose $A, B, C$ are three spaces such that $A \cap B = \{0\}$ and $C \subseteq A \oplus B$. Then sets of vectors can be chosen in the following manner:

- $a_1, \ldots, a_p, a'_1, \ldots, a'_r$ are in $A$ and are linearly independent.
- $b_1, \ldots, b_q, b'_1, \ldots, b'_r$ are in $B$ and are linearly independent.
- $(a_1, \ldots, a_p, b_1, \ldots, b_q, a'_1 + b'_1, \ldots, a'_r + b'_r)$ is a basis for $C$.

Moreover:

- If in addition, $A \subseteq B \oplus C, B \subseteq A \oplus C$,
  - $a_1, \ldots, a_p, a'_1, \ldots, a'_r$ is already a basis for $A$.
  - $b_1, \ldots, b_q, b'_1, \ldots, b'_r$ is already a basis for $B$.
- Otherwise, the sets can (clearly) be extended further so that
  - $a_1, \ldots, a_p, a'_1, \ldots, a'_r, a''_1, \ldots, a''_m$ is a basis for $A$.
  - $b_1, \ldots, b_q, b'_1, \ldots, b'_r, b''_1, \ldots, b''_n$ is a basis for $B$.

Proof. Let $(a_1, \ldots, a_p)$ be a basis for $A \cap C$ and $(b_1, \ldots, b_q)$ be a basis for $B \cap C$. Let $c_1, \ldots, c_r \in C$ be such that $(a_1, \ldots, a_p, b_1, \ldots, b_q, c_1, \ldots, c_r)$ is a basis for $C$. Since $C \subseteq A \oplus B$, $c_j = a'_j + b'_j$ for some $a'_j \in A, b'_j \in B$.

We now prove that $a_1, \ldots, a_p, a'_1, \ldots, a'_r$ are linearly independent. Suppose (on the contrary) that for some index sets $\Phi \subseteq \{1, \ldots, p\}$ and $\Psi \subseteq \{1, \ldots, r\}$, we have $\bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Psi} a'_j = 0$. Consider

$$v = \bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Psi} (a'_j + b'_j) = \left( \bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Psi} a'_j \right) \bigoplus_{j \in \Psi} b'_j = \bigoplus_{j \in \Psi} b'_j.$$ 

Thus we have $v \in C$ as well as $v \in B$ and hence $v \in B \cap C$. Therefore $v = \bigoplus_{j \in \Sigma} b_j$ for some index set $\Sigma \subseteq \{1, \ldots, q\}$ and substituting above

$$\bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Sigma} b_j \bigoplus_{j \in \Psi} (a'_j + b'_j) = 0.$$

This contradicts, unless $\Phi = \Sigma = \Psi = \emptyset$, the assumption that $a_1, \ldots, a_p, b_1, \ldots, b_q, a'_1 + b'_1, \ldots, a'_r + b'_r$ is a basis for $C$ and hence linearly independent.

Finally, we show that if $A \subseteq B \oplus C$, then $a_1, \ldots, a_p, a'_1, \ldots, a'_r$ is in fact a basis for $A$. Indeed, consider any $a \in A$. Since $A \subseteq B \oplus C$, $a = b + c$ for some $b \in B, c \in C$. We write $c$ in the basis for $C$ as $\bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Sigma} b_j \bigoplus_{j \in \Psi} (a'_j + b'_j)$ and hence

$$a = b \bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Sigma} b_j \bigoplus_{j \in \Psi} (a'_j + b'_j).$$

Since $A \cap B = \{0\}$, it follows that $a = \bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Psi} a'_j$. \qed
Lemma 4.8.2. Suppose \( A, Y, Z \) are independent spaces and \( W \subseteq A \oplus Y \oplus Z \). Then there is a basis for \( W \) of the following form \( \bigcup_{s=1}^{7} A_s \) where

\[
A_1 = \{ a_i + y_j + z_k \mid i \in \Sigma_1, \ j \in \Phi_1, \ k \in \Psi_1 \}
A_2 = \{ a_i + y_j \mid i \in \Sigma_2, \ j \in \Phi_2 \}
A_3 = \{ a_i + z_k \mid i \in \Sigma_3, \ k \in \Psi_3 \}
A_4 = \{ a_i + \sigma \mid i \in \Sigma_4 \}
A_5 = \{ y_j + z_k \mid j \in \Phi_5, \ k \in \Psi_5 \}
A_6 = \{ y_j \mid j \in \Phi_6 \}
A_7 = \{ z_k \mid k \in \Psi_7 \}
\]

and we have

- \( \{ a_i \mid i \in \Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4 \} \) are linearly independent vectors in \( A \).
- \( \{ y_j \mid j \in \Phi_1 \cup \Phi_2 \cup \Phi_5 \cup \Phi_6 \} \) are linearly independent vectors in \( Y \).
- \( \{ z_k \mid k \in \Psi_3 \cup \Psi_5 \cup \Psi_7 \} \) are linearly independent vectors in \( Z \).
- The \( \sigma \) are arbitrary linear forms in \( \{ y_j \mid j \in \Phi_1 \cup \Phi_5 \} \), not necessarily all same.

Proof. We start choosing a basis for \( W \subseteq A \oplus Y \oplus Z \), picking one vector at a time, and adding it to \( S = \bigcup_{s=1}^{7} A_s \) as below. We note that we do not add vectors to \( A_4 \) yet (this will be done after the process below ends):

**Initialize** \( A_1 = A_2 = A_3 = A_5 = A_6 = A_7 = \phi \). \( \quad S = \bigcup_{s=1}^{7} A_s \).

**Initialize** \( \Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4 = \phi \). \( \Phi = \Phi_1 \cup \Phi_2 \cup \Phi_5 \cup \Phi_6 = \phi \). \( \Psi = \Psi_1 \cup \Psi_2 \cup \Psi_5 \cup \Psi_7 = \phi \).

**Initialize** \( i^* = j^* = k^* = 1 \).

Repeat as long as possible:

- Pick a vector \( w \in W \), if possible, that fits any of the six cases below.
- If \( w = a + y + z \) where \( a \not\in \operatorname{Span}\{a_i\mid i \in \Sigma\} \), \( y \not\in \operatorname{Span}\{y_j\mid j \in \Phi\} \), \( z \not\in \operatorname{Span}\{z_k\mid k \in \Psi\} \), then
  - Let \( a_{i^*} = a \), \( y_{j^*} = y \), \( z_{k^*} = z \).
  - Add \( w = a_{i^*} + y_{j^*} + z_{k^*} \) to \( A_1 \) as well as \( S \).
  - add \( i^* \) to \( \Sigma \) and \( \Sigma_1 \), \( j^* \) to \( \Phi \) and \( \Phi_1 \), \( k^* \) to \( \Psi \) and \( \Psi_1 \).
  - Increment \( i^*, j^*, k^* \) each.
- \( \ldots \) 5 more similar cases \( \ldots \)

We hope that the process is clear to the reader. The sets of vectors and indices grow as the process continues. The six cases correspond to the six types of \( w : a + y + z, a + y, a + z, y + z, y, z \), which are added to \( A_1, A_2, A_3, A_5, A_6, A_7 \) respectively. In each case, we pick the vector \( w \) only if each of its components is linearly independent of vectors of the same “kind” that have already been “used” before (i.e. those indexed in \( \Sigma, \Phi, \Psi \) respectively). The indices \( i^*, j^*, k^* \) are the next available indices. The sets \( \Sigma, \Phi, \Psi \) maintain all the indices used so far (of the three kinds respectively).

We assume henceforth that the process above has ended. Let \( \operatorname{Span}(S) \subseteq W \) be the span of all the vectors chosen so far. A small modification of the process above ensures that \( (W \cap (Y \oplus Z)) \subseteq \operatorname{Span}(S) \). This is simply by considering the vectors in \( W \) in the order

\[
W \cap Y, \quad W \cap Z, \quad W \cap (Y \oplus Z), \quad \text{rest},
\]
and using Lemma 4.8.1. Hence we may assume henceforth that \((W \cap (Y \oplus Z)) \subseteq \text{Span}(S)\).

We now finish the argument by completing the basis for \(W\) and showing that every vector remaining in \(W \setminus \text{Span}(S)\) is of \(A_1\)-type (possibly after adding a vector in \(\text{Span}(S)\)). Indeed let \(w = a + y + z\) be any “remaining” vector in \(W \setminus \text{Span}(S)\). We observe that:

- It must be that \(a \notin \text{Span}\{a_i | i \in \Sigma\}\). This is because, otherwise we can cancel out \(a\) by adding back appropriate vectors in \(S\). This would result in a vector in \(W \setminus (Y \oplus Z) \subseteq \text{Span}(S)\), a contradiction.
- It must be that \(y \in \text{Span}\{y_j | j \in \Phi\}\) as well as \(z \in \text{Span}\{z_k | k \in \Phi\}\). This is because, otherwise we can keep the one (or both) for which this condition fails and cancel out the other (if any) by adding back appropriate vectors in \(S\). This would result in a vector of the type \(a + y + z\) or \(a + y\) or \(a + z\), contradicting the end of the above process.
- Finally, we can cancel out \(z\) as well as “part of \(y\) that occurs in \(\{y_j | j \in \Phi_2 \cup \Phi_0\}\)” by adding back appropriate vectors in \(S\).

\(\square\)

**Lemma 4.8.3.** Suppose \(P, Q\) are independent spaces, \(\dim(P) = \dim(Q) = n\), and \(W \subseteq P \oplus Q\). Suppose moreover that \(p_1, \ldots, p_n\) and \(q_1, \ldots, q_n\) are given as bases of \(P, Q\) respectively. Then there is an \(n \times n\) invertible matrix \(M\) such that after a change of basis (reusing the names)

\[
(p_1, \ldots, p_n) \leftarrow M(p_1, \ldots, p_n), \quad (q_1, \ldots, q_n) \leftarrow M(q_1, \ldots, q_n),
\]

there is a partition of the index set \(\{1, \ldots, n\} = \Delta_0 \cup \Delta_1 \cup \Delta_2 \cup \ldots \cup \Delta_m \cup \Omega_1 \cup \Omega_2\) and a basis for \(W\) of the form:

\[
\{p_i + \sigma(q_i) | i \in \Delta_1\} \cup B_2 \cup \ldots \cup B_m \cup C_1 \cup C_2,
\]

\[
B_k = \left\{q_j + \sigma(p_{\Delta_{[k+1:m]}}) | j \in \Delta_k\right\}
\]

where \(\Delta_{[k+1:m]} = \Delta_{k+1} \cup \ldots \cup \Delta_m \cup \Omega_1 \cup \Omega_2\),

\[
C_1 = \{p_i + \sigma(q_{\Omega_1}) | i \in \Omega_1 \cup \Omega_2\},
\]

\[
C_2 = \{q_j | j \in \Omega_2\}.
\]

We recall that \(\sigma(\cdot)\) are arbitrary linear forms in respective variables (not necessarily the same).

**Proof.** The proof is iterative. Let \(W_P, W_Q\) denote the projections of \(W\) onto \(P\) and \(Q\) respectively, i.e.

\[
W_P = \{p \in P | \exists q \in Q, p + q \in W\},
\]

\[
W_Q = \{q \in Q | \exists p \in P, p + q \in W\}.
\]

It is clearly possible to choose matched bases \((p_1, \ldots, p_d, p_{d+1}, \ldots, p_n)\) and \((q_1, \ldots, q_d, q_{d+1}, \ldots, q_n)\) for \(P\) and \(Q\) respectively such that

\[
W_P = \text{Span of} \quad p_1, \ldots, p_s, \quad p_{t+1}, \ldots, p_d
\]

\[
W_Q = \text{Span of} \quad q_1, \ldots, q_s, \quad q_{s+1}, \ldots, q_t
\]

The “unused” indices \(\{d + 1, \ldots, n\}\) are placed in \(\Delta_0\). We choose arbitrary linear forms \(\sigma_i(q)\) so that

\[
p_{t+1} + \sigma_{t+1}(q), \ldots, p_d + \sigma_d(q) \in W.
\]

142
These vectors are added to a partial basis for $W$ and the indices $\{t + 1, \ldots, d\}$ are added to $\Delta_1$. Letting $W' = W \cap \text{Span}(p_1, \ldots, p_t, q_1, \ldots, q_t)$, clearly

$$W = W' \oplus \text{Span}(p_{t+1} + \sigma_{t+1}(q), \ldots, p_d + \sigma_d(q)).$$

Since the latter are already added to a partial basis, we only need to find a further basis for $W'$. Moreover, our index-space is now reduced to $\{1, \ldots, t\}$ and we can continue iteratively. This process makes progress unless $p_{t+1}, \ldots, p_d$ are absent, i.e. if

$$W_p = \text{Span of } p_1, \ldots, p_s, \quad W_Q = \text{Span of } q_1, \ldots, q_s, \quad q_{s+1}, \ldots, q_t.$$

In this case, the iterative process is stopped, and we begin a new iterative process. We set $m = 2$ and choose

$$q_{s+1} + \sigma(p_{1, \ldots, s}), \ldots, q_t + \sigma(p_{1, \ldots, s}) \in W.$$

These vectors are added to a partial basis for $W$ letting $\Delta_m = \{s + 1, \ldots, t\}$. We increment $m$ by one and iterate the process on $W' = W \cap \text{Span}(p_1, \ldots, p_s, q_1, \ldots, q_s)$. We note that $W'_Q = \text{Span}(q_1, \ldots, q_s)$ while $W'_P \subseteq \text{Span}(p_1, \ldots, p_s)$ (it may shrink to a proper subspace). After appropriate change of matched basis, $W'_P = \text{Span}(p_1, \ldots, p_r)$ for $r \leq s$. This process makes progress unless we have

$$W_P = \text{Span of } p_1, \ldots, p_s, \quad W_Q = \text{Span of } q_1, \ldots, q_s.$$

At this point, a basis for $W$ is completed by first taking elements

$$p_1 + \sigma(q_{1, \ldots, s}), \ldots, p_s + \sigma(q_{1, \ldots, s}) \in W,$$

and adding to it $q_{r+1}, \ldots, q_s \in W \cap Q$. We can eliminate dependency of the former on the latter by elimination. The proof is completed by setting $\Omega_1 = \{1, \ldots, r\}$ and $\Omega_2 = \{r + 1, \ldots, s\}$.

Lemma 4.8.4. Let $X_1, \ldots, X_n$ be uniformly and independently distributed variables over $\{0, 1\}^k$. Let

$$\lambda_i \{ Y_{ij} \mid 1 \leq j \leq s_i \}, \quad 1 \leq i \leq m,$$

be real-valued functions of its arguments where:

- Each $Y_{ij} = X_r$ for some $r \in \{1, \ldots, n\}$.
- In the collection $\{ Y_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq s_i \}$, each $X_r$ appears exactly twice, as $Y_{i'j'}$ and $Y_{i''j''}$ for $i' \neq i''$.
- It (therefore) holds that $\sum_{i=1}^m s_i = 2n$.

Then

$$\mathbb{E}_{X_1, \ldots, X_n} \left[ \prod_{i=1}^m |\lambda_i(Y_{i1}, \ldots, Y_{is_i})| \right] \leq \prod_{i=1}^m \sqrt{\mathbb{E}_{Y_{i1, \ldots, Y_{is_i}}} \left[ \lambda_i^2(Y_{i1}, \ldots, Y_{is_i}) \right]}.$$

Proof. By induction. For $n = 1$, the only scenario and its proof is Cauchy-Schwartz:

$$\mathbb{E}_{X_1} \left[ |\lambda_1(X_1)\lambda_2(X_1)| \right] \leq \sqrt{\mathbb{E}_{X_1} \left[ \lambda_1^2(X_1) \right]} \sqrt{\mathbb{E}_{X_1} \left[ \lambda_2^2(X_1) \right]}.$$
Otherwise, we assume w.l.o.g. that \( n \geq 2 \) and \( Y_{11} = Y_{21} = X_n \). Applying Cauchy-Schwartz on \( X_n \),

\[
\mathbb{E}_{X_1,\ldots,X_n} \left[ \prod_{i=1}^{m} |\lambda_i(Y_{i1}, \ldots, Y_{is_i})| \right] \leq \mathbb{E}_{X_1,\ldots,X_n} \left[ \sqrt{\mathbb{E}_{X_n} [\lambda_1^2(X_n, \cdot)]} \sqrt{\mathbb{E}_{X_n} [\lambda_2^2(X_n, \cdot)]} \prod_{i=3}^{m} |\lambda_i(Y_{i1}, \ldots, Y_{is_i})| \right].
\]

We get a further desired upper bound by induction hypothesis applied to functions \( \lambda'_1, \lambda'_2, \lambda_3, \ldots, \lambda_m \) where

\[
\lambda'_1(Y_{12}, \ldots, Y_{1s_1}) = \mathbb{E}_{X_n} [\lambda_1^2(X_n, Y_{12}, \ldots, Y_{1s_1})], \quad \lambda'_2(Y_{22}, \ldots, Y_{2s_2}) = \mathbb{E}_{X_n} [\lambda_2^2(X_n, Y_{12}, \ldots, Y_{1s_1})].
\]

\[\square\]

**Lemma 4.8.5.** Suppose \( x \in \{0,1\}^k \) is a uniformly distributed input and \( A, B \subseteq \{1, \ldots, k\} \) such that \( A \cup B = \{1, \ldots, k\} \). Let \( x_A, x_B \) denote the restricted input to \( A \) and \( B \) respectively. Suppose \( \lambda, \psi \) are non-negative functions of \( x_A \) and \( x_B \) respectively. Then

\[
\mathbb{E}_x [\lambda(x_A)\psi(x_B)] \leq \left( \max_{x_{A \cap B}} \mathbb{E}_{x_A \setminus B} [\lambda(x_A)] \right) \cdot \mathbb{E}_{x_B} [\psi(x_B)].
\]

**Proof.** This is self-evident. Suppose \( \beta \) is the maximum above. Then

\[
\mathbb{E}_x [\lambda(x_A)\psi(x_B)] = \mathbb{E}_{x_{A \cap B}} \left[ \mathbb{E}_{x_A \setminus B} [\lambda(x_A)] \cdot \mathbb{E}_{x_B \setminus A} [\psi(x_B)] \right] \leq \beta \cdot \mathbb{E}_{x_B \setminus A} \left[ \mathbb{E}_{x_{B \setminus A}} [\psi(x_B)] \right] = \beta \cdot \mathbb{E}_{x_B} [\psi(x_B)].
\]

\[\square\]
Chapter 5

The Biased Long Code and Hardness of Vertex-Cover

In this chapter we show a reduction from the 2-to-2 Games problem to the Independent Set problem, proving Theorem 1.2.7. The reduction is along the lines of [DS05], using the Biased Long Code and analytic theorems of Russo, Margulis and Friedgut, introduced in [DS05].

5.1 Biased Long Code

While the Biased Long Code can be viewed as an encoding scheme, it is more convenient to take a combinatorial view and treat it as a weighted Kneser graph. Valid codewords correspond to certain large (in fact the largest) independent sets in this graph.

Definition 5.1.1. For a bias parameter $p \in (0, 1)$ and alphabet $\Sigma$, the vertex set of weighted Kneser graph $G_p[\Sigma]$ is $\mathcal{P}(\Sigma)$, the family of all subsets of $\Sigma$. The weight of a vertex $A \subseteq \Sigma$ is $\mu_p(A) = p^{|A|}(1 - p)^{|\Sigma| - |A|}$. The edge set is $\{ (A, B) \mid A, B \subseteq \Sigma, A \cap B = \emptyset \}$.

For a family $F \subseteq \mathcal{P}(\Sigma)$, let $\mu_p(F)$ denote its weight under $\mu_p$.

5.1.1 Independent Sets in the Kneser Graph

Let $p \leq \frac{1}{2}$. An element $\sigma \in \Sigma$ can be encoded by $F_{\sigma}$, an independent set in the weighted Kneser graph defined by

$$F_{\sigma} = \{ A \subseteq \Sigma \mid \sigma \in A \}.$$

Such families are called dictatorships, since the membership of $A$ to the family $F_{\sigma}$ depends only on a single element. Since the probability a random vertex $A$ in the weighted Kneser graph is $p$, we have $\mu_p(F_{\sigma}) = p$, and these are actually the largest independent sets in $G_p[\Sigma]$. We begin with an elegant proof in a special case:

Claim 5.1.2. Suppose $p = \frac{1}{m}$ for $m \in \mathbb{N}$, and let $F \subseteq \mathcal{P}(\Sigma)$ be an independent set in $G_p[\Sigma]$. Then $\mu_p(F) \leq p$.

Proof. Assume by way of contradiction that $\mu_p(F) > p$. Sample a partition $A_1, \ldots, A_m$ of $\Sigma$ by the following process: for each $\sigma \in \Sigma$, choose a single $i \in [m]$ and put $\sigma$ in $A_i$. Observe that $A_1, \ldots, A_m$ are
pairwise disjoint and therefore form a clique, and marginally each $A_i$ is distributed according to $\mu_p$, and thus $\Pr[A_i \in \mathcal{F}] = \mu_p(\mathcal{F}) > p$. Therefore, by linearity of expectation,

$$E \left[ \sum_{i=1}^{m} 1_{A_i \in \mathcal{F}} \right] = \sum_{i=1}^{m} E[1_{A_i \in \mathcal{F}}] > \sum_{i=1}^{m} p = mp \geq 1,$$

and in particular there are random choices in the process such that at least two of the sets $A_1, \ldots, A_m$ are in $\mathcal{F}$, and contradiction. \hfill \Box

The above lemma holds for any $p \leq \frac{1}{2}$ and is a corollary of Hoffman’s eigenvalue bound on the size of an independent set. Below, after making a few definitions, we give an alternative proof.

**Definition 5.1.3.** Let $A \sim \mu_p$ denote the process of picking a set $A \subseteq \Sigma$ according to the distribution $\mu_p$. For a fixed element $\sigma \in \Sigma$, let $\text{Infl}_\sigma(\mathcal{F})$ denote its influence on the family $\mathcal{F}$ defined as

$$\text{Infl}_\sigma(\mathcal{F}) = \Pr_{A \sim \mu_p} [\text{Exactly one of the pair } A \text{ and } A \Delta \{\sigma\} \text{ is in } \mathcal{F}].$$

The average sensitivity of a family $as_p(\mathcal{F})$ is the sum of all influences, i.e.

$$as_p(\mathcal{F}) = \sum_{\sigma \in \Sigma} \text{Infl}_\sigma(\mathcal{F}).$$

**Fact 5.1.4 (Poincaré Inequality).** For a set family $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$, $as_p(\mathcal{F}) \geq \frac{1}{p(1-p)}(\mu_p(\mathcal{F}) - \mu_p(\mathcal{F})^2).$ \hfill \blacksquare

A family $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$ is called monotone if: $A \in \mathcal{F}, A \subseteq B$ implies that $B \in \mathcal{F}$. The following fundamental lemma relates the average sensitivity of $\mathcal{F}$ to the rate in which its size increase with $p$.

**Theorem 5.1.5 (Russo - Margulis [Rus82, Mar74]).** Suppose $\mathcal{F}$ is a monotone family. Then $\mu_q(\mathcal{F})$ is an increasing function of $q$ and

$$\frac{d\mu_q(\mathcal{F})}{dq} = as_q(\mathcal{F}).$$

In particular, if $\mathcal{F}$ is monotone and $q \geq p$ then $\mu_q(\mathcal{F}) \geq \mu_p(\mathcal{F})$ (as the average sensitivity, and thus the derivative, is non-negative).

**Claim 5.1.6.** Suppose $p \leq \frac{1}{2}$ and let $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$ be an independent set in $G_p[\Sigma]$. Then $\mu_p(\mathcal{F}) \leq p$.

**Proof.** We may assume without loss of generality that $\mathcal{F}$ is monotone: otherwise we may consider

$$\mathcal{F}' = \{ B \mid \exists A \in \mathcal{F}, A \subseteq B \}$$

that has larger weight than $\mathcal{F}$, is monotone and an independent set.

Assume by way of contradiction that $\mu_p(\mathcal{F}) > p$, and let

$$q = \inf \{ a \geq p \mid \mu_a(\mathcal{F}) > a \}.$$

Since $q$ is the infimum, we have $\mu_q(\mathcal{F}) = q$; we next show that $\mu_q(\mathcal{F}) > q$, a contradiction that finishes the proof.

Clearly, $q > p$ and note that $q \leq \frac{1}{2}$; otherwise, we would have that $\mu_{1/2}(\mathcal{F}) > \frac{1}{2}$ in contradiction to Claim 5.1.2. Also, note that we have $\mu_a(\mathcal{F}) \leq \frac{1}{2}$ for all $a < q$; otherwise, we would have $\mu_{1/2}(\mathcal{F}) > \mu_a(\mathcal{F}) > \frac{1}{2}$, and again contradiction to Claim 5.1.2.

146
By Theorem 5.1.5, for every \(a \in [p, q]\) we have
\[
\frac{d\mu_a(F)}{da}(a) \geq as_q(F) \geq \frac{1}{a(1-a)}(\mu_a(F) - \mu_a(F)^2) \geq 1,
\]
where the second inequality is by Fact 5.1.4 and the third inequality is since \(a \leq \mu_a(F) \leq \frac{1}{2}\). Therefore,
\[
\mu_q(F) \geq \mu_p(F) + 1 \cdot (q-p) > p + (q-p) = q.
\]

With a bit more effort, one can show that the only independent set of weight \(p\) are dictatorships. A natural question is whether these are the only large independent sets in \(G_p[\Sigma]\), i.e. to prove structural results on independent sets in \(G_p[\Sigma]\) that have non-negligible weight.

At the heart of the soundness analysis (following Dinur and Safra [DS05] and ignoring some subtleties) lies a similar question, and it is shown that any such set is close to a junta (a family that depends only on a few elements), which is sufficient for the purposes of the reduction herein. We remark that this question has been further studied in [DFR08, DMR09, FR18].

### 5.2 The Reduction

Let \(G = (V, E, \Phi, \Sigma)\) be the instance of a 2-to-2 Game as in Theorem 1.2.6. The parameter \(\delta\) therein will be chosen later. The Independent Set instance \(G' = (V', E')\) is defined as follows. Set the parameter \(p = 1 - \frac{1}{\sqrt{2}} - \delta\). The vertex set of the instance is
\[
V' = \{ (x, A) \mid x \in V, A \subseteq \Sigma \}.
\]
The weight of the vertex \((x, A)\) is \(\frac{1}{|V|} \cdot \mu_p(A)\), so that the total weight of all the vertices is 1. The edge set is
\[
E' = \{ (\{(x_1, A_1), (x_2, A_2)\}) \mid (x_1, x_2) \in E \wedge \forall \sigma_1 \in A_1, \sigma_2 \in A_2, (\sigma_1, \sigma_2) \not\in \Phi(x_1, x_2) \}.
\]
In words, there is a cloud of vertices for every \(x \in V\). For every constraint \((x_1, x_2) \in E\), there are cross edges between the respective clouds. There is an edge between \((x_1, A_1)\) and \((x_2, A_2)\) if there is no pair of colors in the sets \(A_1, A_2\) that satisfy the constraint on \((x_1, x_2)\).

#### 5.2.1 Completeness

Let \(C : X \rightarrow \Sigma\) be a \((1, 1 - \delta)\)-coloring of the game \(G = (V, E, \Phi, \Sigma)\) where \(X \subseteq V\), \(|X| = (1 - \delta)|V|\). The coloring satisfies all the constraints inside \(X\). Consider the set of vertices in \(G'(V', E')\),
\[
I \overset{\text{def}}{=} \{ (x, A) \mid x \in X, C(x) \in A \}.
\]

Clearly, the set \(I\) includes a weight \(p\) of the vertices inside the cloud for every \(x \in X\). Hence the weight of \(I\) is \((1 - \delta)p \geq 1 - \frac{1}{\sqrt{2}} - 2\delta\). We argue that \(I\) is an independent set. For every pair of vertices \((x, A), (x', A') \in I\), we show that there is no edge between them in \(G'\). Since the coloring \(C\) satisfies the constraint \((x, x')\), we have \((C(x), C(x')) \in \Phi(x, x')\). By definition of the set \(I\), we have \(C(x) \in A, C(x') \in A'\). Thus \(A, A'\) contain a consistent pair of colors, so there is no edge between \((x, A)\) and \((x', A')\).

---

\(^{21}\) One could add edges inside each cloud according to the Kneser graph. The reduction does not require it though.
5.2.2 Soundness

We begin by stating two auxiliary lemmas towards the soundness analysis. The relevance of the 2-to-2-ness of the constraints and the choice of \( p \approx 1 - \frac{1}{\sqrt{2}} \) is apparent from the statements of these lemmas. Let \( \Sigma \) and \( \Gamma \) be alphabets such that \( |\Gamma| = \frac{|\Sigma|}{2} \) and \( \pi : \Sigma \to \Gamma \) be a 2-to-1 map. For \( F \subseteq \Sigma \), its projection \( \pi(F) \subseteq \Gamma \) is defined naturally as \( \{ \pi(\sigma) | \sigma \in F \} \). For a family \( F \subseteq \mathcal{P}(\Sigma) \), the projected family \( \pi(F) \subseteq \mathcal{P}(\Gamma) \) is defined naturally as \( \{ \pi(F) | F \in F \} \). For a subset \( H \subseteq \Gamma \), the set \( \pi^{-1}(H) \), \( |\pi^{-1}(H)| = 2|H| \) is defined naturally as \( \{ \sigma | \sigma \in \Sigma, \pi(\sigma) \in H \} \).

**Lemma 5.2.1.** For \( q \in (0, 1) \), \( \mu_1(\pi(F)) \geq \mu_q(F) \).

**Proof.** For every \( H \subseteq \Gamma \), we define \( \pi^+(H) = \{ F \subseteq \Sigma | \pi(F) = H \} \). We observe that

- \( \mu_1(F) = \mu_q(\pi^+(H)) \).
- The families \( \pi^+(H) \) over all \( H \subseteq \Gamma \) is a disjoint partition of the family \( \mathcal{P}(\Sigma) \).

The lemma follows by noting that

\[
\mu_1(F) = \sum_{H \in \pi(F)} \mu_1(F) = \sum_{H \in \pi(F)} \mu_q(\pi^+(H)) = \sum_{H \in \pi(F)} \mu_q(\pi^+(H) \cap F) = \mu_q(F).
\]

**Lemma 5.2.2.** Let \( F \subseteq \mathcal{P}(\Sigma) \), \( F' \subseteq \mathcal{P}(\Sigma') \) be two families, each of weight strictly larger than \( \frac{1}{2} \) under the distribution \( \mu_q \) with \( q = 1 - \frac{1}{\sqrt{2}} \). Let \( \pi : \Sigma \to \Gamma \), \( \pi' : \Sigma' \to \Gamma \) be 2-to-1 maps (so \( |\Sigma| = |\Sigma'| = 2|\Gamma| \)). Then there exist \( F \in F \), \( F' \in F' \) such that \( \pi(F) \cap \pi'(F') = \phi \).

**Proof.** We note that \( 1 - (1 - q)^2 = \frac{1}{2} \) and from Lemma 5.2.1, \( \mu_1(F) \geq \mu_q(F) > \frac{1}{2} \) and similarly \( \mu_1(F') > \frac{1}{2} \). Thus \( \pi(F) \) and \( \pi'(F') \) are families, each containing more than half (in the usual counting sense) of the sets from \( \mathcal{P}(\Gamma) \). Hence there must exist \( H \in \pi(F) \), \( H' \in \pi'(F') \) that are complements of each other and in particular \( H \cap H' = \phi \).

We now present the soundness analysis. Given a maximal independent set \( I \) of weight at least \( \varepsilon \) in \( G' \), we show how to construct an assignment for \( G = (V, E, \Phi, \Sigma) \) that satisfies at least \( C(\varepsilon) > 0 \) constraints. For every \( x \in G \), consider the part of \( I \) inside the cloud of \( x \),

\[
F_x = \{ A | A \subseteq \Sigma, (x, A) \in I \}.
\]

**Claim 5.2.3.** The family \( F_x \) is monotone.

**Proof.** Otherwise, there are \( A \subseteq B \) such that \( A \in F_x \) and \( B \notin F_x \). Then \( I \cup \{(x, B)\} \) is an independent set larger than \( I \), contradicting the maximality of \( I \).

Since the independent set \( I \) has weight \( \varepsilon \), by an averaging argument, there is a set \( X \subseteq V \), \( |X| \geq \frac{\varepsilon}{2} \cdot |V| \) such that \( I \) includes a weight \( \geq \frac{\varepsilon}{2} \) of vertices from the cloud of \( x \), i.e. \( \mu_p(F_x) \geq \frac{\varepsilon}{2} \) for \( x \in X \).

**Claim 5.2.4.** There exists \( p' \in (p, p + \delta) \) such that

\[
\mathbb{E}_{x \in X} [\text{as}_{p'}(F_x)] \leq \frac{1}{\delta}.
\]
Proof. By Russo-Margulis 5.1.5 and Lagrange’s mean value theorem, it follows that there exists \( p' \in (p, p + \delta) \) such that

\[
\mathbb{E}_{x \in X} \left[ \text{as}_{p'}(F_x) \right] = \frac{d}{dq} \left( \mathbb{E}_{x \in X} \left[ \mu_q(F_x) \right] \right) \bigg|_{q = p'} = \mathbb{E}_{x \in X} \left[ \mu_{p + \delta}(F_x) \right] - \mathbb{E}_{x \in X} \left[ \mu_{p}(F_x) \right] \leq \frac{d}{\delta}.
\]

From Claim 5.2.4 and an averaging argument, there is a set \( X' \subseteq X \) such that \( |X'| \geq \frac{|X|}{2} \) and Friedgut’s Theorem states that families with bounded average sensitivity are well-approximated by “juntas”.

**Definition 5.2.5 (Junta).** A family \( F \subseteq \mathcal{P}(\Sigma) \) is called a \( j \)-junta, if there exists \( J \subseteq \Sigma, |J| = j \) such that the membership of a set \( A \) in \( F \) is determined by only \( A \cap J \).

**Theorem 5.2.6 (Friedgut [Fri98]).** There exists a constant \( C(q) \geq 1 \) such that for every \( F \subseteq \mathcal{P}(\Sigma) \) and an accuracy parameter \( \eta > 0 \), there exists \( F' \subseteq \mathcal{P}(\Sigma) \) that is a \( j \)-junta and

- \( j = C(q) \text{as}_q(F)/\eta \).
- \( \mu_q(F \Delta F') \leq \eta \).

Fix \( x \in X' \) and set \( \eta \overset{\text{def}}{=} \frac{\varepsilon}{2n} \). Since \( \text{as}_{p'}(F_x) \leq \frac{2}{\delta} \), it follows from Friedgut’s Theorem that \( F_x \) is \( \eta \)-close to a \( k \)-junta with \( k = C(p')^2/(\delta n) \). Let \( J_x \subseteq \Sigma \) denote the set of elements on which the junta depends. Clearly, the set-family that is a junta on \( J_x \) and is closest to \( F_x \) is the “majority vote” on each setting of \( J_x \), namely

\[
[F_x]_{\frac{1}{2}} \overset{\text{def}}{=} \left\{ F \cup F' \mid F \subseteq \Sigma \setminus J_x, F' \subseteq J_x, \sum_{A \subseteq \Sigma \setminus J_x, \sum_{A \sim \mu_{p'}} \Pr_{A} [A \cup F' \in F_x] > \frac{1}{2} \right\}.
\]

The following claim shows that the family \([F_x]_{\frac{1}{2}} \) is also close to \( F_x \) (and will be more useful to work with):

\[
[F_x]_{\frac{3}{4}} \overset{\text{def}}{=} \left\{ F \cup F' \mid F \subseteq \Sigma \setminus J_x, F' \subseteq J_x, \sum_{A \subseteq \Sigma \setminus J_x, \sum_{A \sim \mu_{p'}} \Pr_{A} [A \cup F' \in F_x] > \frac{3}{4} \right\}.
\]

**Claim 5.2.7.** \( \mu_{p'}(F_x \Delta [F_x]_{\frac{3}{4}}) \leq 5\eta \).

**Proof.** Let \( F^* \subseteq \mathcal{P}(J_x) \) be the family of subsets \( F \subseteq J_x \) such that

\[
\frac{1}{2} < \Pr_{A \subseteq \Sigma \setminus J_x, \sum_{A \sim \mu_{p'}} [A \cup F \in F_x]} \leq \frac{3}{4}.
\]

Note that for each such \( F \): (a) at least \( \frac{1}{4} \) (weighted) fraction of its extensions to \( \Sigma \) are not in \( F_x \) (b) each extension is in \([F_x]_{\frac{1}{2}} \) (c) no extension is in \([F_x]_{\frac{3}{4}} \). Hence

\[
\frac{1}{4} \cdot \sum_{F \subseteq J_x, F \sim \mu_{p'}} [F \in F^*] \leq \mu_{p'}(F_x \Delta [F_x]_{\frac{1}{2}}) \leq \eta.
\]

It follows that

\[
\mu([F_x]_{\frac{1}{2}} \Delta [F_x]_{\frac{3}{4}}) = \Pr_{F \subseteq J_x, F \sim \mu_{p'}} [F \in F^*] \leq 4\eta.
\]

The proof is concluded by the triangle inequality,

\[
\mu_{p'}(F_x \Delta [F_x]_{\frac{1}{2}}) \leq \mu_{p'}(F_x \Delta [F_x]_{\frac{3}{4}}) + \mu_{p'}([F_x]_{\frac{1}{2}} \Delta [F_x]_{\frac{3}{4}}) \leq 5\eta.
\]

\[
\square
\]
Claim 5.2.8. \([F_x]_{\frac{3}{4}} \neq \phi\).

Proof. Using the triangle inequality, the previous claim, and that \(\eta = \frac{\varepsilon}{20}\),

\[
\mu_{p'} \left( [F_x]_{\frac{3}{4}} \right) \geq \mu_{p'}(F_x) - \mu_{p'}(F_x \Delta [F_x]_{\frac{3}{4}}) \geq \frac{\varepsilon}{2} - 5\eta > 0.
\]

Definition 5.2.9. The extended junta \(EJ(x)\) of \(x\) is defined by (recall that \(k = C(p')^{2/(\delta n)}\) is an upper bound on the size of the junta)

\[
EJ(x) = J_x \cup \left\{ \sigma \in \Sigma \mid \text{Infl}_\sigma(F_x) \geq 2^{-10k} \right\}.
\]

We note that for \(x \in X'\), as \(p'(F_x) \leq \frac{2}{5}\) and since the average sensitivity is the sum of all influences, \(|EJ(x)| \leq j = k + \frac{2^{10k}}{5}\). Our coloring to the game \(G = (V, E, \Phi, \Sigma)\) will assign, to every \(x \in X'\), a random element from \(EJ(x)\). We show that expectedly, this coloring satisfies \(\delta(\varepsilon)\) fraction of the constraints in \(G\). First, \(|X'| \geq |X| \geq \frac{\varepsilon}{4}|V|\) and thus \(X'\) contains \(\Omega(\varepsilon^2)\) fraction of the edges of \(G\). Secondly, we show that every constraint \((x_1, x_2)\) inside \(X'\) is satisfied with probability \(\geq \frac{1}{j}\). Combining the these two claims, we conclude that there is an assignment satisfying at least \(\Omega(\varepsilon^2) = C(\varepsilon) > 0\) fraction of the constraints of \(G\). Thus, if we were to choose \(\delta < C(\varepsilon)\) in Theorem 1.2.6, we would conclude that \(G'\) has no Independent Set of weight \(\varepsilon\).

Lemma 5.2.10. Suppose \(x_1, x_2 \in X'\) are connected by an edge. Then there exist consistent colors for \(x_1, x_2\) in their respective extended juntas. I.e. there exist \(\sigma_1 \in EJ(x_1)\), \(\sigma_2 \in EJ(x_2)\) such that \((\sigma_1, \sigma_2) \in \Phi(x_1, x_2)\).

Proof. It will be convenient to think of the 2-to-2 constraint in terms of a pair of 2-to-1 maps \(\pi_1 : \Sigma_1 \rightarrow \Gamma, \pi_2 : \Sigma_2 \rightarrow \Gamma\). Here \(\Sigma_1 = \Sigma_2 = \Sigma\) are the same alphabet, but it will be convenient to think of them as separate. A coloring \((a_1, a_2)\) to vertices \((x_1, x_2)\) satisfies the 2-to-2 constraint if and only if \(\pi_1(a_1) = \pi_2(a_2)\). Assume by way of contradiction that there is no pair of consistent colors in the extended juntas for \(x_1\) and \(x_2\). The assumption can be stated as

\[
\pi_1(EJ(x_1)) \cap \pi_2(EJ(x_2)) = \phi.
\]

Note in particular that \(J_{x_2} \subseteq EJ(x_2)\) and hence

\[
\pi_1(EJ(x_1)) \cap \pi_2(J_{x_2}) = \phi. \tag{5.1}
\]

Our goal is to prove that there are \(F_1 \in F_{x_1}\), \(F_2 \in F_{x_2}\) such that \(\pi_1(F_1) \cap \pi_2(F_2) = \phi\). We will zero-in on such \(F_1, F_2\) in progressive manner. We consider the case of \(F_1\), the other case being similar. We zero-in on a sequence of sets

\[
A_1 \subseteq B_1 \subseteq B_1 \subseteq F_1,
\]

that are contained, respectively, in progressively expanding “universe of focus”

\[
J_{x_1} \subseteq \pi^{-1}_1(\pi_1(J_{x_1})) \subseteq \pi^{-1}_1(\pi_1(J_{x_1})) \cup \pi^{-1}_1(\pi_2(J_{x_2})) \subseteq \Sigma_1.
\]

We clarify that the set \(\pi^{-1}_1(\pi_1(J_{x_1}))\) is a superset of \(J_{x_1}\), and can have size up to \(2|J_{x_1}|\) since that map \(\pi_1\) is 2-to-1. The weights (sizes) of set-families are with respect to \(\mu_{p'}\), unless stated otherwise.
Finally, letting $F$ so that we are almost done. Denote such that at least $\frac{3}{4}$ of its extensions outside $J_{x_1}$ are in $F_{x_1}$.

- We let $B_1 = A_1 \cup (\pi_1^{-1}(\pi_1(J_{x_1})) \setminus J_{x_1})$. Due to monotonicity of $F_{x_1}$, at least $\frac{3}{4}$ of the extensions of $B_1$ outside $\pi_1^{-1}(\pi_1(J_{x_1}))$ are in $F_{x_1}$.

- We now retain $B_1$ as is, but consider it as subset of enlarged universe $\pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2}))$. The elements added to the enlarged universe, namely $\pi_1^{-1}(\pi_2(J_{x_2}))$ are outside of $EJ(x_1)$ (using Equation (5.1)), hence have influence at most $2^{-10k}$, and are at most $2k$ in number. The fraction of extensions of $B_1$ outside $\pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2}))$ remains at least
\[
\frac{3}{4} - 2^{-10k} \cdot p'^{-2k-1}(1 - p')^{-2k-1} \geq \frac{5}{8}.
\]

Using a similar argument for $x_2$, to summarize, there exist
\[
B_1 \subseteq D_1 = \pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2})), \quad B_2 \subseteq D_2 = \pi_2^{-1}(\pi_1(J_{x_1})) \cup \pi_2^{-1}(\pi_2(J_{x_2}))
\]
such that at least $\frac{5}{8}$ of their extensions outside $D_1$ and $D_2$ are in $F_{x_1}$ and $F_{x_2}$ respectively. Note that
\[
\pi_1(B_1) \cap \pi_2(B_2) = \phi. \tag{5.2}
\]

We are almost done. Denote
\[
F_1 = \{ S_1 \subseteq \Sigma_1 \setminus D_1 \mid B_1 \cup S_1 \in F_{x_1} \}, \quad F_2 = \{ S_2 \subseteq \Sigma_2 \setminus D_2 \mid B_2 \cup S_2 \in F_{x_2} \},
\]
so that $\mu_{p'}(F_1) \geq \frac{5}{8}$ and due to monotonicity, letting $q = 1 - \frac{1}{\sqrt{2}} \geq p'$, $\mu_q(F_1) \geq \frac{5}{8}$, and similarly $\mu_q(F_2) \geq \frac{5}{8}$. Applying Lemma 5.2.2 to $F_1$, $F_2$ along with 2-to-1 maps $\pi_1 : \Sigma_1 \setminus D_1 \to \Gamma \setminus \pi_1(D_1)$ and $\pi_2 : \Sigma_2 \setminus D_2 \to \Gamma \setminus \pi_2(D_2)$ (we have $\pi_1(D_1) = \pi_2(D_2) = J_{x_1} \cup J_{x_2}$), there exist $F_1^* \subseteq \Sigma_1 \setminus D_1$, $F_2^* \subseteq \Sigma_2 \setminus D_2$ such that
\[
\pi_1(F_1^*) \cap \pi_2(F_2^*) = \phi. \tag{5.3}
\]
Finally, letting $F_1 = B_1 \cup F_1^*$ and $F_2 = B_2 \cup F_2^*$, and using Equations (5.2),(5.3), we conclude that $F_1 \in F_{x_1}$, $F_2 \in F_{x_2}$, $\pi_1(F_1) \cap \pi_2(F_2) = \phi$ as desired. \qed
Chapter 6

Conclusions and Open Problems

The main results of this thesis are:

1. Monotonicity Testing ([GGL+00]): Optimal (non-adaptive) testing algorithm that makes $\tilde{O}(\sqrt{n}\varepsilon^{-2})$ queries. We believe that we used to achieve this tester — a directed version of Talagrand’s Isoperimetric Inequality — may be of independent interest.

2. 2-to-2 Games ([Kho02]): NP-hardness of distinguishing between instances with value $> 1 - \varepsilon$, and instances whose value with value $\leq \varepsilon$ (for every $\varepsilon > 0$). Aside from the result itself and its application to hardness of approximation, we believe that the ideas used in the reduction, as well as the analytical machinery for the Grassmann graph developed in this thesis, may be of independent interest.

We conclude this thesis with several open problems.

**Problem 1.** Prove that there exists $c > \frac{1}{2}$, such that for every $\varepsilon > 0$, $\text{GapUG}[c, \varepsilon]$ is NP-hard on instances with alphabet size $q(\varepsilon) \in \mathbb{N}$.

**Problem 2** (2-to-2 Games with perfect completeness). Prove that for every $\varepsilon > 0$, $\text{Gap2-to-2-Games}[1, \varepsilon]$ is NP-hard on instances with alphabet size $q(\varepsilon) \in \mathbb{N}$.

**Problem 3.** Prove more hardness results based on the hardness of 2-to-2 Games.

For various problems such as Vertex-Cover and Max-Cut, the best known NP-hardness result is worse than the best known hardness result assuming the Unique-Games Conjecture. An interesting open problem is to prove an implication in the other direction: e.g. if a certain gap for Max-Cut or Vertex-Cover is NP-hard, then the Unique-Games Conjecture is true.

**Problem 4.** Show that if $\text{GapIS}[\frac{1}{2} - \varepsilon, \varepsilon]$ is NP-hard $\forall \varepsilon > 0$, then the Unique-Games Conjecture holds.

Currently, the best known adaptive algorithm for monotonicity testing is the tester given herein, and the best known lower bound due to [CWX17a] is $\tilde{\Omega}(n^{-1/3})$, leaving a polynomial gap.

**Problem 5.** Settle the query complexity of monotonicity testing for adaptive testers.

---

We remark that for the Max-Cut problem, this is a well known research direction related to optimal parallel repetition theorem.
Bibliography


ניתן להיווכח בעובדה שב_absolute השואול הקושי ששוורץ את מקודקודים בשפה 
והזוהר, או ששוורץ קורא ש.Void קסם של כל קודקודים בעומק 
המקודקודים בשפה.


The unique games conjecture (UGC) by [Bar17] states that if for some graph \( G \), there exists a \( 1-\epsilon \)-approximation algorithm for the Unique Games problem, then for all \( \alpha, \beta > 0 \), there exists an \( \alpha \)-approximation algorithm for the satisfiability problem \( SAT \) with a clause-to-variable ratio of \( \beta \).

For a graph \( G \), let \( \Phi_G(S) \) denote the number of edges between vertices in \( S \). Then, for any subset \( S \subseteq U \), we have

\[
\Phi_G(S) \leq \frac{|E(S, U \setminus S)|}{d|S|}
\]

where \( E(S, U \setminus S) \) denotes the edges between \( S \) and \( U \setminus S \).

This result implies that for any \( \epsilon > 0 \), there exists a \( (1-\epsilon) \)-approximation algorithm for the satisfiability problem if there is a \( \epsilon \)-approximation algorithm for the Unique Games problem. This is because any satisfying assignment to a satisfiable instance can be used to construct a unique game instance with the same solution, which can then be approximated to within \( \epsilon \) in polynomial time.

In particular, this result provides a reduction from the satisfiability problem to the unique games problem, which has important implications for the complexity of these problems. For example, if we could find a \( \epsilon \)-approximation algorithm for the unique games problem, then we could use this to obtain a \( (1-\epsilon) \)-approximation algorithm for the satisfiability problem, which would imply that \( P = NP \).
\[ G = (V, E, \Phi, \Sigma) \]

2-to-2 Games

\[
\begin{align*}
\text{Problem: 2-to-2 Games} & \quad \forall G = (V, E, \Phi, \Sigma),
\end{align*}
\]

\[ k \in \mathbb{N} \quad \delta > 0 \]

Theorem 2.2

\[ \delta - \epsilon \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} + \epsilon \]

\[ \frac{1}{2} + \frac{\epsilon}{\log(1/\epsilon)} \]

\[ \frac{1}{2} + \frac{\epsilon}{\log(1/\epsilon)} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} + \epsilon \]

\[ \frac{1}{2} + \frac{\epsilon}{\log(1/\epsilon)} \]

\[ \frac{1}{2} + \frac{\epsilon}{\log(1/\epsilon)} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} + \epsilon \]

\[ \frac{1}{2} + \frac{\epsilon}{\log(1/\epsilon)} \]

\[ \frac{1}{2} + \frac{\epsilon}{\log(1/\epsilon)} \]

\[ \frac{1}{2} \]

\[ \frac{1}{2} \]
NP-complete.

In the more general case, the Unique-Games Conjecture states that it is NP-hard to find a coloring of the graph that satisfies a fraction of the constraints, even if a fraction of the constraints are satisfied.

The Unique-Games Conjecture was first proposed by [Ko05] and later refined by [KMM02, K08, CK+06, DMR09]. The conjecture states that it is NP-hard to find a coloring of the graph that satisfies a fraction of the constraints, even if a fraction of the constraints are satisfied.

The Unique-Games Conjecture has been extensively studied and has led to a number of important results in theoretical computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

Another important result is the Unique-Games Conjecture's implication for the MAX-LIN problem. The MAX-LIN problem is a relaxation of the MAX-CUT problem, where the edges are replaced with linear constraints.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-LIN problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

The Unique-Games Conjecture has also been studied in the context of graph colorings and other problems in computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

Another important result is the Unique-Games Conjecture's implication for the MAX-LIN problem. The MAX-LIN problem is a relaxation of the MAX-CUT problem, where the edges are replaced with linear constraints.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-LIN problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

The Unique-Games Conjecture has also been studied in the context of graph colorings and other problems in computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

Another important result is the Unique-Games Conjecture's implication for the MAX-LIN problem. The MAX-LIN problem is a relaxation of the MAX-CUT problem, where the edges are replaced with linear constraints.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-LIN problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

The Unique-Games Conjecture has also been studied in the context of graph colorings and other problems in computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

Another important result is the Unique-Games Conjecture's implication for the MAX-LIN problem. The MAX-LIN problem is a relaxation of the MAX-CUT problem, where the edges are replaced with linear constraints.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-LIN problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

The Unique-Games Conjecture has also been studied in the context of graph colorings and other problems in computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

Another important result is the Unique-Games Conjecture's implication for the MAX-LIN problem. The MAX-LIN problem is a relaxation of the MAX-CUT problem, where the edges are replaced with linear constraints.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-LIN problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

The Unique-Games Conjecture has also been studied in the context of graph colorings and other problems in computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

Another important result is the Unique-Games Conjecture's implication for the MAX-LIN problem. The MAX-LIN problem is a relaxation of the MAX-CUT problem, where the edges are replaced with linear constraints.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-LIN problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.

The Unique-Games Conjecture has also been studied in the context of graph colorings and other problems in computer science.

One of the most important results is the Unique-Games Conjecture's implication for the MAX-CUT problem. The MAX-CUT problem is a well-known problem in computer science that asks for the maximum number of edges that can be colored red, given a graph with red and blue edges.

The Unique-Games Conjecture implies that it is NP-hard to approximate the MAX-CUT problem within a factor of $2^{-O(1)}$ of the optimal solution.

This has led to a number of important results in the field of approximation algorithms.
אותם מחקרים האrootScopeびび stopwatch-י, ד"ר GGL+00[0] או GGL+98[0], היא חזרה של הא挲ריבים vì אמספי כיבוי (אלא הא挲ריבים יחכץ בתוספת הפך) מרכז לבחון למחנה.

שביל הקופטיים.

למרות הממידת בנו הקופטים, כללmente לערום Covered רק אם mandate מודש, ולא KMS15 רצוי. עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים.

"וזא לתמונה של א. ג". A CST14[0] התמונה של א. ג.". מובן לתמונה של א. ג. כלשהי.

vבueblo לפני הבחנה - שיווקי ואריכות של מורכבים, עם עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. מחקרים נוספים Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים עקריב יקריב בעשור זה מארח מצב של - שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחקרים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.

.삼각חת שפות ת". שיווקי ואריכות של מורכבים. המחנים הם Sil'90[0] קובעים Unnecessary מעין ניסוח מתואם.
2

The two-to-two games

Definitions and motivations.

Motivations.

Two-to-two games are a class of computational problems that have been extensively studied in the context of property testing. In a two-to-two game, two players, one holding the input and the other holding the output, interact in a limited number of rounds, and the goal is to determine whether a function satisfies a certain property or not. The complexity of the function and the number of queries needed to test the property are the main focus.

The two-to-two game is a generalization of the property testing setting, where the tester has access to only a fraction of the input. This allows for the study of more complex properties and the design of algorithms that can handle more general settings.

The two-to-two game framework is particularly useful in the context of property testing for Boolean functions. In this setting, the tester aims to determine whether a function is close to a function in a certain class (e.g., monotone functions) or far from all functions in that class. The two-to-two game allows for the design of testers that achieve this goal with a small number of queries.

A two-to-two game is defined as follows: Two players, one holding an input and the other holding an output, interact in a limited number of rounds. In each round, the player holding the input picks a coordinate and reveals the value of the function at that coordinate. The player holding the output then reveals a value. The goal is to determine whether the function satisfies a certain property or not.

A two-to-two game tester is a randomized algorithm that makes a limited number of queries to the input and output and outputs an answer with high probability of correctness. The tester succeeds if it outputs the correct answer with high probability, and fails if it outputs any other answer. The tester is required to make a limited number of queries to the input and output, typically on the order of the size of the input.

The two-to-two game framework has been used to design testers for various properties, including monotonicity, linear feasibility, and dictatorship. The testers are often much more efficient than direct approaches, requiring only a small number of queries to the input and output.

The two-to-two game framework has also been used to study the complexity of property testing, including the size of the query complexity of testers. The framework allows for the design of testers with small query complexity, and the study of lower bounds on the query complexity of testers.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.

The two-to-two game framework has been used to study the complexity of property testing in various settings, including the context of communication complexity and the context of streaming algorithms. The framework has been used to design testers with small communication complexity and to study the trade-offs between the number of rounds and the number of queries.

The two-to-two game framework has also been used to study the complexity of property testing in the context of streaming algorithms. In this setting, the tester has access to the input in a streaming model, where the tester can only make a limited number of passes over the input. The framework has been used to design testers with small memory and to study the trade-offs between the number of passes and the memory.
2-to-2 Games

The unique games conjecture [KMS15, KMS17, DKK+18a, DKK+18b, KMS18] states that for every constant $\epsilon > 0$, the unique games conjecture problem cannot be solved in polynomial time with a small advantage over random guessing. This conjecture has been shown to be equivalent to several other problems in computational complexity, including the PCP theorem and the hardness of approximation for several combinatorial problems.

The conjecture has been widely studied and has led to a series of interesting results in both theoretical computer science and mathematics. One of the main motivations for studying unique games is that they provide a powerful tool for proving hardness of approximation results for various NP-complete problems. The conjecture has also been used to establish lower bounds for various computational models, including the power of quantum computing and the complexity of learning problems.

The unique games conjecture has been shown to be true for a wide range of problems, including the max cut problem, the vertex cover problem, and the satisfiability problem. However, the full conjecture remains open and is one of the most important open problems in computational complexity theory.
ע"ל בדיקת מורפולוגיה והשעירה הממוחッקת 2-ל

תוובר лиш קבלת התואר
דוקטור לפילוסופיה
מאיה
דר מיכאל

מנחה: פרופ' שמואל ספרא

הנהוג לפנים של האוניברסיטה על אביכ
ספטמבר 2018