



# Pseudorandom Sets in Grassmann Graph have Near-Perfect Expansion

Subhash Khot\*

Dor Minzer<sup>†</sup>Muli Safra<sup>‡</sup>

## Abstract

We prove that pseudorandom sets in Grassmann graph have near-perfect expansion as hypothesized in [14]. This completes the proof of the 2-to-2 Games Conjecture (albeit with imperfect completeness) as proposed in [34, 13, 14], along with contributions from [39, 9, 33].

The Grassmann graph  $Gr_{\text{global}}$  contains induced subgraphs  $Gr_{\text{local}}$  that are themselves isomorphic to Grassmann graphs of lower orders. A set is called pseudorandom if its density is  $o(1)$  inside all subgraphs  $Gr_{\text{local}}$  whose order is  $O(1)$  lower than that of  $Gr_{\text{global}}$ . We prove that pseudorandom sets have expansion  $1 - o(1)$ , greatly extending the results and techniques in [14].

---

\*Department of Computer Science, Courant Institute of Mathematical Sciences, New York University. Research supported by NSF CCF-1422159, Simons Collaboration on Algorithms and Geometry, and Simons Investigator Award.

<sup>†</sup>School of Computer Science, Tel Aviv University. Research supported by Clore Fellowship, ISF 2013/17, BSF 2016414 and Blavatnik fund.

<sup>‡</sup>School of Computer Science, Tel Aviv University. Research supported by ISF 2013/17, BSF 2016414 and Blavatnik fund.

# 1 Introduction

This paper completes the proof of the 2-to-2 Games Conjecture (albeit with imperfect completeness) proposed by Dinur, Kindler, and the authors of this paper [34, 13, 14], along with contributions from Barak, Kothari, and Steurer [9] and Moshkovitz and the authors of this paper [39, 33]. The 2-to-2 Games Conjecture has several implications towards approximability of NP-hard problems (and more widely to computational complexity, optimization, and combinatorics). Its proof is completed by proving a certain combinatorial hypothesis proposed in [14] regarding the expansion properties of Grassmann graph. While the main focus of this paper is on the combinatorial hypothesis (stated as Theorem 1.12), we present the broader context here for reader's benefit.

## 1.1 PCPs, Vertex Cover, Independent Set

An approximation algorithm for an NP-hard problem is an efficient algorithm that computes a solution that is guaranteed to be within a certain multiplicative factor of the optimum (known as the approximation factor). It turns out that for several NP-hard problems, even computing an approximate solution, within a certain multiplicative factor of the optimum, remains an NP-hard problem (known as the hardness factor). The complementary study of approximation algorithms and hardness of approximation aims at characterizing precise approximation thresholds for NP-hard problems of interest, i.e. the threshold at which the approximation factor and the hardness factor (essentially) match.

The hardness of approximation results build on a foundational result known as the Probabilistically Checkable Proofs (PCP) Theorem [18, 3, 2]. The theorem can be viewed from a hardness viewpoint as well as from a proof checking viewpoint. From the hardness viewpoint, it states that there exists an absolute constant  $\beta < 1$  such that, given a 3SAT formula  $\phi$ , it is NP-hard to distinguish whether it is satisfiable or whether it is at most  $\beta$ -satisfiable (i.e. no assignment satisfies more than  $\beta$  fraction of the clauses). From the proof checking viewpoint, it states that every NP-statement has a polynomial size proof that can be checked efficiently by a probabilistic verifier that reads only a constant number of bits from the proof. The verifier is complete and sound in the sense that a correct proof of a correct statement is accepted with probability 1 and any proof of an incorrect statement is accepted with probability at most, say  $\frac{1}{2}$ .

The equivalence between the hardness and the proof checking viewpoints, though not difficult to see, has led to many illuminating insights and strong hardness results over the last three decades. The proof checking viewpoint (whose roots go back to the work on interactive proofs) played a decisive role in the discovery of the PCP Theorem. However, for the sake of uniformity and ease of presentation, we adopt the hardness viewpoint here. We refer the reader to the surveys [1, 25, 47, 30] for an overview of the extensive and influential body work on PCPs and hardness results. While we know, by now, optimal hardness results for a handful of problems, e.g. 3SAT, Clique, Set Cover [23, 24, 17], for a vast majority of problems there remains a significant gap between the best known approximation factor and the best known hardness factor.

One such problem is Vertex Cover. Given a graph  $G = (V, E)$ , a subset  $C \subseteq V$  is called a vertex cover if for every edge  $e = (u, v) \in E$ , either  $u$  or  $v$  is in  $C$ . Finding a minimum vertex cover is a well known NP-hard problem [27]. It admits a simple 2-approximation, namely an efficient algorithm that outputs a cover  $C$  of size at most twice the minimum. The algorithm picks an arbitrary edge  $e = (u, v) \in E$ , adds both  $u, v$  to  $C$ , removes all edges that are incident on either  $u$  or  $v$ , and repeats this step. Whether there exists an algorithm with approximation factor strictly below 2 is among the flagship questions in approximability. Surprisingly, there is now good reason to believe that no such algorithm exists, i.e. it is conceivable that

approximating Vertex Cover within factor  $2 - \varepsilon$  is NP-hard!<sup>1</sup>

An independent set in a graph is complement of a vertex cover, i.e. a subset of vertices that contains no edge inside. Towards proving hardness results for Vertex Cover, it is more convenient to prove hardness results for the Independent Set problem (in the special case where the concern is independent sets of *linear* size). Let  $0 < s < c \leq \frac{1}{2}$  be constants and let  $\text{GapIS}(c, s)$  be a “promise gap problem” where an  $n$ -vertex graph is given with the promise that either it contains an independent set of size  $cn$  or contains no independent set of size  $sn$  and the algorithmic task is to distinguish between the two cases. It is easy to see that if  $\text{GapIS}(c, s)$  is NP-hard, then approximating Vertex Cover within factor  $\frac{1-s}{1-c}$  is NP-hard. Thus, to prove  $2 - \varepsilon$  hardness result for Vertex Cover, it is sufficient (and turns out necessary)<sup>2</sup> to prove hardness of  $\text{GapIS}(c, s)$  where  $s \rightarrow 0$  and  $c \rightarrow \frac{1}{2}$ .

There is a simple but nice connection between hardness of the Independent Set problem and PCPs that is worth pointing out. As noted in [11], the hardness of  $\text{GapIS}(c, s)$  is equivalent to a PCP with “zero free bits”, namely a PCP where the verifier has completeness  $c$ , soundness  $s$ , and has exactly one accepting answer to her queries (and thus “knows” the answer before reading the queries!). This seems contradictory: if the verifier knows the answer(s) beforehand, why wouldn’t she be able to construct the correct proof by herself? The subtle point is that the PCP has imperfect completeness (i.e.  $c < 1$ ), so even for a correct proof of a correct statement, only a fraction  $c$  of the answers that the verifier “knows” are actually correct and the verifier cannot tell which  $c$  fraction of the answers (among polynomially many) are correct. Indeed, the task of determining which of the answers are correct amounts to finding an independent set in a related graph (hence the connection).

Håstad, building on the works of Bellare, Goldreich, and Sudan [11] and Raz [45], proved that  $\text{GapIS}(\frac{1}{4} - \varepsilon, \frac{1}{8} + \varepsilon)$  is NP-hard, implying a hardness factor  $\frac{7}{6} - \varepsilon \approx 1.16$  for Vertex Cover. Dinur and Safra [16] proved that  $\text{GapIS}(p - \varepsilon, 4p^3 - 3p^4 + \varepsilon)$  is NP-hard for  $p \leq \frac{3 - \sqrt{5}}{2}$ , implying a hardness factor  $10\sqrt{5} - 21 - \varepsilon \approx 1.36$  for Vertex Cover. Their paper introduced several techniques, e.g. the Biased Long Code, application of Fourier analytic theorems on Boolean hypercube, and implicitly, the notion of 2-to-2 Games. Fourier analysis has since become ubiquitous in analysis of PCP constructions. The hardness result of [16] remained the best hardness result for Vertex Cover (until the present work).

As noted, the quest towards proving optimal hardness results was stalled after the remarkable but relatively few successes. In [29], the author introduced the Unique Games Conjecture as a plausible avenue to make further progress and presented a hardness result for the Min-2SAT-Deletion problem as a demonstration. In addition, motivated by the Dinur-Safra paper (where the 2-to-2 Games appeared implicitly), the author also introduced the  $d$ -to- $d$  Conjecture for  $d \geq 2$  and showed that it implies that  $\text{GapIS}(1 - 2^{-1/d} - \varepsilon, \varepsilon)$  is NP-hard. For  $d = 2$ , this would give a hardness factor of  $\sqrt{2} - \varepsilon \approx 1.42$  for Vertex Cover, an improvement over Dinur-Safra result.<sup>3</sup> Subsequently, in [38], it was shown that the Unique Games Conjecture implies NP-hardness of  $\text{GapIS}(\frac{1}{2} - \varepsilon, \varepsilon)$  and hence hardness of approximating Vertex Cover within  $2 - \varepsilon$ .

<sup>1</sup>Unless stated otherwise,  $\varepsilon > 0$  will denote an arbitrarily small constant and the statements are meant to hold for every such  $\varepsilon$ .

<sup>2</sup>One notes that if a graph contains an independent set of fractional size  $c \geq \frac{1}{2} + \varepsilon$ , then by obtaining a 2-approximation for Vertex Cover and taking the complement, one finds an independent set of fractional size  $\geq 2\varepsilon$ . Therefore, the best  $(c, s)$  hardness gap one can hope for Independent Set is  $c = \frac{1}{2} - \varepsilon, s = \varepsilon$ .

<sup>3</sup>For  $d \geq 3$ , while one does not get improvement for Vertex Cover, one does get NP-hardness for  $\text{GapIS}(c, s)$  for a fixed  $c$  and  $s \rightarrow 0$ , which in authors’ opinion is a more fundamental issue.

## 1.2 The Unique Games Conjecture

For the purposes of this paper, it suffices to define the Unique Game as the following computational problem. 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 1.1.** *An instance  $\mathcal{U}$  of the UniqueGame $[\mathbb{F}_2^\ell]$  problem consists of  $n$  variables  $x_1, \dots, x_n$  taking values over (the alphabet)  $\mathbb{F}_2^\ell$  and  $m$  constraints  $C_1, \dots, C_m$  where each constraint  $C_i$  is a linear equation of form  $x_{i_1} \oplus x_{i_2} = b_i$  and  $b_i \in \mathbb{F}_2^\ell$ . Let  $\text{OPT}(\mathcal{U})$  denote the maximum fraction of the constraints that can be satisfied by any assignment to the instance.*

The term “unique” refers to the specific nature of the constraints: for every assignment to the variable  $x_{i_1}$ , there is a unique assignment to the variable  $x_{i_2}$  that satisfies the constraint and vice versa (the Unique Game problem was studied earlier by Feige and Lovász in the context of parallel repetition [19]). For constants  $0 < s < c < 1$ , let  $\text{GapUG}[\mathbb{F}_2^\ell](c, s)$  be the gap-version where the instance  $\mathcal{U}$  of the UniqueGame $[\mathbb{F}_2^\ell]$  problem is promised to have either  $\text{OPT}(\mathcal{U}) \geq c$  or  $\text{OPT}(\mathcal{U}) \leq s$ . The Unique Games Conjecture states that<sup>4</sup>.

**Conjecture 1.2.** *For every constant  $\varepsilon > 0$ , there exists a sufficiently large integer  $\ell = \ell(\varepsilon)$  such that  $\text{GapUG}[\mathbb{F}_2^\ell](1 - \varepsilon, \varepsilon)$  is NP-hard.*

The Unique Games Conjecture has since been established as a prominent open question in theoretical computer science (please see the surveys [48, 30, 28, 31]). The conjecture has facilitated numerous connections among algorithm design, computational complexity, geometry, and analysis. In particular, assuming the conjecture, optimal hardness results are known for Vertex Cover, Max Cut, Max Acyclic Subgraph [38, 32, 21], super-constant hardness results for Sparsest Cut [12, 40], and one can even deduce optimality of generic approximation algorithms for the entire class of Constraint Satisfaction Problems [43]. In spite of this rather large body of work, the correctness of the conjecture itself remains open.

Over the last decade, several arguments were put forward against the Unique Games Conjecture. We sketch these arguments here to the best of our knowledge.<sup>5</sup>

- There is no known distribution over instances of the Unique Game problem that is plausibly hard (more on this below). In fact, results in [5, 41] showed that the problem is easy on “semi-random” instances (for a rather generous interpretation of the term semi-random), thus indicating otherwise.
- No integrality gap instances are known for the Unique Game problem for a constant number, say 10, of rounds of the Sum-of-Squares (a.k.a. Lasserre, Parrilo) hierarchy of semi-definite programming relaxation. Stated differently: arguably, a constant number of rounds of the Sum-of-Squares hierarchy might already qualify as an efficient algorithm for the problem, disproving the Unique Games Conjecture.
- Results in [7] showed that the known integrality gap instances of the problem (that hold for a very basic SDP relaxation)<sup>6</sup> do not survive a few additional rounds of the Sum-of-Squares hierarchy. The

---

<sup>4</sup>The original statement in [29] refers to more general constraints. However it follows from [32] that the original conjecture is equivalent to the statement here, i.e. when the constraints are linear equations over the group  $\mathbb{F}_2^\ell$ .

<sup>5</sup>Since these arguments were not made by us, we are not taking responsibility as to whether these arguments were indeed made or whether these are/were considered pressing arguments. We present these only for reader’s benefit.

<sup>6</sup>Qualitatively speaking, there is just one example, from [40]. Lack of qualitatively different examples might itself have been an argument against the Unique Games Conjecture.

authors argued that there is a barrier for the “common” techniques used to construct SDP integrality gaps (in the spirit of the natural proofs barrier for circuit lower bounds).

- Arora-Barak-Steurer [4] presented an algorithm that solves  $(1 - \varepsilon)$ -satisfiable instances of the Unique Game problem (i.e. finds say  $\frac{1}{10}$ -satisfying assignment) in time  $2^{n^{\varepsilon'}}$  where  $\varepsilon'$  depends on  $\varepsilon$ . The algorithm was later incorporated into the Sum-of-Squares framework as well [10]. If the running time of the algorithm is improved so that  $\varepsilon' \rightarrow 0$  independently of  $\varepsilon$ , then the Unique Game problem would not be NP-hard.<sup>7</sup> The improvement could arguably come from a quantitative improvement in the connection between the number of large eigenvalues and expansion of small sets in graphs.
- The Arora-Barak-Steuter algorithm and the Unique Games Conjecture together imply that the Unique Game problem has “intermediate complexity” (see Section 1.3), a behavior one might not expect for a constraint satisfaction problem.

Given these arguments, the situation was remedied somewhat (from the viewpoint of a believer in the Unique Games Conjecture) in the recent works on “PCPs over reals” [36, 35]. The authors therein construct a candidate explicit family of mildly hard Unique Game instances (the qualifier “candidate” acknowledges that the authors are unable to provide a soundness analysis for the construction). The construction is best viewed as a candidate integrality gap for the problem  $\text{GapUG}[\mathbb{F}_2](1 - \varepsilon, 1 - K(\varepsilon) \cdot \varepsilon)$  for  $K(\varepsilon)$  rounds of the Sum-of-Squares relaxation. Here the Unique Game problem is over the binary alphabet and  $K(\varepsilon) \rightarrow \infty$  as  $\varepsilon \rightarrow 0$ , i.e. the gap is (hypothesized to be) super-constant in terms of the fraction of the unsatisfied constraints and for a super-constant number of Sum-of-Squares rounds. One notes that even though the (hypothesized) gap would be mild, it nevertheless would be a big progress. It is known for instance that if the Unique Game problem were shown NP-hard with a mild gap  $(1 - \varepsilon, 1 - K(\varepsilon) \cdot \sqrt{\varepsilon})$ , the gap could then be boosted by parallel repetition [45, 26, 44] to the gap  $(1 - \varepsilon', \varepsilon')$ , proving the Unique Games Conjecture in its fullness.

The context hitherto leads finally to the present line of work: a sequence of papers [34, 13, 14] presented an approach towards proving the related 2-to-2 Games Conjecture<sup>8</sup> (or alternately the Unique Games Conjecture with completeness  $\frac{1}{2}$ , if reader finds it more convenient to think about). In conjunction with a missing link provided in [9], the approach finally reduced the 2-to-2 Games Conjecture to a concrete combinatorial hypothesis regarding the expansion properties of the Grassmann graph (stated as Theorem 1.12). In the present paper, we prove this combinatorial hypothesis, thus completing the proof of the 2-to-2 Games Conjecture (now a theorem).<sup>9</sup> The 2-to-2 Games Theorem gives, among other things, a strong evidence towards the Unique Games Conjecture (in our opinion). All the arguments against the Unique Games Conjecture that we described apply equally well to the 2-to-2 Games Conjecture and in spite of it, the 2-to-2 Games Conjecture, at the end of the day, does happen to be correct!

### 1.3 The 2-to-2 Games Theorem and its Significance

We now state the 2-to-2 Games Theorem formally, describe its significance, and sketch the developments that led to it. This is followed by the description of the Grassmann graph and its role in these developments.

<sup>7</sup>Under the rather standard hypothesis  $\text{NP} \not\subseteq \cap_{\gamma>0} \text{DTIME}(2^{n^\gamma})$ .

<sup>8</sup>Albeit with imperfect completeness; we will ignore this issue henceforth.

<sup>9</sup>As for proving the Unique Games Conjecture itself, in the opinion of the first-named author, the approach in [36, 35] is a more viable approach.

**Definition 1.3.** An instance  $\mathcal{U}_{2\leftrightarrow 2}$  of the 2-to-2 Game $[\mathbb{F}_2^\ell]$  problem consists of  $n$  variables  $x_1, \dots, x_n$  taking values over (the alphabet)  $\mathbb{F}_2^\ell$  and  $m$  constraints  $C_1, \dots, C_m$  where each constraint is of the form  $T_{ij}x_i \oplus T'_{ij}x_j \in \{b_{ij}, b'_{ij}\}$ ,  $T_{ij}, T'_{ij}$  are  $\ell \times \ell$  invertible matrices, and  $b_{ij}, b'_{ij} \in \mathbb{F}_2^\ell$ . Let  $\text{OPT}(\mathcal{U}_{2\leftrightarrow 2})$  denote the maximum fraction of the constraints that can be satisfied by any assignment to the instance.

The term “2-to-2” refers to the specific nature of the constraints: for every assignment to the variable  $x_i$ , there are exactly two assignments to the variable  $x_j$  that satisfy the constraint and vice versa. For constants  $0 < s < c \leq 1$ , let  $\text{Gap 2-to-2}[\mathbb{F}_2^\ell](c, s)$  be the gap-version where the instance  $\mathcal{U}_{2\leftrightarrow 2}$  of the 2-to-2 Game $[\mathbb{F}_2^\ell]$  problem is promised to have either  $\text{OPT}(\mathcal{U}_{2\leftrightarrow 2}) \geq c$  or  $\text{OPT}(\mathcal{U}_{2\leftrightarrow 2}) \leq s$ .<sup>10</sup> The 2-to-2 Games Theorem is stated below along with an immediate corollary for the hardness of the Unique Game problem with completeness  $\frac{1}{2}$ . The latter is obtained by writing each 2-to-2 Game constraint as a pair of Unique Game constraints so that in the completeness case, there is a  $\frac{1}{2}(1 - \varepsilon)$ -satisfying assignment. The completeness can be increased artificially to precisely  $\frac{1}{2}$  by adding a small fraction of constraints that are always satisfied.

**Theorem 1.4.** For every constant  $\varepsilon > 0$ , there exists a sufficiently large integer  $\ell = \ell(\varepsilon)$  such that  $\text{Gap 2-to-2}[\mathbb{F}_2^\ell](1 - \varepsilon, \varepsilon)$  is NP-hard.

**Theorem 1.5.** For every constant  $\varepsilon > 0$ , there exists a sufficiently large integer  $\ell = \ell(\varepsilon)$  such that  $\text{Gap UG}[\mathbb{F}_2^\ell](\frac{1}{2}, \varepsilon)$  is NP-hard.

### 1.3.1 Implications of the 2-to-2 Games Theorem

We now summarize the main implications of the 2-to-2 Games Theorem (with imperfect completeness; some of these implications depend on its specific proof obtained in the present and previous works). As before,  $\varepsilon > 0$  denotes constant that can be taken as arbitrarily small.

- **Hardness Results**

The following results were already known based on the 2-to-2 Games Conjecture (as indicated in the references; perfect completeness in the last two results if 2-to-2 Games Conjecture holds with perfect completeness). These represent a big progress, in our opinion, on flagship problems in approximability.

- [37]:  $\text{Gap Max Cut}(\frac{1}{2} + \Omega(\varepsilon), \frac{1}{2} + \frac{\varepsilon}{\log(1/\varepsilon)})$  is NP-hard. This is optimal up to the constant in the  $\Omega$ -notation.
- [29]:  $\text{Gap Independent Set}(1 - \frac{1}{\sqrt{2}} - \varepsilon, \varepsilon)$  is NP-hard and as a corollary, Vertex Cover is NP-hard to approximate within a factor strictly less than  $\sqrt{2}$ . Between these two implications, the “correct gap-location” (arbitrarily low soundness) for the Independent Set problem is more fundamental.

---

<sup>10</sup>Comments regarding the original formulation of this conjecture in [29]: (1) It was proposed with *perfect completeness*, i.e. stating that  $\text{Gap 2-to-2}[\mathbb{F}_2^\ell](1, \varepsilon)$  is NP-hard. However, as far as this paper is considered, we view the issue of perfect versus imperfect completeness as being relatively minor. (2) It was proposed with more general constraints (rather than the special case with linear structure described herein) and with “2-to-1” constraints (rather than with “2-to-2” constraints described herein; the conjecture was referred to as the 2-to-1 Conjecture). Both these are non-issues however: the current and preceding works [34, 13, 14] now prove the conjecture *with* linear structure and the constraints are easily reinterpreted as being 2-to-1 constraints (hence proving the 2-to-1 Conjecture as well, which in any case is morally equivalent to the 2-to-2 Conjecture).

- [15]: It is NP-hard to distinguish whether a graph has four disjoint independent sets of (relative) size  $\frac{1}{4} - \varepsilon$  each (and hence is almost 4-colorable) or whether there is no independent set of (relative) size  $\varepsilon$  (and hence is not almost  $(\frac{1}{\varepsilon})$ -colorable).
- [22]: It is NP-hard to properly color (using  $k$  colors) more than a fraction  $1 - \frac{1}{k} + O(\frac{\log k}{k^2})$  edges of an almost  $k$ -colorable graph. This is optimal up to the constant in the  $O$ -notation.

- **Integrality Gaps, Plausibly Hard Distributions, Cart before the Horse**

The present line of work gives a reduction from the 3Lin problem to the 2-to-2 Game problem and subsequently to the Unique Game problem with completeness  $\frac{1}{2}$  and to the graph theoretic problems mentioned above. Denoting any of these problems by  $\mathcal{P}$ , the reduction can be used

- To “translate” an integrality gap instance of the 3Lin problem (see below) to an integrality gap instance of the problem  $\mathcal{P}$  (the idea to use a reduction to construct integrality gap was used earlier in [40]).
- To “translate” a distribution over 3Lin instances that is plausibly hard (e.g. random instances with appropriate parameters) to a distribution over  $\mathcal{P}$  instances that is plausibly hard.

In both cases, we do not know an alternate construction, i.e. without having to go through a NP-hardness reduction (and lack of any construction so far was an argument against the Unique Games Conjecture as discussed before). On the other hand, “logically”, integrality gap construction (and maybe construction of a plausibly hard distribution as well) ought to *precede* an NP-hardness reduction. We find this “cart before the horse” phenomenon quite interesting.

- **Sum-of-Squares Integrality Gaps with Perfect Completeness**

- If one concerns integrality gap (say up to a polynomial number of rounds of the Sum-of-Squares relaxation), the previous result for graph coloring holds with perfect completeness. I.e. there is a graph along with an SDP solution such that (a) the SDP solution pretends as if the graph is 4-colorable whereas (b) in actuality, the graph has no independent set of size  $\varepsilon$ .
- Integrality gap (say up to a polynomial number of rounds of the Sum-of-Squares relaxation) for the 2-to-2 Games problem holds with perfect completeness and soundness  $\varepsilon$ .

These results are a consequence of the integrality gap known for the 3Lin problem with perfect completeness [20, 46]. The integrality gap instance for 3Lin can be translated via the reduction as remarked above.

- **Intermediate Complexity Theorem**

Barak [6] pointed out that Theorem 1.5, along with the Arora-Barak-Steurer algorithm and the (rather standard) hypothesis  $\text{NP} \not\subseteq \bigcap_{\gamma > 0} \text{DTIME}(2^{n^\gamma})$ , implies the Intermediate Complexity Theorem:

*For every  $\varepsilon > 0$ , there exist  $\varepsilon' > \varepsilon'' > 0$  and integer  $\ell$  such that  $\varepsilon' \rightarrow 0$  as  $\varepsilon \rightarrow 0$  and the promise constraint satisfaction problem  $\text{GapUG}[\mathbb{F}_2^\ell](\frac{1}{2}, \varepsilon)$  on  $n$  variables can be solved in time  $2^{n^{\varepsilon'}}$  but not in time  $2^{n^{\varepsilon''}}$ .*

This is perhaps surprising (and was perhaps cited as an argument against the Unique Games Conjecture). The past experience (e.g. the Dichotomy Conjecture/Theorem, the Exponential Time Hypothesis, near-linear sized PCPs etc) perhaps suggested (and if so, incorrectly as it turns out) that

time complexity of  $n$ -variable CSPs ought to be either polynomial or truly exponential, i.e.  $2^{\Omega(n)}$  amounting to a brute-force search over all assignments.

- **Evidence towards the Unique Games Conjecture**

$\text{GapUG}(\frac{1}{2}, \varepsilon)$  is NP-hard, i.e. a weaker form of the Unique Games Conjecture holds with completeness  $\frac{1}{2}$ . As far as the authors know (and we have consulted the algorithmic experts), the known algorithmic attacks on the Unique Game problem work equally well whether the completeness is  $\approx 1$  or whether it is  $\frac{1}{2}$ . Thus, the implication that  $\text{GapUG}(\frac{1}{2}, \varepsilon)$  is NP-hard is a compelling evidence, in our opinion, that the known algorithmic attacks are (far) short of disproving the Unique Games Conjecture.

Moreover, as remarked before, all the arguments against the Unique Games Conjecture, sketched in Section 1.2, apply equally well to its weaker form with completeness  $\frac{1}{2}$ . In spite of all these arguments, the  $\text{GapUG}(\frac{1}{2}, \varepsilon)$  problem, at the end of the day, does happen to be NP-hard, circumventing all the arguments mentioned!

- **Unique Games Conjecture versus the Small Set Expansion Conjecture**

Raghavendra and Steurer [42] proposed the Small Set Expansion Conjecture and showed that it implies the Unique Games Conjecture. Roughly speaking, it states that  $\text{GapSSE}(\varepsilon, 1 - \varepsilon)$ , the problem of distinguishing whether a graph has a small set of expansion at most  $\varepsilon$  or whether every small set has expansion at least  $1 - \varepsilon$ , is NP-hard.

The 2-to-2 Games Theorem arguably supports the authors’ 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  $\text{GapSSE}[\varepsilon, 1 - \varepsilon]$  to  $\text{GapUG}[1 - \varepsilon', \varepsilon']$ . The same reduction also shows that  $\text{GapSSE}[\beta, 1 - \varepsilon]$  reduces to  $\text{GapUG}[\frac{1}{2}, \varepsilon']$  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 authors’ suspicion. Indeed, this is precisely what happens in the proof of the 2-to-2 Games Theorem. One gets a reduction to  $\text{GapUG}[\frac{1}{2}, \varepsilon']$  without getting a reduction to  $\text{GapSSE}$ ; the graphs in the reduction *always* have small non-expanding sets.

### 1.3.2 Works Leading to the 2-to-2 Games Theorem

As noted, the 2-to-2 Games Theorem is proven in a sequence of works [34, 13, 14] and completed in the present work. In addition, contributions from [39, 9, 33] have been crucial in the overall proof. We summarize these developments below. At a high level, the proof involves the chain of implications (all these are now theorems):

$$\text{Grassmann Expansion Hypothesis} \stackrel{[14]}{\implies} \text{Linearity Testing Hypothesis} \stackrel{[9]}{\implies} \text{2 to 2 Games Conjecture} \stackrel{[34, 13]}{\implies} \text{2 to 2 Games Conjecture}.$$

- The Grassmann graphs and their potential application to the 2-to-2 Game problem were proposed in [34]. The contributions therein were: (1) introducing a certain linearity testing primitive based on the Grassmann graph/code (2) using a certain “sub-code covering” property of this code analogous to a similar property of the Hadamard code, previously introduced in [39] (3) proposing a reduction to the 2-to-2 Game problem (4) proposing a Weak Linearity Testing Hypothesis and showing that it implies, via the reduction, a Weak 2-to-2 Games Conjecture. We do not elaborate on the qualifier “weak” here.

It refers to a rather awkward variant that is nevertheless quite natural and useful as far as application to Independent Set and Vertex Cover is concerned, which was the main motivation in [34].

- In [13], the authors formulated a Linearity Testing Hypothesis (the “right” and clean formulation in hindsight) and showed that it implied, via a reduction that is very similar to that in [34], the 2-to-2 Games Conjecture (with imperfect completeness).
- In [34, 13], it was already clear that the connectivity and expansion properties of the Grassmann graph would be crucial towards proving the Linearity Testing Hypotheses therein. In [14], the authors proposed (let’s call it) Grassmann Expansion Hypothesis (stated as Theorem 1.12 in the present paper), and argued that it would at least be necessary towards proving the Linearity Testing Hypothesis. The authors presented a Fourier analytic framework and a preliminary set of results (for the first and second Fourier levels) towards proving the Grassmann Expansion Hypothesis.
- Barak, Kothari, and Steurer [9] proved that the Grassmann Expansion Hypothesis (almost immediately) implies the Linearity Testing hypothesis. While simple, this link is nevertheless important and was missed by the authors of [14].
- Finally, the Grassmann Expansion Hypothesis is proved in the present paper, stated as Theorem 1.12. A similar Johnson Expansion Hypothesis is proved in [33] and the technical insight therein has been useful in the present paper.

We now discuss the Grassmann Linearity Testing Hypothesis, the Grassmann Expansion Hypothesis, and indicate how the latter implies the former. We do not attempt here an overview of how the Linearity Testing Hypothesis is used to analyze a PCP/reduction and prove the 2-to-2 Games Conjecture. The reader is instead referred to the introductory section of the paper [13].

## 1.4 The Grassmann Graph/Code/Test and Linearity Testing Hypothesis

In the following, one thinks of the parameter  $\ell$  as a sufficiently large integer and (after fixing it) the parameter  $k$  as a sufficiently large integer.

**Definition 1.6.** *The Grassmann graph  $Gr_{k,\ell}$  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  $\dim(L \cap L') = \ell - 1$ .*

Associated with the Grassmann graph is the Grassmann code that encodes linear functions  $f: \mathbb{F}_2^k \rightarrow \mathbb{F}_2$ . The encoding of a linear function  $f$  is given by a word/table  $F[\cdot]$  that assigns to each vertex  $L$  of the graph, the restriction of  $f$  to  $L$ , i.e.  $F[L] = f|_L$ . Since there are  $2^\ell$  linear functions on an  $\ell$ -dimensional space, the alphabet for the encoding has size  $2^\ell$ . The Grassmann code is equipped with a natural testing primitive that we call the Grassmann Linearity Test: given a word  $F[\cdot]$  (not necessarily a codeword), the test picks an edge  $(L, L')$  uniformly at random from the graph and checks that  $F[L]|_{L \cap L'} = F[L']|_{L \cap L'}$ , i.e. that the linear functions  $F[L]$  and  $F[L']$  are consistent on the common intersection of  $L, L'$ .

It is observed immediately that the test is a “2-to-2 test” in the sense that for every assignment/answer  $F[L]$  there are exactly two answers to  $F[L']$  so that the test accepts (and vice versa). This is because a linear function on  $L \cap L'$  can be extended to  $L$  (and similarly to  $L'$ ) in exactly two ways (and this is how the test eventually leads to hardness of 2-to-2 Games). By design, the test has perfect completeness: if  $F[\cdot]$  is a codeword, then the test passes with probability 1 since  $F[L], F[L']$  are then restrictions of the same global linear function. The question of interest is what about the soundness of the test? I.e. if a given word  $F[\cdot]$

passes the test with (small) probability  $\geq \delta$ , what “decoding” could we infer? Could we infer that the given word  $F[\cdot]$  necessarily has good consistency with some codeword (and if so, list-decode)? Based on the past experience (e.g. the Low Degree Test and the Blum-Luby-Rubinfeld Test that are well-understood and are crucial building blocks of PCPs), one is tempted to speculate that the answer is positive, formally stated below.<sup>11</sup> Here  $\delta, \varepsilon$  are thought of as constants independent of the parameters  $k, \ell$ .

**Speculation 1.7.** *For every  $\delta > 0$ , there exists  $\varepsilon > 0$  such that if a table  $F[\cdot]$  passes the Grassmann Linearity Test with probability  $\delta$ , then there exists a global linear function  $f: \mathbb{F}_2^k \rightarrow \mathbb{F}_2$  such that*

$$\Pr_L [F[L] = f|_L] \geq \varepsilon.$$

It turns out however that the speculation is false, the key reason being that the Grassmann graph has small sets whose (edge-)expansion is strictly bounded away from 1.

**Definition 1.8.** *Let  $G = (V, E)$  be an  $n$ -vertex,  $d$ -regular graph. For a non-empty set of vertices  $S \subseteq V$  with  $|S| \leq \frac{n}{2}$ , its (edge-)expansion is defined as*

$$\Phi(S) = \frac{|E(S, \bar{S})|}{d \cdot |S|},$$

where  $E(S, \bar{S})$  denotes the set of edges with one endpoint in  $S$  and the other in  $\bar{S} = V \setminus S$ .

Alternately,  $\Phi(S)$  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  $S$ . We will be interested in whether a set  $S$  has expansion very close to 1 (near-perfect expansion) or has expansion strictly bounded away from 1.

### Counter-example to Speculation 1.7

Consider the following construction (it will be clear soon what the sets  $S_i$  would be):

1. Let  $S_1, \dots, S_m$  be disjoint subsets of vertices of the Grassmann graph  $\text{Gr}_{k, \ell}$ , all of equal size, such that their union constitutes a constant  $\alpha$  fraction of vertices of the graph.
2. The sets  $S_i$  are very small. Specifically,  $m = m(k, \ell) \rightarrow \infty$  as  $k, \ell \rightarrow \infty$ .
3. Suppose that  $\Phi(S_i) \leq 1 - \beta$  for every  $1 \leq i \leq m$  for a constant  $\beta$ .
4. For each  $1 \leq i \leq m$ , select a global linear function  $f_i: \mathbb{F}_2^k \rightarrow \mathbb{F}_2$  at random.
5. Define  $F[L] = f_i|_L$  for every  $L \in S_i$ . For  $L \notin \cup_{i=1}^m S_i$ ,  $F[L]$  is defined at random.

We show that the word/table  $F[\cdot]$  passes the Grassmann test with probability  $\alpha\beta$ , but has negligible consistency with any global linear function. Firstly, since  $S_i$  cover  $\alpha$  fraction of vertices and each  $S_i$  has expansion at most  $1 - \beta$ , the fraction of edges of the Grassmann graph that are inside some  $S_i$  is at least  $\alpha\beta$ . Since on each  $S_i$ , the table  $F[\cdot]$  is consistent with the global function  $f_i$ , the table passes the test for all edges  $(L, L')$  that are inside some  $S_i$ . Secondly, since the functions  $f_i$  on different pieces  $S_i$  are random and unrelated to each other, no single global function has non-negligible consistency with  $F[\cdot]$ . This completes the description of the counter-example.

<sup>11</sup>Moreover, a positive answer would lead to a very straightforward analysis of the PCP/reduction to 2-to-2 Games, avoiding most of the complications in [34, 13].

How does one get around this counter-example, i.e. reformulate Speculation 1.7 so that it is *correct* as well as *sufficient* towards analysis of the PCP/reduction to 2-to-2 Games? With regards to the specific counter-example above, here is a vacuous statement: if we restrict our attention to *only* the subset of vertices in say  $S_1$ , then  $F[\cdot]$  indeed has full consistency with a global linear function, namely the function  $f_1$ . Moreover, as we will see, a canonical example of a small set with expansion strictly bounded away from 1 is  $S = \text{Gr}_{k,\ell}[A, B]$  where  $A \subseteq B \subseteq \mathbb{F}_2^k$  are subspaces with  $\dim(A) + \text{codim}(B) \leq r$  and

$$S = \text{Gr}_{k,\ell}[A, B] = \{L \mid A \subseteq L \subseteq B\}.$$

In this case,  $\phi(S) = 1 - 2^{-r}$  which is strictly bounded away from 1 for small integer  $r$  (say  $r = 4$ ). These observations motivated the following Linearity Testing Hypothesis in [13].

**Hypothesis 1.9.** *For every constant  $\delta > 0$ , there exists a constant  $\varepsilon > 0$  and an integer  $r$  such that for all sufficiently large integers  $\ell$  and (after fixing it) for all sufficiently large integer  $k$ , the following holds. If a table  $F[\cdot]$  passes the Grassmann Linearity Test with probability  $\delta$ , then there exist subspaces  $A \subseteq B \subseteq \mathbb{F}_2^k$  with  $\dim(A) + \text{codim}(B) \leq r$  and a linear function  $f: B \rightarrow \mathbb{F}_2$ , such that*

$$\Pr_{A \subseteq L \subseteq B} [F[L] = f|_L] \geq \varepsilon.$$

In words, while  $F[\cdot]$  need not have good consistency with a global linear function on the *entire* graph  $\text{Gr}_{k,\ell}$ , there must be a *structured subgraph*  $\text{Gr}_{k,\ell}[A, B]$  on which it does have good consistency with a global linear function and moreover this subgraph is of constant “co-order”, defined as  $\dim(A) + \text{codim}(B)$ . The Linearity Testing Hypothesis above was shown to be sufficient towards analysis of the PCP/reduction to 2-to-2 Games in [13]. Towards proving the hypothesis itself, the authors (naturally) proposed to study structure of sets with expansion strictly bounded away from 1, formulated the Grassmann Expansion Hypothesis (essentially) characterizing such sets, and then made partial progress towards proving the Expansion Hypothesis [14]. The authors also argued that proving the Expansion Hypothesis is at least necessary towards proving the Linearity Testing Hypothesis. The missing link, namely that it is also sufficient, was provided in [9].

## 1.5 Grassmann Expansion Hypothesis

**Definition 1.10.** *Suppose  $A \subseteq B \subseteq \mathbb{F}_2^k$  are subspaces. Let  $\dim(A) = a$ ,  $\text{codim}(B) = b$  and think of  $a, b$  as small constants (say  $a = b = 2$ ). Then (as introduced before) the subgraph  $\text{Gr}_{k,\ell}[A, B]$  is an induced subgraph of  $\text{Gr}_{k,\ell}$  induced on precisely the set of vertices  $L$  such that  $A \subseteq L \subseteq B$ . It is easily seen 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  $a + b$  as 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  $\text{Gr}_{k,\ell}[A, B]$ , when seen as a subset of  $\text{Gr}_{k,\ell}$ , has expansion precisely  $1 - 2^{-(a+b)}$  (up to an error  $O(2^{-\ell})$  which is thought of as negligible and ignored for the 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 (we skip the easy proof) 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 1.11.** (Informal): *A subset of constant density inside a constant co-order copy of Grassmann graph inside a Grassmann graph has expansion strictly bounded away from 1.*

(Formal): *Let  $S \subseteq \text{Gr}_{k,\ell}[A, B] \subseteq \text{Gr}_{k,\ell}$  be such that  $\dim(A) = a$ ,  $\text{codim}(B) = b$  and the density of  $S$  inside  $\text{Gr}_{k,\ell}[A, B]$  is  $\varepsilon$ . Then  $\Phi(S) \leq 1 - \varepsilon \cdot 2^{-(a+b)}$ .*

The authors of [14] hypothesize, essentially, that the converse of the above fact is true. Informally, their hypothesis is 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. A precise statement appears below (now as a theorem and the main result in this paper):

**Theorem 1.12.** *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  $\dim(A) = a$ ,  $\text{codim}(B) = b$ ,  $a + b \leq r$  and*

$$\frac{|S \cap \text{Gr}_{k,\ell}[A, B]|}{|\text{Gr}_{k,\ell}[A, B]|} \geq \varepsilon.$$

Following [14], Barak-Kothari-Steurer [9] showed that Theorem 1.12 (a hypothesis at the time) implies Hypothesis 1.9, and by the work of [34, 13] the 2-to-2 Games Conjecture. It remained therefore to prove Theorem 1.12. Partial progress towards its proof was already made in [14] where the authors prove the theorem when  $\alpha < \frac{7}{8}$ , 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  $\alpha < 1 - 2^{-(s+1)}$ , it is easily observed that the indicator vector of the set  $\mathbf{1}_S$  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 [14] when  $s = 2$ , corresponding to  $\alpha < \frac{7}{8}$ . 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. Instead, we are able to argue in a more systematic fashion and avoid the explosion in potential case analysis (easier said than done of course).

We end this section with some remarks on Theorem 1.12. Firstly, the subspaces  $A$  and  $B$  therein are referred to as “zoom-in” and “zoom-out” spaces respectively [34, 13, 14]. This makes sense if one imagines searching for the appropriate subgraph  $\text{Gr}_{k,\ell}[A, B]$  where the set  $S$  happens to have significant density. Secondly, 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. Thirdly, 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  $r = s$  is needed. This is tight (i.e. a lesser zoom-in-out dimension is not sufficient) since we know that subgraphs  $\text{Gr}_{k,\ell}[A, B]$  have expansion  $1 - 2^{-(a+b)}$  and the combined zoom-in-out dimension (obviously)  $a + b$ . Finally, we note that towards proving the theorem, it will be easier to work with the contra-positive: 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).

The phenomena as in Theorem 1.12 occurs also in the Johnson graph and has been analyzed in [33]. In a Johnson graph, the vertices are  $\ell$ -subsets of a  $k$ -set and the edges are  $t$ -wise intersecting pairs (we are concerned with the case when  $t = \lfloor \frac{\ell}{2} \rfloor$ ). Therein the notion of zoom-out and Fourier analysis are not needed. The Johnson case can informally be seen as a special case of the Grassmann case and the analysis of the former in [33] has been insightful in the analysis of the later in the current paper.

## 2 Preliminaries

In this section, we recall and summarize the high-level plan towards proving Theorem 1.12 (or rather the contra-positive), developed already in [14]. The task boils down to upper-bounding the fourth moment of the “projection of the indicator function  $\mathbf{1}_S$  onto the Fourier level- $r$ ”. While this was accomplished in [14] for  $r \leq 2$ , the authors therein were unable to extend it further for two reasons:

- It relied on rather ad hoc case analysis.
- The Fourier analysis on  $\text{Gr}_{k,\ell}$  is unfriendly. It is futile to write down the eigenvectors explicitly and one instead works with the (eigen)space spanned by all eigenvectors with a specific eigenvalue, referred to as the “Fourier level”.
  - The  $r^{\text{th}}$  level has eigenvalue very close to  $2^{-r}$ , but there is a error term.
  - The “union” of the eigenspaces at levels  $0, 1, \dots, r$ , has a clean description: it is spanned by the indicator functions of the subgraphs  $\text{Gr}_{k,\ell}[A, B]$  of co-order at most  $r$ . However to get the hands on precisely the  $r^{\text{th}}$  eigenspace requires “subtracting” the contribution of the previous levels. This leads to rather unfriendly inclusion-exclusion type recursive formulas even for  $r = 2$  and it is not clear how to extend these to higher levels.

For these and additional reasons, authors of [14] worked with approximations to all the quantities and formulas of interest. Extending these approximate formulas to higher Fourier levels seems to incur error terms that are unaffordable.

We are able to circumvent both these obstacles (which to some extent go hand in hand). Firstly, while we still have to consider a large number of elaborate cases, the proof is systematic and works simultaneously for all Fourier levels  $r$  (i.e. without the number of cases exploding with  $r$ ). Secondly, instead of working with the Grassmann graph  $\text{Gr}_{k,\ell}$ , we instead work with a related graph  $\text{H}_{k,\ell}$  (see the definition below). Surprisingly (or perhaps not so surprisingly in hindsight) we are able to write down exact recursive formulas relating quantities at successive Fourier levels. The eigenvalues are exactly  $2^{-i}$  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 3. The recursive formulas therein are systematic extensions, to higher Fourier levels, of the approximate and ad hoc formulas for the second level in [14].

We now define the new graph  $\text{H}_{k,\ell}$  and show that the task of proving Theorem 1.12 reduces to the task of proving an analogous theorem for  $\text{H}_{k,\ell}$  (i.e. Theorem 2.6). Then we describe the very basics of Fourier analysis, just enough to recall the high-level plan (from [14]) that reduces the task further to showing that the indicator function  $\mathbf{1}_S$  of a pseudorandom set  $S \subseteq \text{H}_{k,\ell}$  has low Fourier weight on low Fourier levels (see Section 2.4 and Theorem 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  $\mathbf{1}_S$  (see Sections 2.5, 2.6), which finally is reduced to our main technical result, Theorem 2.15, about upper-bounding closely related “4-wise correlations”.

The main contribution of the paper is Section 3 onwards: the full Fourier analytic machinery is developed in Section 3, upper-bounds on pairwise and 3-wise correlations are presented in Section 4 as a warm-up, and the heart of the paper, the desired upper-bound on the 4-wise correlations, is presented in Sections 5, 6, 7.

## 2.1 Switching to the Graph $H_{k,\ell}$

**Definition 2.1.** 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  $x = (x_1, \dots, x_\ell)$  where  $x_1, \dots, x_\ell \in \{0, 1\}^k$ . Sample  $y \leftarrow \{0, 1\}^k$ ,  $b_1, \dots, b_\ell \leftarrow \{0, 1\}$  independently and uniformly at random. Let the neighbor of  $x$  be  $z = (x_1 + b_1 \cdot y, x_2 + b_2 \cdot y, \dots, x_\ell + b_\ell \cdot y)$ .

The two graphs  $H_{k,\ell}$  and  $\text{Gr}_{k,\ell}$  are closely related as follows:

- The vertices of  $H_{k,\ell}$  are  $\ell$ -tuples of vectors in  $\mathbb{F}_2^k$ . The vertices of  $\text{Gr}_{k,\ell}$  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  $x = (x_1, \dots, x_\ell)$  in  $H_{k,\ell}$  is sampled and then a random edge incident on it is sampled by sampling  $y \leftarrow \{0, 1\}^k$  and  $b_1, \dots, b_\ell \in \{0, 1\}$ , with probability  $\approx 2^{-k} + 2^{-\ell}$ , either  $y = 0$  or  $b = (b_1, \dots, b_\ell) = 0$ , and the edge is a self-loop. Otherwise  $y \neq 0, b \neq 0$  and the other endpoint is  $z = (x_1 + b_1 \cdot y, \dots, x_\ell + b_\ell \cdot y)$ . Provided that both  $x, z$  have full  $\ell$ -dimensional linear span (which happens with probability except  $\approx 2^{\ell-k}$  and we think of  $\ell \ll k$ ), the edge  $(x, z)$  corresponds to a uniformly random edge of the Grassmann graph.

**Remark 2.2.** Barak, Kothari, and Steuter [9] have made a similar suggestion. They consider a graph whose vertices are  $k \times \ell$  matrices and the edges are pairs of matrices that differ by a rank 1 matrix. In terms of our notation, this amounts to an edge  $(x, z)$  with  $z = x \oplus y \otimes b, y \neq 0, b \neq 0$  and  $x, z$  are thought of as  $k \times \ell$  matrices. They seem to be interested in “reducing” the  $k \times \ell$  case to the  $k \times k$  case; the latter is same as the graph of the “degree-2 short code test” as in [8].

## 2.2 It Suffices to Work with $H_{k,\ell}$

We show that Theorem 1.12 for  $\text{Gr}_{k,\ell}$  follows easily from the corresponding Theorem 2.6 for  $H_{k,\ell}$  (see below) and then we work with the graph  $H_{k,\ell}$  for the rest of the paper. It will be convenient to restate Theorem 1.12 in the contra-positive and in terms of “pseudo-random sets”.

**Definition 2.3.** A subset of vertices  $S \subseteq \text{Gr}_{k,\ell}$  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

$$\mu_{\text{in}(A), \text{out}(B)}(S) \stackrel{\text{def}}{=} \frac{|S \cap \text{Gr}_{k,\ell}[A, B]|}{|\text{Gr}_{k,\ell}[A, B]|} \leq \varepsilon.$$

**Theorem 2.4. (Theorem 1.12 restated)** 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  $S \subseteq \text{Gr}_{k,\ell}$  is  $(r, \varepsilon)$ -pseudorandom, then  $\Phi(S) \geq 1 - \zeta$ .

Now we show how to reduce this theorem to Theorem 2.6 below. The reasoning is straightforward. We will show that for every  $S \subseteq \text{Gr}_{k,\ell}$ , there is a natural corresponding set  $S^* \subseteq \text{H}_{k,\ell}$  such that

- Lemma 2.8 below: If  $S$  is  $(r, \varepsilon)$ -pseudorandom, then  $S^*$  is  $(r, \varepsilon)$ -pseudorandom (for a similar notion of being pseudorandom in  $\text{H}_{k,\ell}$  and up to a negligible additive change in the parameter  $\varepsilon$ ).
- Theorem 2.6 below: If  $S^*$  is  $(r, \varepsilon)$ -pseudorandom, then  $\Phi(S^*) \geq 1 - \zeta$ .
- Lemma 2.7 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. The reader may wish to skip the self-evident proofs of Lemmas 2.7 and 2.8. For a set  $S \subseteq \text{Gr}_{k,\ell}$ , the corresponding set  $S^* \subseteq \text{H}_{k,\ell}$  is defined naturally as

$$S^* \stackrel{\text{def}}{=} \{(x_1, \dots, x_\ell) \mid \dim(\text{Span}(x_1, \dots, x_\ell)) = \ell, \text{Span}(x_1, \dots, x_\ell) \in S\}.$$

We note that  $S^*$  is invariant under change of basis, i.e. if  $\text{Span}(x_1, \dots, x_\ell) = \text{Span}(y_1, \dots, y_\ell)$ , then  $(x_1, \dots, x_\ell) \in S^*$  if and only if  $(y_1, \dots, y_\ell) \in S^*$ . We call such subsets of  $\text{H}_{k,\ell}$  basis-invariant. Throughout the paper, we will only concern ourselves with basis-invariant subsets of  $\text{H}_{k,\ell}$ . 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 2.7 and 2.8). The notion of pseudorandom sets in  $\text{H}_{k,\ell}$  is defined in a similar manner.

**Definition 2.5.** A basis-invariant subset of vertices  $S^* \subseteq \text{H}_{k,\ell}$  is called  $(r, \varepsilon)$ -pseudorandom if for any sequence  $Q = (x_1, \dots, x_q)$  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

$$\mu_{\text{in}(Q), \text{out}(W)}(S^*) \stackrel{\text{def}}{=} \Pr_{z_{q+1}, \dots, z_\ell \in W} [(x_1, \dots, x_q, z_{q+1}, \dots, z_\ell) \in S^*] \leq \varepsilon.$$

There is a slight difference between Definitions 2.3 and 2.5. 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 paper. As noted, together with Lemmas 2.7 and 2.8 below, it implies Theorem 1.12 and hence proves the 2-to-2 Games Conjecture.

**Theorem 2.6.** 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  $S^* \subseteq \text{H}_{k,\ell}$  is a basis-invariant  $(r, \varepsilon)$ -pseudorandom set, then  $\Phi(S) \geq 1 - \zeta$ .

**Lemma 2.7.**

$$|\Phi(S^*) - \Phi(S)| \leq 2^{-\ell} + 2 \cdot 2^{\ell-k}.$$

*Proof.* Towards proving the lemma, let  $x = (x_1, \dots, x_\ell) \in S^*$ . Denote its random neighbor by  $z = (x_1 + b_1 \cdot y, \dots, x_\ell + b_\ell \cdot y)$  and  $b = (b_1, \dots, b_\ell)$ . Then

$$\Phi(S^*) = \Pr_{x \in S^*, y, b} [z \notin S^*] = 2^{-\ell} \cdot 0 + (1 - 2^{-\ell}) \Pr_{x \in S^*, y, b \neq 0} [z \notin S^*]. \quad (1)$$

Note that for any  $x \in S^*$ , the vectors  $x_1, \dots, x_\ell$  are linearly independent. Let  $L = \text{Span}(x_1, \dots, x_\ell)$  and  $L' = \text{Span}(x_1 + b_1 \cdot y, \dots, x_\ell + b_\ell \cdot y)$ . Conditioned on  $y \notin L$ ,  $L'$  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

$$\Pr_{\substack{x \in S^*, y, \\ b \neq 0}} [z \notin S^*] = \Pr_{x \in S^*, y} [y \notin L] \cdot \Pr_{\substack{x \in S^*, y, \\ b \neq 0}} [z \notin S^* \mid y \notin L] + \Pr_{x \in S^*, y} [y \in L] \cdot \Pr_{\substack{x \in S^*, y, \\ b \neq 0}} [z \notin S^* \mid y \in L].$$

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  $\Pr_{x \in S^*, y} [y \in L] \leq 2^{\ell-k}$ . Combining everything finishes the proof.  $\square$

**Lemma 2.8.** *If  $S$  is  $(r, \varepsilon)$ -pseudorandom, then  $S^*$  is  $(r, \varepsilon + 2^{\ell+r-k})$ -pseudorandom.*

*Proof.* Towards proving the lemma, we recall Definition 2.5 and consider any sequence  $Q = (x_1, \dots, x_q) \subseteq \mathbb{F}_2^k$  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.

$$\mu_{\text{in}(Q), \text{out}(W)}(S^*) = \Pr_{z_{q+1}, \dots, z_\ell \in W} [(x_1, \dots, x_q, z_{q+1}, \dots, z_\ell) \in S^*].$$

Denoting by  $\mathcal{E}$  the event that  $\{x_1, \dots, x_q, z_{q+1}, \dots, z_\ell\}$  are linearly independent, we have

$$\begin{aligned} \mu_{\text{in}(Q), \text{out}(W)}(S^*) &\leq \Pr[\mathcal{E}] \cdot \Pr_{z_{q+1}, \dots, z_\ell \in W} [(x_1, \dots, x_q, z_{q+1}, \dots, z_\ell) \in S^* \mid \mathcal{E}] + \Pr[\bar{\mathcal{E}}] \\ &= \Pr[\mathcal{E}] \cdot \mu_{\text{in}(\text{Span}(Q)), \text{out}(Q \oplus W)}(S) + \Pr[\bar{\mathcal{E}}]. \end{aligned}$$

The last equality follows from the fact that conditioned on  $\mathcal{E}$ ,  $L = \text{Span}(x_1, \dots, x_q, z_{q+1}, \dots, z_\ell)$  is distributed uniformly among all  $\ell$ -dimensional subspaces containing  $Q$  and contained in  $Q \oplus W$ , and  $L \in S$  if and only if  $(x_1, \dots, x_q, z_{q+1}, \dots, z_\ell) \in S^*$ . 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

$$\Pr_{z_{q+1}, \dots, z_\ell \in W} [\mathcal{E}] \geq 1 - \sum_{i=q}^{\ell-1} 2^{i-(k-r)} \geq 1 - 2^{\ell+r-k}.$$

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  $S$ .  $\square$

### 2.3 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 2.9.** For  $T_1, \dots, T_k \in \{0, 1\}^k$ , define  $\chi_{T_1, \dots, T_\ell} : (\{0, 1\}^k)^\ell \rightarrow \{-1, 1\}$  by

$$\chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell) = \prod_{i=1}^{\ell} \chi_{T_i}(x_i),$$

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  $\chi_{T_1, \dots, T_\ell}$ .

**Lemma 2.10.** *If  $\dim(\text{Span}(T_1, \dots, T_\ell)) = r$ , then  $\chi_{T_1, \dots, T_\ell}$  is a eigenvector of  $H_{k,\ell}$  with eigenvalue  $2^{-r}$ , i.e.*

$$H_{k,\ell} \cdot \chi_{T_1, \dots, T_\ell} = 2^{-r} \cdot \chi_{T_1, \dots, T_\ell}.$$

*Proof.* Considering a random choice of  $y \in \{0, 1\}^k$  and  $b = (b_1, \dots, b_\ell) \in \{0, 1\}^\ell$ ,

$$\begin{aligned} H_{k,\ell} \cdot \chi_{T_1, \dots, T_\ell}(x) &= \mathbb{E}_{y,b} [\chi_{T_1, \dots, T_\ell}(x_1 + b_1 y, \dots, x_\ell + b_\ell(y))] \\ &= \chi_{T_1, \dots, T_\ell}(x) \cdot \mathbb{E}_{y,b} [\chi_{\oplus_{i=1}^\ell b_i T_i}(y)]. \end{aligned}$$

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  $\text{Span}(T_1, \dots, T_\ell)$ , the probability over the choice of  $b$  that  $\oplus_{i=1}^\ell b_i T_i = 0$  is precisely  $2^{-r}$ .  $\square$

**Definition 2.11.** (Clearly) any function  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  can be written as

$$F[x_1, \dots, x_\ell] = \sum_{T_1, \dots, T_\ell \in \{0, 1\}^k} \widehat{F}(T_1, \dots, T_\ell) \cdot \chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell).$$

Its  $i^{\text{th}}$  level component is defined as its projection onto the eigenspace with eigenvalue  $2^{-i}$ , i.e.

$$F_{=i}[x_1, \dots, x_\ell] = \sum_{\substack{T_1, \dots, T_\ell \in \{0, 1\}^k \\ \dim(\text{Span}(T_1, \dots, T_\ell)) = i}} \widehat{F}(T_1, \dots, T_\ell) \cdot \chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell).$$

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 2.12.** A function  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  is called basis-invariant if for every  $x_1, \dots, x_\ell \in \{0, 1\}^k$  and an invertible  $\ell \times \ell$  matrix  $M$  over  $\mathbb{F}_2$ , we have that

$$F[x_1, \dots, x_\ell] = F[M(x_1, \dots, x_\ell)].$$

Here  $M(x_1, \dots, x_\ell) = (y_1, \dots, y_\ell)$  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 paper are basis invariant and in particular the indicators of sets  $S \subseteq H_{k,\ell}$  that “arise” from corresponding sets in  $\text{Gr}_{k,\ell}$ .

## 2.4 Pseudorandomness implies Low Weight at Low Levels implies 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  $i^{\text{th}}$  Fourier level”. We note that the weight at the  $0^{\text{th}}$  level is  $\delta^2$  and the sum of the weights at all Fourier levels equals  $\delta$ . Theorem 2.6 requires us to show that if  $S$  is pseudorandom, then it has near-perfect 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 a quick proof of the second step below for the sake of completeness. 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  $\langle F_1, F_2 \rangle = \mathbb{E}_x [F_1(x)F_2(x)]$ . We have

$$1 - \Phi(S) = \Pr_{x \in S, z \sim x} [z \in S] = (1/\delta) \cdot \Pr_{x, z \sim x} [x \in S \wedge z \in S] = (1/\delta) \cdot \langle F, \mathbb{H}_{k,\ell} F \rangle.$$

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 2.6, it suffices to prove (hence this is our main result):

**Theorem 2.13.** *Let  $S$  be a basis-invariant set of vertices in  $\mathbb{H}_{k,\ell}$  that has density  $\delta$  and is  $(r, \varepsilon)$  pseudo-random. Let  $F : \mathbb{H}_{k,\ell} = (\{0,1\}^k)^\ell \rightarrow \{0,1\}$  be the indicator function of  $S$ . Then for any  $i = 0, 1, \dots, r$ ,*

$$\eta = \|F_{=i}\|_2^2 \leq 2^{7r^3+3} \varepsilon^{\frac{1}{4}} \delta.$$

We now summarize the high-level plan to prove Theorem 2.13 as in [14]. 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  $(r, \varepsilon)$  pseudo-random 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  $2^9\delta^4$  on the right and then takes a fourth root)

$$\frac{\eta^5}{2^9 \cdot \delta^4} \leq \mathbb{E} [F_{=i}^4] \leq 2^{25r^3} \eta \varepsilon. \quad (2)$$

## 2.5 Lower-bounding the Fourth Moment of $F_{=i}$

**Lemma 2.14.** *Under the condition and notation of Theorem 2.13,  $\mathbb{E} [F_{=i}^4] \geq \frac{\eta^5}{2^9 \cdot \delta^4}$ .*

*Proof.* We note the decomposition  $F = \sum_{j=0}^{\ell} F_{=j}$  into mutually orthogonal components and that  $\|F\|_2^2 = \delta$ ,  $\|F_{=i}\|_2^2 = \eta$ . Hence  $\mathbb{E}[(F - F_{=i})^2] = \delta - \eta$ . By Markov's inequality,

$$\Pr\left[(F - F_{=i})^2 \geq 1 - \frac{\eta}{2\delta}\right] \leq \delta - \frac{\eta}{2}.$$

On the other hand,  $F$  is Boolean and  $\Pr[F = 1] = \delta$ . Thus with probability at least  $\frac{\eta}{2}$ , both the events below occur:

$$F = 1, \quad (F - F_{=i})^2 \leq 1 - \frac{\eta}{2\delta},$$

in which case it holds that  $(1 - F_{=i})^2 \leq 1 - \frac{\eta}{2\delta}$  and in turn that  $F_{=i} \geq \frac{\eta}{4\delta}$ . Hence as claimed,

$$\mathbb{E}[F_{=i}^4] \geq \frac{\eta}{2} \cdot \left(\frac{\eta}{4\delta}\right)^4.$$

□

## 2.6 Upper-bounding the Fourth Moment of $F_{=i}$

To summarize, the task of proving Theorem 2.13 is now reduced to proving the upper bound in Equation (2), i.e. under the condition and notation of Theorem 2.13, to prove that, for  $0 \leq i \leq r$ ,

$$\mathbb{E}[F_{=i}^4] \leq 2^{25r^3} \eta \varepsilon, \quad \eta \stackrel{def}{=} \mathbb{E}[F_{=i}^2].$$

Proving this upper bound is really the main result of this paper. We describe the first step of the proof below, take a lengthy detour in Section 3 to develop the required analytic machinery, and then return to the proof in Section 5. As shown in Section 3, Lemma 3.13,  $F_{=i}$  has an alternate characterization (in addition to that in Definition 2.11 and the two characterizations are related): there exists a (unique) function  $f_{=i} : (\{0, 1\}^k)^i \rightarrow \mathbb{R}$  such that for all  $x = (x_1, \dots, x_\ell) \in (\{0, 1\}^k)^\ell$ ,

$$F_{=i}[x] = \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i,\ell]} f_{=i}(Mx).$$

Let's explain the notation: here  $\mathcal{M}[i, \ell]$  is the set of all  $i \times \ell$  matrices over  $\mathbb{F}_2$  that have full row-rank  $i$ . For  $M \in \mathcal{M}[i, \ell]$ ,  $Mx \in (\{0, 1\}^k)^i$  is a  $i$ -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. To compute (or rather upper bound)  $\mathbb{E}[F_{=i}^4]$ , we simply take the sum to the fourth power, expand, and take the expectation over  $x$ :

$$\mathbb{E}[F_{=i}^4] = \frac{1}{\beta_{i,i}^4} \sum_{M_1, M_2, M_3, M_4 \in \mathcal{M}[i,\ell]} \mathbb{E}_x [f_{=i}(M_1x) f_{=i}(M_2x) f_{=i}(M_3x) f_{=i}(M_4x)].$$

We partition the sum according to the direct sum of row spaces of  $M_1, \dots, M_4$ , that is according to  $A = \bigoplus_{s=1}^4 \text{rowspan}(M_s)$ . We note that  $A \subseteq \{0, 1\}^\ell$  is a subspace and  $i \leq d = \dim(A) \leq 4i$ .

$$\mathbb{E}[F_{=i}^4] = \frac{1}{\beta_{i,i}^4} \sum_{d=i}^{4i} \sum_{A: \dim(A)=d} \sum_{\substack{M_1, M_2, M_3, M_4 \in \mathcal{M}[i,\ell] \\ \bigoplus_{s=1}^4 \text{rowspan}(M_s)=A}} \mathbb{E}_x [f_{=i}(M_1x) f_{=i}(M_2x) f_{=i}(M_3x) f_{=i}(M_4x)].$$

The main task is to upper bound each individual expectation above. A crude upper bound on the sum is taken thereafter. We note that the number of choices for  $A$  is at most  $2^{d\ell}$  (the number of  $d$ -dimensional subspaces of an  $\ell$ -dimensional space), for 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  $\frac{2^{7r^3} \eta \varepsilon}{2^{d\ell}}$  on each individual expectation. The main technical result in the paper is therefore:

**Theorem 2.15 (Main Technical Theorem).** *Let  $S$  be a basis-invariant set of vertices in  $\mathbb{H}_{k,\ell}$  that is  $(r, \varepsilon)$  pseudo-random. Let  $F: \mathbb{H}_{k,\ell} = (\{0, 1\}^k)^\ell \rightarrow \{0, 1\}$  be the indicator function of  $S$  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  $M_1, \dots, M_4 \in \mathcal{M}[i, \ell]$  such that  $\bigoplus_{s=1}^4 \text{rowspan}(M_s) = A$ , we have that*

$$\left| \mathbb{E}_{x \in (\{0,1\}^k)^\ell} [f_{=i}(M_1 x) f_{=i}(M_2 x) f_{=i}(M_3 x) f_{=i}(M_4 x)] \right| \leq 2^{7r^3} \frac{\eta \varepsilon}{2^{d\ell}}. \quad (3)$$

### 3 Analytic Machinery

In this section, we present the Fourier analytic machinery needed towards our main results. Unfortunately, we are unable to provide extra insight into the statements of various lemmas in addition to what may be inferred per se from their statements (but please do see Section 3.1 for a high-level picture). In terms of which of these lemmas are to be considered central and which ones more auxiliary in nature, we recommend that Lemmas 3.13, 3.19, 3.20 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,  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  is a basis-invariant function in the sense of Definition 2.12. Much of what is said applies to all such functions and not necessarily Boolean functions that are indicators of a basis-invariant set  $S \subseteq \mathbb{H}_{k,\ell} = (\{0, 1\}^k)^\ell$ . 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 2.11 of the Fourier representation and the decomposition into Fourier levels,  $F = \sum_{r=0}^\ell F_{=r}$ :

$$F[x_1, \dots, x_\ell] = \sum_{T_1, \dots, T_\ell \in \{0,1\}^k} \widehat{F}(T_1, \dots, T_\ell) \cdot \chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell).$$

$$F_{=r}[x_1, \dots, x_\ell] = \sum_{\substack{T_1, \dots, T_\ell \in \{0,1\}^k \\ \dim(\text{Span}(T_1, \dots, T_\ell)) = r}} \widehat{F}(T_1, \dots, T_\ell) \cdot \chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell).$$

**Lemma 3.1.**  $\widehat{F}(T_1, \dots, T_\ell)$  depends only on  $\text{Span}(T_1, \dots, T_\ell)$ .

*Proof.* Suppose  $\dim(\text{Span}(T_1, \dots, T_\ell)) = r$  and let  $A_1, \dots, A_r$  be a basis for the span. Then there is a  $r \times \ell$  matrix of row-rank  $r$  such that  $(T_1, \dots, T_\ell) = M^{\text{Tr}}(A_1, \dots, A_r)$ , where  $M^{\text{Tr}}$  is the  $\ell \times r$  transposed matrix. Moreover, in this case, defining vectors  $(y_1, \dots, y_r)$  such that  $(y_1, \dots, y_r) = M(x_1, \dots, x_\ell)$ ,

$$\prod_{j=1}^\ell \chi_{T_j}(x_j) = (-1)^{\bigoplus_{j=1}^\ell T_j \cdot x_j} = (-1)^{\bigoplus_{s=1}^r A_s \cdot y_s} = \prod_{s=1}^r \chi_{A_s}(y_s).$$

We extend  $M$  to a  $\ell \times \ell$  invertible matrix  $M'$  by appropriately appending  $\ell - r$  rows. Let  $(y_1, \dots, y_r, y_{r+1}, \dots, y_\ell) = M'(x_1, \dots, x_\ell)$ . It follows, using basis-invariance of  $F$ , that

$$\begin{aligned} \widehat{F}(T_1, \dots, T_\ell) &= \mathbb{E}_{x_1, \dots, x_\ell} \left[ F[x_1, \dots, x_\ell] \prod_{j=1}^{\ell} \chi_{T_j}(x_j) \right] \\ &= \mathbb{E}_{x_1, \dots, x_\ell} \left[ F[x_1, \dots, x_\ell] \prod_{s=1}^r \chi_{A_s}(y_s) \right] \\ &= \mathbb{E}_{y_1, \dots, y_\ell} \left[ F[y_1, \dots, y_\ell] \prod_{s=1}^r \chi_{A_s}(y_s) \right] \\ &= \widehat{F}(A_1, \dots, A_r, 0, \dots, 0). \end{aligned}$$

□

Thanks to this lemma, we write  $\widehat{F}(T_1, \dots, T_r)$  instead of  $\widehat{F}(T_1, \dots, T_\ell)$  if  $\dim(\text{Span}(T_1, \dots, T_\ell)) = r$  and the first  $r$  characters  $T_1, \dots, T_r$  are linearly independent.

### 3.1 High-level Picture

We 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 at 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,  $\|F_{=r}\|_2^2$ .

And then there is a third notion: as mentioned in Section 2.6,  $F_{=r}$  has an alternate characterization: there exists a (unique) function  $f_{=r} : (\{0, 1\}^k)^r \rightarrow \mathbb{R}$  such that for all  $x = (x_1, \dots, x_\ell) \in (\{0, 1\}^k)^\ell$ ,

$$F_{=r}[x] = \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i,\ell]} f_{=i}(Mx).$$

The functions  $f_{=r}$  will play a crucial role in our analysis. We will avoid giving them a name. While these functions do capture the zoom-in-out densities, unfortunately we do not have a good intuition as to how.

A large part of our Fourier analytic machinery is devoted to relating the three notions,  $\mu_{\text{in}(Q), \text{out}(W)}(S)$ ,  $F_{=r}$ ,  $f_{=r}$  to each other. 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 enter the picture only in an indirect fashion, as Fourier sums of  $f_{=r}$  (see Lemmas 3.19, 3.20). The relationship between  $f_{=r}$  and the zoom-in densities is somewhat clearer, see Definition 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 up to  $r$  points.

### 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 3.2.** Let  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  be a function. For a subspace  $W \subseteq \{0, 1\}^k$ , we define the function  $F_W: W^\ell \rightarrow \mathbb{R}$  to be the restriction of  $F$  to  $W^\ell$  (referred to as the zoom-out function).

**Definition 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 3.4.** Let  $A, T_1, \dots, T_r$  be linearly independent characters. Then

$$\widehat{F}(A, T_1, \dots, T_r) = \frac{1}{2^\ell - 2^r} \left( \widehat{F}_{W_A}(T_1, \dots, T_r) - \sum_{\substack{D \subseteq \text{Span}(A, T_1, \dots, T_r) \\ \dim(D)=r, A \notin D}} \widehat{F}(D) \right).$$

*Proof.* The computation proceeds as below where 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  $\mathbb{E}_{y \in W_A} [\chi_R(y)]$  equals 1 when  $R = 0$  or when  $R = A$  and vanishes otherwise.

$$\begin{aligned} \widehat{F}_{W_A}(T_1, \dots, T_r) &= \mathbb{E}_{x_1, \dots, x_\ell \in W_A} [F_{W_A}(x_1, \dots, x_\ell) \cdot \chi_{T_1, \dots, T_r}(x_1, \dots, x_r)] \\ &= \mathbb{E}_{x_1, \dots, x_\ell \in W_A} \left[ F(x_1, \dots, x_\ell) \cdot \prod_{i=1}^r \chi_{T_i}(x_i) \right] \\ &= \mathbb{E}_{x_1, \dots, x_\ell \in W_A} \left[ \sum_{Q_1, \dots, Q_\ell} \widehat{F}(Q_1, \dots, Q_\ell) \prod_{i=1}^r \chi_{T_i \oplus Q_i}(x_i) \cdot \prod_{j=r+1}^{\ell} \chi_{Q_j}(x_j) \right] \\ &= \sum_{1 \leq i \leq r: Q_i \in \{T_i, T_i \oplus A\}} \sum_{r+1 \leq j \leq \ell: Q_j \in \{0, A\}} \widehat{F}(Q_1, \dots, Q_\ell) \\ &= (2^\ell - 2^r) \widehat{F}(A, T_1, \dots, T_r) + \sum_{\substack{D \subseteq \text{Span}(A, T_1, \dots, T_r) \\ \dim(D)=r, A \notin D}} \widehat{F}(D). \end{aligned}$$

The last equality is justified as follows. There are  $2^\ell$  terms in the summation which 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  $\text{Span}(Q_1, \dots, Q_\ell)$  is same as  $\text{Span}(A, T_1, \dots, T_r)$  and since the Fourier coefficients depend only on this span,  $\widehat{F}(Q_1, \dots, Q_\ell) = \widehat{F}(A, T_1, \dots, T_r)$ .
- For the remaining  $2^r$  terms, for all  $j \geq r + 1$ ,  $Q_j = 0$ . In this case,  $\text{Span}(Q_1, \dots, Q_\ell)$  is same as  $\text{Span}(Q_1, \dots, Q_r)$  which is an  $r$ -dimensional subspace of  $\text{Span}(A, T_1, \dots, T_r)$  that does not contain  $A$ . Moreover each subspace of this kind is counted exactly once.

□

### 3.3 Defining $f_{=r}$ and Relating $F_{=r}$ , $f_{=r}$ and Zoom-in Densities

For a (basis-invariant) function  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$ , we have the decomposition  $F = \sum_{r=0}^\ell F_{=r}$  where

$$F_{=r}[x_1, \dots, x_\ell] = \sum_{\substack{T_1, \dots, T_\ell \subseteq [k] \\ \dim(\text{Span}(T_1, \dots, T_\ell))=r}} \widehat{F}(T_1, \dots, T_\ell) \cdot \chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell).$$

As mentioned, we will need an alternate formula for  $F_{=r} : (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  in terms of related functions  $f_{=r} : (\{0, 1\}^k)^r \rightarrow \mathbb{R}$ . Deriving this formula turns out to be rather cumbersome (but quite interesting at the same time). Next few subsections are devoted to this derivation. Sometimes we write  $f_{=r, F}$  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  $|\mathcal{M}[i, r]| = \prod_{j=0}^{i-1} (2^r - 2^j) \stackrel{\text{def}}{=} \beta_{i, r}$ .
- For  $r \geq 0$ , we will pretend that  $\beta_{0, r} = 1$  and that there is a single matrix  $\{0\}$  in  $\mathcal{M}[0, r]$ .
- For  $x = (x_1, \dots, x_r)$  and  $M \in \mathcal{M}[i, r]$ ,  $Mx$  denotes the tuple  $(y_1, \dots, y_i)$  where  $y_j = \sum_{t=1}^r M_{jt}x_t$ .

#### Defining $f_{=r}$ in terms of Zoom-in Densities

**Definition 3.5.** For  $0 \leq r \leq \ell$ , define  $f_{=r} : (\{0, 1\}^k)^r \rightarrow \mathbb{R}$  inductively as

$$f_{=0}(\{0\}) = \mu(F) \stackrel{\text{def}}{=} \mathbb{E}_{x_1, \dots, x_\ell} [F[x_1, \dots, x_\ell]],$$

$$f_{=r}(x_1, \dots, x_r) = \mu_{\text{in}(\{x_1, \dots, x_r\})}(F) - \sum_{d=0}^{r-1} \frac{1}{\beta_{d, d}} \sum_{M \in \mathcal{M}[d, r]} f_{=d}(Mx).$$

We note that

- $\mu_{\text{in}(\{x_1, \dots, x_r\})}(F) = \mathbb{E}_{z_{r+1}, \dots, z_\ell} [F[x_1, \dots, x_r, z_{r+1}, \dots, z_\ell]]$  is the zoom-in density.
- The term corresponding to  $d = 0$  in the summation above equals  $\mu(F)$ .
- In the case  $r = 1$ ,

$$f_{=1}(x_1) = \mu_{\text{in}(\{x_1\})}(F) - \mu(F).$$

**Lemma 3.6.** The function  $f_{=r}$  is basis-invariant, i.e. for every  $x = (x_1, \dots, x_r) \in (\{0, 1\}^k)^r$  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  $x_1, \dots, x_r$  and an  $r \times r$  invertible matrix  $M$  as in the claim. By definition

$$f_{=r}(Mx) = \mu_{\text{in}(Mx)}(F) - \sum_{d=0}^{r-1} \frac{1}{\beta_{d, d}} \sum_{M' \in \mathcal{M}[d, r]} f_{=d}(M'Mx).$$

Observe that for any  $d \geq 0$ , the mapping  $M' \rightarrow M'M$  is a bijection on  $\mathcal{M}[d, r]$ . Also observe that  $\mu_{\text{in}(Mx)}(F) = \mu_{\text{in}(x)}(F)$ , since  $F$  is basis invariant. Thus the last expression equals

$$\mu_{\text{in}(x)}(F) - \sum_{d=0}^{r-1} \frac{1}{\beta_{d,d}} \sum_{M' \in \mathcal{M}[d,r]} f_{=d}(M'x) = f_{=r}(x).$$

□

### Zoom-in Restriction Lemma

**Definition 3.7.** Let  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  be a function and let  $Q = \{a_1, \dots, a_j\} \subseteq \{0, 1\}^k$  where  $j \leq \ell - 1$ . We define the function  $F_Q: (\{0, 1\}^k)^{\ell-j} \rightarrow \mathbb{R}$  (the zoom-in restriction function) by

$$F_Q[x_1, \dots, x_{\ell-j}] = F[a_1, \dots, a_j, x_1, \dots, x_{\ell-j}].$$

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 3.8.** Let  $F: (\{0, 1\}^k)^\ell \rightarrow \mathbb{R}$  and  $r \geq 0$ . Let  $D = (a, x_1, \dots, x_r) = (a, x)$ . Then

$$f_{=r+1, F}(D) = f_{=r, F_{\{a\}}}(x) - \frac{1}{\beta_{r,r}} \sum_{\substack{M' \in \mathcal{M}[r, r+1] \\ e_1 \notin \text{rowspan}(M')}} f_{=r, F}(M'D).$$

*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 convenience, we write  $f_{=r}$  instead of  $f_{=r, F}$ , but do write  $f_{=r, F_a}$  when it is the zoom-in function that is concerned. From Definition 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

$$f_{=r+1}(D) = \mu_{\text{in}(x)}(F_a) - \mu(F) - \sum_{j=1}^r \frac{1}{\beta_{j,j}} \left( \sum_{\substack{M^* \in \mathcal{M}[j, r+1] \\ e_1 \in \text{rowspan}(M^*)}} f_{=j}(M^*D) + \sum_{\substack{M^* \in \mathcal{M}[j, r+1] \\ e_1 \notin \text{rowspan}(M^*)}} f_{=j}(M^*D) \right). \quad (4)$$

For fixed  $j$ , let's call the two terms above  $\Gamma_j$  and  $\Delta_j$  respectively. Below, computation of  $\Gamma_j$  results in an "extra"  $-\Delta_{j-1}$  term that cancels with the previous  $\Delta$ -term in a telescoping manner.

$$\begin{aligned} \Gamma_j &= \frac{1}{\beta_{j,j}} \sum_{\substack{M^* \in \mathcal{M}[j, r+1] \\ e_1 \in \text{rowspan}(M^*)}} f_{=j}(M^*D) = \frac{\beta_{j, r+1}}{\beta_{j,j}} \frac{2^j - 1}{2^{r+1} - 1} \cdot \mathbb{E}_{\substack{M^* \in \mathcal{M}[j, r+1] \\ e_1 \in \text{rowspan}(M^*)}} [f_{=j}(M^*D)] \\ &= \frac{\beta_{j-1, r}}{\beta_{j-1, j-1}} \cdot \mathbb{E}_{\substack{M^* \in \mathcal{M}[j, r+1] \\ e_1 \in \text{rowspan}(M^*)}} [f_{=j}(M^*D)], \end{aligned} \quad (5)$$

where we replaced summation by expectation for the sake of convenience 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  $\frac{2^j-1}{2^{r+1}-1}$  of them will contain  $e_1$  in their row-span (all non-zero vectors being symmetric in this regard).

Using Lemma 3.9, we see that  $M^*$  can be sampled by sampling  $M \in \mathcal{M}[j-1, r]$ , constructing  $M'$  from  $M$ , sampling  $M''$ , and then letting  $M^* = M'' \cdot M'$ . Since  $M''$  is invertible and  $f_{=j}$  is basis invariant,

$$f_{=j}(M^*D) = f_{=j}(M''M'D) = f_{=j}(M'D) = f_{=j}(M'(a, x)) = f_{=j}(a, Mx).$$

Hence the expectation in (5) is same as  $\mathbb{E}_M[f_{=j}(a, Mx)]$ . Applying the induction hypothesis (note that  $Mx$  is a  $(j-1)$ -tuple):

$$\begin{aligned} & \mathbb{E}_{M \in \mathcal{M}[j-1, r]} \left[ f_{=j-1, F_a}(Mx) - \frac{1}{\beta_{j-1, j-1}} \sum_{\substack{M'' \in \mathcal{M}[j-1, j] \\ e_1 \notin \text{rowspan}(M'')}} f_{=j-1}(M''(a, Mx)) \right] \\ &= \mathbb{E}_{M \in \mathcal{M}[j-1, r]} [f_{=j-1, F_a}(Mx)] - \frac{\beta_{j-1, j}}{\beta_{j-1, j-1}} \frac{2^j - 2^{j-1}}{2^j - 1} \mathbb{E}_{M \in \mathcal{M}[j-1, r]} \left[ \mathbb{E}_{\substack{M'' \in \mathcal{M}[j-1, j] \\ e_1 \notin \text{rowspan}(M'')}} [f_{=j-1}(M''(a, Mx))] \right], \end{aligned}$$

where the normalization factor is justified as before. Using Lemma 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

$$f_{=j-1}(M''(a, Mx)) = f_{=j-1}(M''M'(a, x)) = f_{=j-1}(M^*D).$$

Using the definition of the  $\beta$ -parameters, the expression can be re-written as

$$\mathbb{E}_{M \in \mathcal{M}[j-1, r]} [f_{=j-1, F_a}(Mx)] - 2^{j-1} \mathbb{E}_{\substack{M^* \in \mathcal{M}[j-1, r+1] \\ e_1 \notin \text{rowspan}(M^*)}} [f_{=j-1}(M^*D)].$$

Substituting in (5), we get

$$\begin{aligned} \Gamma_j &= \frac{\beta_{j-1, r}}{\beta_{j-1, j-1}} \mathbb{E}_{M \in \mathcal{M}[j-1, r]} [f_{=j-1, F_a}(Mx)] - \frac{\beta_{j-1, r}}{\beta_{j-1, j-1}} \cdot 2^{j-1} \mathbb{E}_{\substack{M^* \in \mathcal{M}[j-1, r+1] \\ e_1 \notin \text{rowspan}(M^*)}} [f_{=j-1}(M^*D)] \\ &= \frac{1}{\beta_{j-1, j-1}} \sum_{M \in \mathcal{M}[j-1, r]} f_{=j-1, F_a}(Mx) - \frac{1}{\beta_{j-1, j-1}} \sum_{\substack{M^* \in \mathcal{M}[j-1, r+1] \\ e_1 \notin \text{rowspan}(M^*)}} f_{=j-1}(M^*D) \\ &= \frac{1}{\beta_{j-1, j-1}} \sum_{M \in \mathcal{M}[j-1, r]} f_{=j-1, F_a}(Mx) - \Delta_{j-1}. \end{aligned}$$

Substituting in (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

$$\begin{aligned} f_{=r+1}(D) &= \left( \mu_{\text{in}(x)}(F_a) - \mu(F) + \Delta_0 - \sum_{j=0}^{r-1} \frac{1}{\beta_{j, j}} \sum_{M \in \mathcal{M}[j, r]} f_{=j, F_a}(Mx) \right) - \Delta_r \\ &= f_{=r, F_{\{a\}}}(x) - \Delta_r, \end{aligned}$$

completing the proof. □

## Some Auxiliary Lemmas

**Lemma 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 first column and the first row are 0 and deleting the first column and the first row 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  $L' \subseteq \mathbb{F}_2^{r+1}$  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  $L'$ . Thus it follows that the row-span of  $M'$  is a random  $j$ -dimensional subspace of  $\mathbb{F}_2^{r+1}$  containing  $e_1$ . Now pre-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.  $\square$

**Lemma 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):*

- *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 first column and the first row are 0 and deleting the first column and the first row yields  $M$ .*
- *Pick a matrix  $M''$  that is uniformly distributed over matrices in  $\mathcal{M}[j - 1, j]$  whose row-span does not contain  $e_1$  and output  $M^* = M'' \cdot M'$ .*

*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  $v_2, \dots, v_j$  of  $L$ . From Lemma 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  $v_1 = e_1$ .  $\square$

**Lemma 3.11.** *Let  $v_1, \dots, v_j$  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*

$$(w_1, \dots, w_{j-1}) = M'' \cdot (v_1, \dots, v_j).$$

*Then  $(w_1, \dots, w_{j-1})$  is a random basis of a random  $(j - 1)$ -dimensional subspace of  $\text{Span}(v_1, \dots, v_j)$  that does not contain  $v_1$ .*

*Proof.* It is clear that

- Since the rows of  $M''$  are linearly independent, so are  $w_1, \dots, w_{j-1}$ .
- The matrix  $M''$  and the tuple  $(w_1, \dots, w_{j-1})$  determine each other.
- $e_1 \notin \text{rowspan}(M'')$  if and only if  $v_1 \notin \text{Span}(w_1, \dots, w_{j-1})$ .

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  $(w_1, \dots, w_{j-1})$  that span a  $(j - 1)$ -dimensional subspace of  $\text{Span}(v_1, \dots, v_j)$  that does not contain the vector  $v_1$ .  $\square$

## Relating $F_{=r}$ to Zoom-in Densities

**Lemma 3.12.** For every  $0 \leq r \leq \ell$  and  $x_1, \dots, x_\ell \in \{0, 1\}^k$ ,

$$\left( \sum_{i=0}^r \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,\ell}} F_{=i} \right) [x_1, \dots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{\substack{M \in \mathcal{M}[r,\ell] \\ y=Mx}} \mu_{\text{in}((y_1, \dots, y_r))}(F).$$

*Proof.* By definition,

$$\begin{aligned} F_{=i}[x_1, \dots, x_\ell] &= \sum_{\substack{T_1, \dots, T_\ell \\ \dim(T_1, \dots, T_\ell) = i}} \widehat{F}(T_1, \dots, T_\ell) \cdot \chi_{T_1, \dots, T_\ell}(x_1, \dots, x_\ell) \\ &= \sum_{\dim(D)=i} \widehat{F}(Q_1, \dots, Q_i, Q_{i+1}, \dots, Q_\ell) \sum_{\substack{T_1, \dots, T_\ell \\ \text{Span}(T_1, \dots, T_\ell) = D}} \prod_{j=1}^{\ell} \chi_{T_j}(x_j), \end{aligned}$$

where the outer summation is over all  $i$ -dimensional subspaces  $D$  and for given  $D$ ,  $(Q_1, \dots, Q_i)$  is a specific ordered basis for it, and  $Q_{i+1} = \dots = Q_\ell = 0$ . 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  $(T_1, \dots, T_\ell)$  such that  $\text{Span}(T_1, \dots, T_\ell) = D$  are obtained precisely as

$$(T_1, \dots, T_\ell) = M^{\text{Tr}}(Q_1, \dots, Q_i)$$

where  $M^{\text{Tr}}$  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  $(y_1, \dots, y_i)$  such that  $(y_1, \dots, y_i) = M(x_1, \dots, x_\ell)$  (which we abbreviate as  $y = Mx$ )

$$\prod_{j=1}^{\ell} \chi_{T_j}(x_j) = (-1)^{\oplus_{j=1}^{\ell} T_j \cdot x_j} = (-1)^{\oplus_{s=1}^i Q_s \cdot y_s} = \prod_{s=1}^i \chi_{Q_s}(y_s).$$

Thus we can write

$$F_{=i}[x_1, \dots, x_\ell] = \sum_{\dim(D)=i} \widehat{F}(Q_1, \dots, Q_i, Q_{i+1}, \dots, Q_\ell) \sum_{\substack{M \in \mathcal{M}[i,\ell] \\ y=Mx}} \prod_{j=1}^i \chi_{Q_j}(y_j).$$

Using the definition of  $\widehat{F}(Q_1, \dots, Q_i, Q_{i+1} = 0, \dots, Q_\ell = 0)$  and interchanging the order of summation,

$$\begin{aligned} F_{=i}[x_1, \dots, x_\ell] &= \sum_{\dim(D)=i} \mathbb{E}_{z_1, \dots, z_\ell} \left[ F(z_1, \dots, z_\ell) \prod_{j=1}^i \chi_{Q_j}(z_j) \right] \cdot \sum_{\substack{M \in \mathcal{M}[i,\ell] \\ y=Mx}} \prod_{j=1}^i \chi_{Q_j}(y_j) \\ &= \sum_{\substack{M \in \mathcal{M}[i,\ell] \\ y=Mx}} \mathbb{E}_{z_1, \dots, z_\ell} \left[ F(z_1, \dots, z_\ell) \sum_{\dim(D)=i} \prod_{j=1}^i \chi_{Q_j}(y_j \oplus z_j) \right] \\ &= \frac{\beta_{i,\ell}}{\beta_{r,\ell}} \sum_{\substack{M \in \mathcal{M}[r,\ell] \\ y=Mx}} \mathbb{E}_{z_1, \dots, z_\ell} \left[ F(z_1, \dots, z_\ell) \sum_{\dim(D)=i} \prod_{j=1}^i \chi_{Q_j}(y_j \oplus z_j) \right]. \end{aligned}$$

Note that in the last line, the summation is over  $r \times \ell$  matrices instead of  $i \times \ell$  matrices and out of the vectors  $(y_1, \dots, y_r)$ , 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  $r \times r$  invertible matrix  $M'$ , consider the change of basis  $y = M'y'$ ,  $(z_1, \dots, z_r) = M'(z'_1, \dots, z'_r)$  and  $(A_1, \dots, A_r) = M'^{\text{Tr}}(Q_1, \dots, Q_i, Q_{i+1} = 0, \dots, Q_r = 0)$ . 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

$$F_{=i}[x_1, \dots, x_\ell] = \frac{\beta_{i,\ell}}{\beta_{r,\ell}} \sum_{\substack{M \in \mathcal{M}[r,\ell] \\ y=Ms}} \mathbb{E}_{z_1, \dots, z_\ell} \left[ F(z_1, \dots, z_\ell) \sum_{\dim(D)=i} \mathbb{E}_{M'} \left[ \prod_{j=1}^r \chi_{A_j}(y_j \oplus z_j) \right] \right].$$

In the above expression, first an  $i$ -dimensional subspace  $D$  is chosen along with a fixed ordered basis  $Q_1, \dots, Q_i$  and then  $(A_1, \dots, A_r) = M'^{\text{Tr}}(Q_1, \dots, Q_i, Q_{i+1} = 0, \dots, Q_r = 0)$  for a random  $r \times r$  invertible matrix  $M'$ . Up to a factor of  $\frac{1}{\beta_{r,r}}$ , one can instead consider a summation over all  $M'$ , and then every tuple  $(A_1, \dots, A_r)$  such that  $\dim(A_1, \dots, A_r) = i$  occurs exactly  $\frac{\beta_{r,r}}{\beta_{i,r}}$  times. Hence the above equation can be written as

$$F_{=i}[x_1, \dots, x_\ell] = \frac{\beta_{i,\ell}}{\beta_{r,\ell}} \frac{1}{\beta_{i,r}} \sum_{\substack{M \in \mathcal{M}[r,\ell] \\ y=Ms}} \mathbb{E}_{z_1, \dots, z_\ell} \left[ F(z_1, \dots, z_\ell) \sum_{\dim(A_1, \dots, A_r)=i} \prod_{j=1}^r \chi_{A_j}(y_j \oplus z_j) \right].$$

Moving the  $\beta$ -factors to the left hand side and summing over  $i = 0, 1, \dots, r$  counts every  $r$ -tuple  $(A_1, \dots, A_r)$  exactly once (irrespective of its dimension). Hence

$$\left( \sum_{i=0}^r \beta_{i,r} \frac{\beta_{r,\ell}}{\beta_{i,\ell}} F_{=i} \right) [x_1, \dots, x_\ell] = \sum_{\substack{M \in \mathcal{M}[r,\ell] \\ y=Ms}} \mathbb{E}_{z_1, \dots, z_\ell} \left[ F(z_1, \dots, z_\ell) \sum_{A_1, \dots, A_r} \prod_{j=1}^r \chi_{A_j}(y_j \oplus z_j) \right].$$

We observe finally that the inner summation 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

$$\left( \sum_{i=0}^r \beta_{i,r} \frac{\beta_{r,\ell}}{\beta_{i,\ell}} F_{=i} \right) [x_1, \dots, x_\ell] = \sum_{\substack{M \in \mathcal{M}[r,\ell] \\ y=Ms}} \mathbb{E}_{z_{r+1}, \dots, z_\ell} [F(y_1, \dots, y_r, z_{r+1}, z_\ell)].$$

The proof of Lemma 3.12 is completed by dividing both sides by  $\beta_{r,r}$  and noting that the expectation is precisely  $\mu_{\text{in}(y=Ms)}(F)$ .  $\square$

### Relating $F_{=r}$ and $f_{=r}$

**Lemma 3.13.** For every  $0 \leq r \leq \ell$  and  $x_1, \dots, x_\ell \in \{0, 1\}^k$ ,

$$F_{=r}[x_1, \dots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} f_{=r}(Mx).$$

*Proof.* The proof is by induction on  $r$ . The case  $r = 0$  is trivial (both sides equal  $\mu(F)$ ). Otherwise, using Lemma 3.12 we have

$$\left( \sum_{i=0}^r \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,\ell}} F_{=i} \right) [x_1, \dots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} \mu_{\text{in}((Mx))}(F).$$

We note that the coefficient of  $F_{=r}$  on the left side is 1. Therefore we get

$$F_{=r}[x_1, \dots, x_\ell] = \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} \mu_{\text{in}((Mx))}(F) - \left( \sum_{i=0}^{r-1} \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,\ell}} F_{=i} \right) [x_1, \dots, x_\ell]. \quad (6)$$

Using the induction hypothesis, we get that

$$\begin{aligned} \left( \sum_{i=0}^{r-1} \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,\ell}} F_{=i} \right) [x_1, \dots, x_\ell] &= \sum_{i=0}^{r-1} \frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{r,r} \beta_{i,\ell}} \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i, \ell]} f_{=i}(Mx) \\ &= \frac{1}{\beta_{r,r}} \sum_{A \in \mathcal{M}[r, \ell]} \sum_{i=0}^{r-1} \frac{1}{\beta_{i,i}} \sum_{Q \in \mathcal{M}[i, r]} f_{=i}(QAx). \end{aligned} \quad (7)$$

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  $\frac{\beta_{i,r} \beta_{r,\ell}}{\beta_{i,\ell}}$  times. Substituting (7) into (6) yields

$$\begin{aligned} F_{=r}[x_1, \dots, x_\ell] &= \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} \mu_{\text{in}((Mx))}(F) - \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} \sum_{i=0}^{r-1} \frac{1}{\beta_{i,i}} \sum_{Q \in \mathcal{M}[i, r]} f_{=i}(QMx) \\ &= \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} f_{=r}(Mx), \end{aligned}$$

where in the last equality we combined the two sums over  $M$ , and used Definition 3.5 of  $f_{=r}$ , thereby finishing the inductive step.  $\square$

### 3.4 Relating $F_{=r}$ and $f_{=r}$ in the Fourier Domain

**Lemma 3.14.** Let  $0 \leq j \leq r - 1$  and let  $a_1, \dots, a_j \in \{0, 1\}^k$ . Then

$$\mathbb{E}_{y_{j+1}, \dots, y_r \in \{0, 1\}^k} [f_{=r}(a_1, \dots, a_j, y_{j+1}, \dots, y_r)] = 0.$$

*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  $r = 1$ ,

$$\mathbb{E}_{y_1} [f_{=1}(y_1)] = \mathbb{E}_{y_1} [\mu_{\text{in}(y_1)}(F) - \mu(F)] = 0.$$

Now assume  $j = 0$  and  $r \geq 2$ .

$$\begin{aligned} \mathbb{E}_{y_1, \dots, y_r} [f_{=r}(y_1, \dots, y_r)] &= \mathbb{E}_{y_1, \dots, y_r} \left[ \mu_{\text{in}(y_1, \dots, y_r)}(F) - \sum_{i=0}^{r-1} \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i,r]} f_{=i}(My) \right] \\ &= - \sum_{i=1}^{r-1} \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i,r]} \mathbb{E}_{y_1, \dots, y_r} [f_{=i}(My)], \end{aligned}$$

where we used the fact that in the summation, the term with index  $i = 0$  is  $\mu(F)$  and

$$\mathbb{E}_{y_1, \dots, y_r} [\mu_{\text{in}(y_1, \dots, y_r)}(F)] = \mu(F)$$

as well. We note that for any  $1 \leq i \leq r - 1$  and  $M \in \mathcal{M}[i, r]$ , the distribution of  $My$  is uniform over  $(\{0, 1\}^k)^i$  and hence by the induction hypothesis

$$\mathbb{E}_{y_1, \dots, y_r} [f_{=i}(My)] = 0,$$

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 3.8 we see

$$\begin{aligned} &\mathbb{E}_{y_{j+1}, \dots, y_r} [f_{=r}(a_1, \dots, a_j, y_{j+1}, \dots, y_r)] \\ &= \mathbb{E}_{y_{j+1}, \dots, y_r} \left[ f_{=r-1, F_{\{a_1\}}}(a_2, \dots, a_j, y_{j+1}, \dots, y_r) - \frac{1}{\beta_{r-1, r-1}} \sum_{\substack{M \in \mathcal{M}[r-1, r] \\ e_1 \notin \text{rowspan}(M)}} f_{=r-1}(M(a, y)) \right]. \quad (8) \end{aligned}$$

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 3.15 below, we may assume that  $M$  is semi-diagonal. Hence

$$M(a, y) = M(a_1, \dots, a_j, y_{j+1}, \dots, y_r) = (a'_2, \dots, a'_j, y'_{j+1}, \dots, y'_r),$$

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

$$\mathbb{E}_{y_{j+1}, \dots, y_r} [f_{=r-1}(M(a, y))] = \mathbb{E}_{y'_{j+1}, \dots, y'_r} [f_{=r-1}(a'_2, \dots, a'_j, y'_{j+1}, \dots, y'_r)] = 0,$$

completing the proof. □

**Lemma 3.15.** *A  $(r - 1) \times r$  matrix is called semi-diagonal if deleting the first column gives a square matrix that is diagonal and has ones 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  $W_2 \setminus \{e_1\}$ ,  $W_3 \setminus W_2$ ,  $\dots$ ,  $W_r \setminus W_{r-1}$  respectively. Moreover, in this process, when we choose a vector  $v_s \in W_s \setminus W_{s-1}$ , the  $s^{\text{th}}$  coordinate of  $v$  equals one, and we can zero-out its coordinates  $2, \dots, s - 1$  by adding to it  $v_2, \dots, v_{s-1}$  if necessary.  $\square$

We now consider the Fourier representation of  $f_{=r} : (\{0, 1\}^k)^r \rightarrow \mathbb{R}$ :

$$f_{=r}(x_1, \dots, x_r) = \sum_{T_1, \dots, T_r} \widehat{f}_{=r}(T_1, \dots, T_r) \chi_{T_1}(x_1) \cdots \chi_{T_r}(x_r).$$

**Lemma 3.16.**  $\widehat{f}_{=r}(T_1, \dots, T_r)$  depends only on  $\text{Span}(T_1, \dots, T_r)$ .

*Proof.* Follows from the basis-invariance of  $f_{=r}$  (Lemma 3.6) and a proof identical to that of Lemma 3.1.  $\square$

**Lemma 3.17.** Suppose  $T_1, \dots, T_r$  are linearly dependent. Then  $\widehat{f}_{=r}(T_1, \dots, T_r) = 0$ .

*Proof.* Suppose w.l.o.g. that  $T_r$  depends on  $T_1, \dots, T_{r-1}$ . By Lemma 3.16,

$$\begin{aligned} \widehat{f}_{=r}(T_1, \dots, T_r) &= \widehat{f}_{=r}(T_1, \dots, T_{r-1}, 0) \\ &= \mathbb{E}_{y_1, \dots, y_r} \left[ f_{=r}(y_1, \dots, y_r) \prod_{j=1}^{r-1} \chi_{T_j}(y_j) \right] \\ &= \mathbb{E}_{y_1, \dots, y_{r-1}} \left[ \mathbb{E}_{y_r} [f_{=r}(y_1, \dots, y_{r-1}, y_r)] \prod_{j=1}^{r-1} \chi_{T_j}(y_j) \right], \end{aligned}$$

and the inner expectation vanishes according to Lemma 3.14.  $\square$

Therefore, we may write

$$f_{=r}(y_1, \dots, y_r) = \sum_{\substack{T_1, \dots, T_r \\ \dim(T_1, \dots, T_r) = r}} \widehat{f}_{=r}(T_1, \dots, T_r) \chi_{T_1}(y_1) \cdots \chi_{T_r}(y_r). \quad (9)$$

**Lemma 3.18.** Let  $0 \leq r \leq \ell$  and let  $T_1, \dots, T_r$  be characters such that  $\dim(T_1, \dots, T_r) = r$ . Then

$$\widehat{f}_{=r}(T_1, \dots, T_r) = \widehat{F}(T_1, \dots, T_r).$$

*Proof.* For any  $x_1, \dots, x_\ell \in \{0, 1\}^k$ , by Lemma 3.13,

$$\begin{aligned} F_{=r}[x_1, \dots, x_\ell] &= \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} f_{=r}(Mx) \\ &= \frac{1}{\beta_{r,r}} \sum_{M \in \mathcal{M}[r, \ell]} \sum_{\substack{T_1, \dots, T_r \\ \dim(T_1, \dots, T_r) = r}} \widehat{f}_{=r}(T_1, \dots, T_r) \prod_{j=1}^r \chi_{T_j}((Mx)_j) \\ &= \frac{1}{\beta_{r,r}} \sum_{\substack{T_1, \dots, T_r \\ \dim(T_1, \dots, T_r) = r}} \widehat{f}_{=r}(T_1, \dots, T_r) \sum_{M \in \mathcal{M}[r, \ell]} \prod_{j=1}^r \chi_{T_j}((Mx)_j). \end{aligned}$$

As we have done previously, if  $T = (T_1, \dots, T_r)$  and  $M^{\text{Tr}}$  is the transposed matrix, we have

$$\prod_{j=1}^r \chi_{T_j}((Mx)_j) = \prod_{i=1}^{\ell} \chi_{(M^{\text{Tr}}T)_i}(x_i).$$

Hence,

$$\begin{aligned} F_{=r}[x_1, \dots, x_{\ell}] &= \frac{1}{\beta_{r,r}} \sum_{\substack{T_1, \dots, T_r \\ \dim(T_1, \dots, T_r) = r}} \widehat{f}_{=r}(T_1, \dots, T_r) \sum_{M \in \mathcal{M}[r, \ell]} \prod_{i=1}^{\ell} \chi_{(M^{\text{Tr}}T)_i}(x_i) \\ &= \sum_{\substack{P_1, \dots, P_{\ell} \\ \dim(P_1, \dots, P_{\ell}) = r}} \widehat{f}_{=r}(\text{basis}(P_1, \dots, P_{\ell})) \prod_{i=1}^{\ell} \chi_{P_i}(x_i), \end{aligned}$$

where it is easily checked that each tuple  $(P_1, \dots, P_{\ell})$  with dimension of the span  $r$  is counted exactly once. On the other hand, by definition

$$F_{=r}[x_1, \dots, x_{\ell}] = \sum_{\substack{P_1, \dots, P_{\ell} \\ \dim(P_1, \dots, P_{\ell}) = r}} \widehat{F}(\text{basis}(P_1, \dots, P_{\ell})) \prod_{i=1}^{\ell} \chi_{P_i}(x_i).$$

By uniqueness of Fourier representation, we conclude the assertion of the lemma.  $\square$

### 3.5 Bounding Restricted Fourier Sums of $f_{=r}$

**Lemma 3.19.** *Let  $S$  be a (basis invariant) set of vertices in  $H_{k, \ell}$  that is  $(r, \varepsilon)$  pseudo-random. Let  $0 \leq j \leq r \leq \frac{\ell}{2}$  and  $F: (\{0, 1\}^k)^{\ell} \rightarrow \{0, 1\}$  be the indicator function of  $S$ . Then for any characters  $A_1, \dots, A_j$ ,*

$$\sum_{T_{j+1}, \dots, T_r} \widehat{f}_{=r}^2(A_1, \dots, A_j, T_{j+1}, \dots, T_r) \leq \frac{2^{4r^2}}{2^{(r+j)\ell}} \varepsilon.$$

*Proof.* We will prove an upper bound of  $\frac{C_r}{2^{(r+j)\ell}} \varepsilon$  with  $C_r = 2^{4r^2}$ . We note first that a  $(r, \varepsilon)$ -pseudorandom set automatically has density 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)

$$\begin{aligned} \sum_{\substack{T_1, \dots, T_r \\ \dim(T_1, \dots, T_r) = r}} \widehat{f}_{=r}^2(T_1, \dots, T_r) &= \sum_{\substack{T_1, \dots, T_r \\ \dim(T_1, \dots, T_r) = r}} \widehat{F}^2(T_1, \dots, T_r) \\ &= \frac{\beta_{r,r}}{\beta_{r,\ell}} \sum_{\substack{Q_1, \dots, Q_{\ell} \\ \dim(Q_1, \dots, Q_{\ell}) = r}} \widehat{F}^2(Q_1, \dots, Q_{\ell}), \end{aligned}$$

since for a fixed  $r$ -dimensional span of the arguments, there are  $\beta_{r,r}$  terms  $(T_1, \dots, T_r)$  in the first summation and  $\beta_{r,\ell}$  terms  $(Q_1, \dots, Q_{\ell})$  in the second summation. The expression now equals and is then upper bounded as ( $C_r = 2^{4r^2}$ ),

$$\frac{\beta_{r,r}}{\beta_{r,\ell}} \|F_{=r}\|_2^2 \leq \frac{2^{r^2}}{\frac{1}{2} \cdot 2^{r\ell}} \|F\|_2^2 \leq \frac{2^{r^2+1}}{2^{r\ell}} \varepsilon \leq \frac{C_r}{2^{r\ell}} \varepsilon.$$

Now assume  $j \geq 1$ . By Lemma 3.18 and 3.4, for any  $\dim(A_1, \dots, A_j, T_{j+1}, \dots, T_r) = r$ ,

$$\begin{aligned}
\widehat{f}_{=r}^2(A_1, \dots, A_j, T_{j+1}, \dots, T_r) &= \widehat{F}^2(A_1, \dots, A_j, T_{j+1}, \dots, T_r) \\
&= \frac{1}{(2^\ell - 2^{r-1})^2} \left( \widehat{F}_{W_{A_1}}(A_2, \dots, A_j, T_{j+1}, \dots, T_r) - \sum_{\substack{D \subseteq \text{Span}(A_1, \dots, A_j, T_{j+1}, \dots, T_r) \\ \dim(D)=r-1, A_1 \notin D}} \widehat{F}(D) \right)^2 \\
&\leq \frac{4 \cdot 2^r}{2^{2\ell}} \left( \widehat{F}_{W_{A_1}}^2(A_2, \dots, A_j, T_{j+1}, \dots, T_r) + \sum_{\substack{D \subseteq \text{Span}(A_1, \dots, A_j, T_{j+1}, \dots, T_r) \\ \dim(D)=r-1, A_1 \notin D}} \widehat{F}^2(D) \right), \tag{10}
\end{aligned}$$

the last inequality is by Cauchy-Schwarz (there are  $2^{r-1}$  choices for  $D$ ). Summing over  $T_{j+1}, \dots, T_r$ , the first term sums up to at most

$$\begin{aligned}
&\sum_{\substack{T_{j+1}, \dots, T_r \\ \dim(A_2, \dots, A_j, T_{j+1}, \dots, T_r)=r-1}} \widehat{F}_{W_{A_1}}^2(A_2, \dots, A_j, T_{j+1}, \dots, T_r) \\
&= \sum_{\substack{T_{j+1}, \dots, T_r \\ \dim(A_2, \dots, A_j, T_{j+1}, \dots, T_r)=r-1}} \widehat{f}_{=r-1, F_{W_{A_1}}}^2(A_2, \dots, A_j, T_{j+1}, \dots, T_r) \\
&\leq \frac{C_{r-1}}{2^{(r-1+j-1)\ell}} \varepsilon,
\end{aligned}$$

using the induction hypothesis and since the  $(r, \varepsilon)$  pseudorandomness of  $F$  implies  $(r-1, \varepsilon)$  pseudorandomness of  $F_{W_{A_1}}$ . For the second term, consider any  $D \subseteq \text{Span}(A_1, \dots, A_j, T_{j+1}, \dots, T_r)$  of dimension  $r-1$  not containing  $A_1$ . By Lemma 3.15, we may assume that  $D$  has basis

$$D = \text{Span}(A'_2, \dots, A'_j, T'_{j+1}, \dots, T'_r)$$

where  $A'_i = A_i + b_i \cdot A_1$  and  $T'_i = T_i + b_i \cdot A_1$  for some  $b = (b_2, \dots, b_r)$ . In the following calculation,  $b$  is thought of as fixed, determining  $D$ . Summing over all  $T_{j+1}, \dots, T_r$ ,

$$\begin{aligned}
\sum_{\substack{T_{j+1}, \dots, T_r \\ \dim(A_1, \dots, A_j, T_{j+1}, \dots, T_r)=r}} \widehat{F}^2(D) &\leq \sum_{\substack{T'_{j+1}, \dots, T'_r \\ \dim(A'_2, \dots, A'_j, T'_{j+1}, \dots, T'_r)=r-1}} \widehat{F}^2(A'_2, \dots, A'_j, T'_{j+1}, \dots, T'_r) \\
&= \sum_{\substack{T'_{j+1}, \dots, T'_r \\ \dim(A'_2, \dots, A'_j, T'_{j+1}, \dots, T'_r)=r-1}} \widehat{f}_{=r-1}^2(A'_2, \dots, A'_j, T'_{j+1}, \dots, T'_r) \\
&\leq \frac{C_{r-1}}{2^{(r-1+j-1)\ell}} \varepsilon,
\end{aligned}$$

using the induction hypothesis. We note that there are  $2^{r-1}$  choices for  $b$  (or equivalently  $D$ ). Combining both the upper bounds, gets us 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 3.20.** *Let  $S$  be a (basis invariant) set of vertices in  $H_{k,\ell}$  that is  $(r, \varepsilon)$  pseudo-random. Let  $s, t, p, q$  be non-negative integers such that  $s + t + p + q = r \leq \frac{\ell}{2}$ . Let  $F: (\{0, 1\}^k)^\ell \rightarrow \{0, 1\}$  be the indicator function of  $S$ . Let  $a_1, \dots, a_s$  be points and  $A_1, \dots, A_p$  be characters. Define the restriction*

$$g_{a_1, \dots, a_s, x_1, \dots, x_t}(y_1, \dots, y_p, z_1, \dots, z_q) = f_{=r}(a_1, \dots, a_s, x_1, \dots, x_t, y_1, \dots, y_p, z_1, \dots, z_q).$$

Then

$$\mathbb{E}_{x_1, \dots, x_t \in \{0, 1\}^k} \left[ \sum_{T_1, \dots, T_q} \widehat{g}_{a_1, \dots, a_s, x_1, \dots, x_t}^2(A_1, \dots, A_p, T_1, \dots, T_q) \right] \leq \frac{2^{2sr^2+4r^2}}{2^{(t+2p+q)\ell}} \varepsilon \leq \frac{2^{6r^3}}{2^{(t+2p+q)\ell}} \varepsilon.$$

*Proof.* We will prove an upper bound of  $\frac{C_{s,r}}{2^{(t+2p+q)\ell}} \varepsilon$  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 to be upper-bounded equals (we denote by  $x, y, z$  the respective tuples of variables)

$$\mathbb{E}_x \left[ \sum_{T_1, \dots, T_q} \left( \mathbb{E}_{y, z} \left[ f_{=r}(x, y, z) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] \right)^2 \right]. \quad (11)$$

Consider the expectation inside. Expanding  $f_{=r}$  into Fourier, it equals

$$\begin{aligned} & \sum_{\substack{Q_1, \dots, Q_t \\ R_1, \dots, R_p, S_1, \dots, S_q}} \mathbb{E}_{y, z} \left[ \widehat{f}_{=r}(Q, R, S) \prod_{i=1}^t \chi_{Q_i}(x_i) \prod_{i=1}^p \chi_{A_i \oplus R_i}(y_i) \prod_{i=1}^q \chi_{T_i \oplus S_i}(z_i) \right] \\ &= \sum_{Q_1, \dots, Q_t} \widehat{f}_{=r}(Q, A, T) \prod_{i=1}^t \chi_{Q_i}(x_i). \end{aligned}$$

Squaring this, taking the expectation over  $x$ , and then summing over  $T_1, \dots, T_q$  shows that (11) equals,

$$\sum_{Q_1, \dots, Q_t, T_1, \dots, T_q} \widehat{f}_{=r}^2(Q_1, \dots, Q_t, A_1, \dots, A_p, T_1, \dots, T_q),$$

which is upper bounded by  $\frac{2^{4r^2}}{2^{(t+2p+q)\ell}} \varepsilon$  by Lemma 3.19. Now consider the case  $s \geq 1$ . As before, the expectation to be upper-bounded equals (with an additional argument  $a = (a_1, \dots, a_s)$ )

$$\mathbb{E}_x \left[ \sum_{T_1, \dots, T_q} \left( \mathbb{E}_{y, z} \left[ f_{=r}(a, x, y, z) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] \right)^2 \right]. \quad (12)$$

Applying Lemma 3.8 we get

$$\begin{aligned} f_{=r}(a_1, \dots, a_s, x, y, z) &= f_{=r-1, F_{a_1}}(a_2, \dots, a_s, x, y, z) \\ &+ \frac{1}{\beta_{r-1, r-1}} \sum_{\substack{M \in \mathcal{M}[r-1, r] \\ e_1 \notin \text{rowspan}(M)}} f_{=r-1, F_{a_1}}(M(a, x, y, z)). \end{aligned}$$

We take expectation over  $y, z$ . For the first term, we have

$$\mathbb{E}_{y,z} \left[ f_{=r-1, F_{a_1}}(a_2, \dots, a_s, x, y, z) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] = \widehat{h}_{a_2, \dots, a_s, x}(A_1, \dots, A_p, T_1, \dots, T_q), \quad (13)$$

where  $h_{a_2, \dots, a_s, x}$  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  $M \in \mathcal{M}[r-1, r]$  whose row-span does not contain  $e_1$ . By Lemma 3.15, we may assume that  $M$  is semi-diagonal and then

$$M(a, x, y, z) = (a' = (a'_2, \dots, a'_s), x', y', z'),$$

where each new coordinate is same as earlier except possibly adding  $a_1$ . Hence

$$\mathbb{E}_{y,z} \left[ f_{=r-1}(M(a, x, y, z)) \prod_{i=1}^p \chi_{A_i}(y_i) \prod_{i=1}^q \chi_{T_i}(z_i) \right] = \text{sign} \cdot \widehat{h}_{a', x'}(A_1, \dots, A_p, T_1, \dots, T_q) \quad (14)$$

for a  $\text{sign} \in \{-1, 1\}$  (which takes into account the possible additions of  $a_1$  to get the new coordinates  $y', z'$ ). Towards upper-bounding (12), we can now sum up the absolute values of (13) and (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  $T = (T_1, \dots, T_q)$  and expectation over  $x$ . We end up with an overall upper bound

$$\left( 1 + \frac{\beta_{r-1, r}}{\beta_{r-1, r-1}^2} \right) \left( \mathbb{E}_x \left[ \sum_T \widehat{h}_{a_2, \dots, a_s, x}^2(A, T) \right] + \sum_{M \in \mathcal{M}[r-1, r]} \mathbb{E}_x \left[ \sum_T \widehat{h}_{a', x'}^2(A, T) \right] \right).$$

We may now apply the induction hypothesis since sequences  $(a_2, \dots, a_s)$  and  $a'$  have length  $s-1$ ,  $x'$  is distributed the same as  $x$ , and furthermore,  $F_{a_1}$  is  $(r-1, \varepsilon)$  pseudo-random. Thus we get an upper bound of

$$\frac{\varepsilon}{2^{(t+2p+q)\ell}} \cdot C_{s-1, r-1} \cdot \left( 1 + \frac{\beta_{r-1, r}}{\beta_{r-1, r-1}^2} \right) (1 + \beta_{r-1, r}).$$

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  $\frac{\varepsilon}{2^{(t+2p+q)\ell}} \cdot C_{s, r}$ . It suffices to have  $C_{s, r} \geq 2^{2r^2} C_{s-1, r-1}$  and  $C_{0, r} = 2^{4r^2}$ , i.e.  $C_{s, r} = 2^{2sr^2 + 4r^2}$ .  $\square$

## 4 Pair-wise and Three-wise Correlations of $f_{=i}$

The rest of the paper is devoted to the proof of our main technical result, Theorem 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  $f_{=i}$ .

### 4.1 Pairwise Correlations

Studying pairwise correlations is simple. There are two cases depending on whether  $\text{rowspan}(M_1)$ ,  $\text{rowspan}(M_2)$  are distinct or the same. In the latter case, due to basis-invariance of  $f_{=i}$ , we may assume that the two matrices are the same.

**Lemma 4.1.** *Let  $M_1, M_2 \in \mathcal{M}[i, \ell]$ . If  $\text{rowspan}(M_1) \neq \text{rowspan}(M_2)$ , then*

$$\mathbb{E}_{x_1, \dots, x_\ell \in \{0, 1\}^k} [f_{=i}(M_1 x) f_{=i}(M_2 x)] = 0.$$

*Proof.* It is clearly possible to choose linearly independent vectors  $v_1, \dots, v_s, u_1, \dots, u_{i-s}, w_1, \dots, w_{i-s}$  in  $\{0, 1\}^\ell$  such that

- $(v_1, \dots, v_s)$  is a basis for  $\text{rowspan}(M_1) \cap \text{rowspan}(M_2)$ ,
- $(v_1, \dots, v_s, u_1, \dots, u_{i-s})$  is a basis for  $\text{rowspan}(M_1)$ , and
- $(v_1, \dots, v_s, w_1, \dots, w_{i-s})$  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  $x = (x_1, \dots, x_\ell)$ , let us denote  $a' = \langle a, x \rangle = \sum_{j=1}^\ell a_j x_j$ . Let us define

$$v'_j = \langle v_j, x \rangle, \quad u'_j = \langle u_j, x \rangle, \quad w'_j = \langle w_j, x \rangle.$$

Thus  $\{v'_j, u'_j, w'_j\}$  are uniformly and independently distributed over  $\{0, 1\}^k$ . Moreover

$$M_1 x = (v'_1, \dots, v'_s, u'_1, \dots, u'_{i-s}), \quad M_2 x = (v'_1, \dots, v'_s, w'_1, \dots, w'_{i-s}).$$

It follows using Lemma 3.14 that

$$\begin{aligned} \mathbb{E}_x [f_{=i}(M_1 x) f_{=i}(M_2 x)] &= \mathbb{E}_{v'_j, u'_j, w'_j} [f_{=i}(v'_1, \dots, v'_s, u'_1, \dots, u'_{i-s}) f_{=i}(v'_1, \dots, v'_s, w'_1, \dots, w'_{i-s})] \\ &= \mathbb{E}_{v'_j, w'_j} \left[ \mathbb{E}_{u'_1, \dots, u'_{i-s}} [f_{=i}(v'_1, \dots, v'_s, u'_1, \dots, u'_{i-s})] \cdot f_{=i}(v'_1, \dots, v'_s, w'_1, \dots, w'_{i-s}) \right] \\ &= 0. \end{aligned}$$

□

**Lemma 4.2.** *Let  $M \in \mathcal{M}[i, \ell]$ . Then*

$$\mathbb{E}_{x_1, \dots, x_\ell \in \{0, 1\}^k} [f_{=i}^2(Mx)] = \frac{\beta_{i,i}}{\beta_{i,\ell}} \|F_{=i}\|_2^2.$$

*Proof.* Denoting  $(y_1, \dots, y_i) = Mx$ , we note that  $y_1, \dots, y_i$  are uniformly and independently distributed in  $\{0, 1\}^k$  and hence the expectation above, call it  $\Gamma$ , does not depend on the choice of  $M \in \mathcal{M}[i, \ell]$ . Also, due to basis variance, 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 3.13, squaring, and taking expectation over  $x$ ,

$$\begin{aligned} F_{=i}[x_1, \dots, x_\ell] &= \frac{1}{\beta_{i,i}} \sum_{M \in \mathcal{M}[i, \ell]} f_{=i}(Mx). \\ \beta_{i,i}^2 \cdot \|F_{=i}\|_2^2 &= \sum_{\substack{M_1, M_2 \in \mathcal{M}[i, \ell] \\ \text{rowspan}(M_1) = \text{rowspan}(M_2)}} \Gamma + \sum_{\substack{M_1, M_2 \in \mathcal{M}[i, \ell] \\ \text{rowspan}(M_1) \neq \text{rowspan}(M_2)}} \mathbb{E}_x [f_{=i}(M_1 x) f_{=i}(M_2 x)]. \end{aligned}$$

The lemma follows by noting that that are  $\beta_{i,\ell} \beta_{i,i}$  pairs  $M_1, M_2$  with the same row-span and by previous Lemma 4.1, the expectation vanishes when the row-spans are distinct. □

The left hand side in the statement of this Lemma equals  $\|f_{=i}\|_2^2$  and we have  $\beta_{i,i} \leq 2^{i^2-1}$ ,  $\beta_{i,\ell} \geq \frac{1}{2} \cdot 2^{i\ell}$ . We record this very useful fact for future:

**Lemma 4.3.**  $\|f_{=i}\|_2^2 \leq \frac{2^{i^2}}{2^{i\ell}} \|F_{=i}\|_2^2$ .

## 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 3. Our formal result is:

**Theorem 4.4.** *Let  $S$  be a basis-invariant set of vertices in  $H_{k,\ell}$  that is  $(r, \varepsilon)$  pseudo-random. Let  $F: H_{k,\ell} = (\{0, 1\}^k)^\ell \rightarrow \{0, 1\}$  be the indicator function of  $S$  and  $\eta = \|F_{=i}\|_2^2$ . Then for any  $0 \leq i \leq r$ ,  $i \leq d \leq 3i$ ,  $A \subseteq \{0, 1\}^\ell$  of dimension  $d$  and  $M_1, M_2, M_3 \in \mathcal{M}[i, \ell]$  such that  $\bigoplus_{s=1}^3 \text{rowspan}(M_s) = A$ , we have that*

$$\left| \mathbb{E}_{x \in (\{0,1\}^k)^\ell} [f_{=i}(M_1x)f_{=i}(M_2x)f_{=i}(M_3x)] \right| \leq 2^{4r^2} \frac{\eta\sqrt{\varepsilon}}{2^{d\ell}}. \quad (15)$$

The rest of this section is devoted to the proof of the above lemma. Fix  $i$ ,  $i \leq d \leq 3i$ ,  $A \subseteq \{0, 1\}^\ell$  of dimension  $d$ , and  $M_1, M_2, M_3 \in \mathcal{M}[i, \ell]$  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 4.5.** *Suppose  $\text{rowspan}(M_3) \not\subseteq \text{rowspan}(M_1) \oplus \text{rowspan}(M_2)$  (or the other two symmetric cases). Then*

$$\mathbb{E}_{x \in (\{0,1\}^k)^\ell} [f_{=i}(M_1x)f_{=i}(M_2x)f_{=i}(M_3x)] = 0.$$

*Proof.* The proof is essentially the same as that of Lemma 4.1. We can choose linearly independent vectors  $w_1, \dots, w_t, v_1, \dots, v_s$  such that  $t \geq 1$  and

- $w_1, \dots, w_t$  are the first  $t$  rows of  $M_3$ .
- $v_1, \dots, v_s$  span the remaining  $i - t$  rows of  $M_3$  as well as  $\text{rowspan}(M_1) \oplus \text{rowspan}(M_2)$ .

We define  $v'_j = \langle v_j, x \rangle$  and  $w'_j = \langle w_j, x \rangle$  so that  $\{v'_j, w'_j\}$  are uniformly and independently distributed in  $\{0, 1\}^k$ . Clearly,  $M_1x, M_2x$  depend only on  $v'_j$  and  $M_3x = (w'_1, \dots, w'_t, y_1, \dots, y_{i-t})$  where  $y_1, \dots, y_{i-t}$  depend only on  $v'_j$ . Fixing an arbitrary choice of  $v'_j$  fixes  $a^* = M_1x, b^* = M_2x$  and  $c^* = (y_1, \dots, y_{i-t})$ . The expectation is then

$$f_{=i}(a^*)f_{=i}(b^*) \mathbb{E}_{w'_1, \dots, w'_t} [f_{=i}(w'_1, \dots, w'_t, c^*)],$$

which vanishes according to Lemma 3.14. □

Thanks to Lemma 4.5, we assume henceforth that the row-span of each matrix is contained in the direct sum of the other two. Let  $H = \bigcap_{j=1}^3 \text{rowspan}(M_j)$ ,  $\dim(H) = s$ , and let  $g_1, \dots, g_s$  be a basis for it. We may assume w.l.o.g. that the first  $s$  rows of each matrix are  $g_1, \dots, g_s$ . 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  $\bigcap_{j=1}^3 \text{rowspan}(M'_j) = \{0\}$  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 4.6.** *The row spans of  $M'_1, M'_2, M'_3$  have the form*

$$\begin{aligned} & \text{Span}(w_1, \dots, w_t, \quad y_1, \dots, y_n, \quad y_{n+1}, \dots, y_{i-s-t}) \\ & \text{Span}(w_1, \dots, w_t, \quad z_1, \dots, z_n, \quad z_{n+1}, \dots, z_{i-s-t}) \\ & \text{Span}(y_1 + z_1, \dots, y_n + z_n, \quad y_{n+1}, \dots, y_{i-s-t}, \quad z_{n+1}, \dots, z_{i-s-t}) \end{aligned}$$

where the vectors  $w_1, \dots, w_t, y_1, \dots, y_{i-s-t}, z_1, \dots, z_{i-s-t}$  are linearly independent.

*Proof.* Let  $\{w_1, \dots, w_t\}$  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  $w_1 = \dots = w_t = 0$  (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 A.1 to get the desired form. One caveat 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  $w_1, \dots, w_t$  (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  $y_j \leftarrow y_j + \sigma(w)$ .  $\square$

We now turn back to the task of upper bounding the expectation in (15). We make a change of variables:  $g'_j = \langle g_j, x \rangle$  for  $j = 1, \dots, s$ ,  $w'_j = \langle w_j, x \rangle$  for  $j = 1, \dots, t$ , and  $y'_j = \langle y_j, x \rangle$  and  $z'_j = \langle z_j, x \rangle$  for  $j = 1, \dots, i - s - t$ . 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 (15) equals

$$\begin{aligned} & \mathbb{E}_{g,w,y,z} \left[ f_{=i}(g_1, \dots, g_s, \quad w_1, \dots, w_t, \quad y_1, \dots, y_n, \quad y_{n+1}, \dots, y_{i-s-t}) \right. \\ & \quad f_{=i}(g_1, \dots, g_s, \quad w_1, \dots, w_t, \quad z_1, \dots, z_n, \quad z_{n+1}, \dots, z_{i-s-t}) \\ & \quad \left. f_{=i}(g_1, \dots, g_s, \quad y_1 + z_1, \dots, y_n + z_n, \quad y_{n+1}, \dots, y_{i-s-t}, \quad z_{n+1}, \dots, z_{i-s-t}) \right]. \end{aligned}$$

Denote  $h_{g_1, \dots, g_s}(a_1, \dots, a_{i-s}) = f_{=i}(g_1, \dots, g_s, a_1, \dots, a_{i-s})$ . To reduce cumbersome notation, we drop the subscript from  $h$  for now and remember that it is  $g_1, \dots, g_s$  throughout. Then our expectation is

$$\begin{aligned} & \mathbb{E}_{g,w,y,z} \left[ h(w_1, \dots, w_t, \quad y_1, \dots, y_n, \quad y_{n+1}, \dots, y_{i-s-t}) \right. \\ & \quad h(w_1, \dots, w_t, \quad z_1, \dots, z_n, \quad z_{n+1}, \dots, z_{i-s-t}) \\ & \quad \left. h(y_1 + z_1, \dots, y_n + z_n, \quad y_{n+1}, \dots, y_{i-s-t}, \quad z_{n+1}, \dots, z_{i-s-t}) \right]. \end{aligned}$$

For a tuple  $(b_1, b_2, \dots, b_n)$  and  $m_1 \leq m_2$ , we will denote by  $b_{[m_1:m_2]}$  the sub-tuple  $(b_{m_1}, b_{m_1+1}, \dots, b_{m_2})$ . Applying the Fourier transform on  $h$  and using the expectation over  $w, y, z$ , we see that the expectation equals

$$\begin{aligned} & \mathbb{E}_g \sum_{\substack{T_1, \dots, T_n \\ W_1, \dots, W_t \\ P_{n+1}, \dots, P_{i-s-t} \\ Q_{n+1}, \dots, Q_{i-s-t}}} \left[ \widehat{h}(W_{[1:t]}, T_{[1:n]}, P_{[n+1:i-s-t]}) \cdot \widehat{h}(W_{[1:t]}, T_{[1:n]}, Q_{[n+1:i-s-t]}) \right. \\ & \quad \left. \widehat{h}(T_{[1:n]}, P_{[n+1:i-s-t]}, Q_{[n+1:i-s-t]}) \right]. \end{aligned}$$

For ease of notation, we will denote by  $T$  the tuple  $(T_1, \dots, T_n)$  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 A.4):

$$\begin{aligned} & \sum_{W, P} |\hat{h}(W, T, P)| \left( \sum_Q |\hat{h}(W, T, Q)| \cdot |\hat{h}(T, P, Q)| \right) \\ & \leq \sum_{W, P} |\hat{h}(W, T, P)| \left( \sqrt{\sum_Q \hat{h}^2(W, T, Q)} \sqrt{\sum_Q \hat{h}^2(T, P, Q)} \right) \\ & = \sum_W \sqrt{\sum_Q \hat{h}^2(W, T, Q)} \left( \sum_P |\hat{h}(W, T, P)| \cdot \sqrt{\sum_Q \hat{h}^2(T, P, Q)} \right) \\ & \leq \sum_W \sqrt{\sum_Q \hat{h}^2(W, T, Q)} \left( \sqrt{\sum_P \hat{h}^2(W, T, P)} \sqrt{\sum_{P, Q} \hat{h}^2(T, P, Q)} \right) \\ & = \sqrt{\sum_{P, Q} \hat{h}^2(T, P, Q)} \sum_W \left( \sqrt{\sum_Q \hat{h}^2(W, T, Q)} \sqrt{\sum_P \hat{h}^2(W, T, P)} \right) \\ & \leq \sqrt{\sum_{P, Q} \hat{h}^2(T, P, Q)} \sqrt{\sum_{W, Q} \hat{h}^2(W, T, Q)} \sqrt{\sum_{W, P} \hat{h}^2(W, T, P)} \\ & \stackrel{def}{=} \sqrt{A_1(T)} \sqrt{A_2(T)} \sqrt{A_2(T)}, \end{aligned}$$

where we labeled 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

$$\mathbb{E}_g \left[ \sum_T \sqrt{A_1(T)} A_2(T) \right] \leq \sqrt{\max_{g, T} A_1(T)} \cdot \mathbb{E}_g \left[ \sum_T A_2(T) \right].$$

By Parseval and by Lemma 4.3,

$$\mathbb{E}_g \left[ \sum_T A_2(T) \right] = \mathbb{E}_g \left[ \sum_{T, W, P} \hat{h}^2(W, T, P) \right] = \mathbb{E}_g [\|h\|_2^2] = \|f_{=i}\|_2^2 \leq \frac{2^{i^2}}{2^{i\ell}} \eta \leq \frac{2^{r^2}}{2^{i\ell}} \eta. \quad (16)$$

By Lemma 3.20, we have

$$\max_{g, T} A_1(T) = \max_{g, T} \sum_{P, Q} \hat{h}^2(T, P, Q) \leq \frac{2^{6i^3}}{2^{2n+(2 \cdot (i-s-t-n))\ell}} \varepsilon \leq \frac{2^{6r^3}}{2^{2(i-s-t)\ell}} \varepsilon. \quad (17)$$

Combining both upper bounds (16) and (17), and noting that  $d = 2i - s - t$ , we get the desired upper bound

$$\left( \frac{2^{6r^3}}{2^{2(i-s-t)\ell}} \varepsilon \right)^{\frac{1}{2}} \frac{2^{r^2}}{2^{i\ell}} \eta \leq 2^{4r^2} \frac{\eta\sqrt{\varepsilon}}{2^{(2i-s-t)\ell}}.$$

## 5 Four-wise Correlations: Getting the Matrices into Convenient Form

We now begin the proof of our main technical Theorem 2.15. This section is devoted to bringing 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 5.1.** *Suppose  $\text{rowspan}(M_4) \not\subseteq \bigoplus_{j=1}^3 \text{rowspan}(M_j)$  (or the other three symmetric cases). Then*

$$\mathbb{E}_{x \in (\{0,1\}^k)^\ell} [f_{=i}(M_1x)f_{=i}(M_2x)f_{=i}(M_3x)f_{=i}(M_4x)] = 0.$$

*Proof.* Essentially the same as that of Lemma 4.5. □

### 5.1 Removing 4-wise and 3-wise Intersections of Rowspaces

Consider the subspace  $\bigcap_{j=1}^4 \text{rowspan}(M_j)$ . 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  $v_1, \dots, v_r$  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  $\bigcap_{j=1}^4 \text{rowspan}(M_j) = \{0\}$  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  $w \in \bigcap_{j=1}^3 \text{rowspan}(M_j)$ . Since we assumed that the 4-wise intersection of the row-spaces is trivial,  $w \notin \text{rowspan}(M_4)$ . We may assume w.l.o.g. that  $w$  is the first row of  $M_1, M_2, M_3$  and their rest of the rows as well as the rows of  $M_4$  are linear combinations of vectors  $v_1, \dots, v_r$  (not necessarily the same as in the previous para) 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 notational convenience, 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  $j^{\text{th}}$  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.

## 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 4.6 and Lemma A.1, we can write  $A, B, C$  as

$$\begin{aligned} & \text{Span}(v_1, \dots, v_t, p_1, \dots, p_n, u, y) \\ & \text{Span}(v_1, \dots, v_t, q_1, \dots, q_n, w, z) \\ & \text{Span}(p_1 + q_1, \dots, p_n + q_n, u, w) \end{aligned}$$

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  $a = (a_1, \dots, a_h)$  from  $\text{rowspan}(M_3) \setminus C$ . Hence the row-spans of  $M_1, M_2, M_3$  can be assumed to be in the form ( $p, q$  have the same length  $n$ ):

$$\begin{aligned} & \text{Span}(v, p, u, y) \\ & \text{Span}(v, q, w, z) \\ & \text{Span}(a, p_1 + q_1, \dots, p_n + q_n, u, w). \end{aligned} \tag{18}$$

## 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 (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 A.2, there is a basis for  $W$  of the following form  $\cup_{s=1}^7 A_s$  where

$$\begin{array}{l|l} A_1 = \{ & a_i + y_j + z_k \\ A_2 = \{ & a_i + y_j \\ A_3 = \{ & a_i + z_k \\ A_4 = \{ & a_i + \sigma(y_{\Phi_1, \Phi_5}) \\ A_5 = \{ & y_j + z_k \\ A_6 = \{ & y_j \\ A_7 = \{ & z_k \end{array} \quad \left| \quad \begin{array}{l} i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \\ i \in \Sigma_2, j \in \Phi_2 \\ i \in \Sigma_3, k \in \Psi_3 \\ i \in \Sigma_4 \\ j \in \Phi_5, k \in \Psi_5 \\ j \in \Phi_6 \\ k \in \Psi_7 \end{array} \right. \quad \left. \begin{array}{l} \} \\ \} \\ \} \\ \} \\ \} \\ \} \\ \} \end{array}$$

Here  $\sigma(y_{\Phi_1, \Phi_5})$  are arbitrary linear forms in  $\{y_j | j \in \Phi_1 \cup \Phi_5\}$  (possibly different ones, but we hide this fact as it will be essentially irrelevant). 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 the sake of ease. 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 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  $\sigma(v, p, u, q, w)$  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 5.2.**  $\text{rowspan}(M_4)$  has a partial basis of the following form  $\cup_{s=1}^7 A_s$  where

$$\begin{aligned}
A_1 &= \{ a_i + y_j + z_k + \sigma(v, p, u, q, w) & | & i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \} \\
A_2 &= \{ a_i + y_j + \sigma(v, p, u, q, w) & | & i \in \Sigma_2, j \in \Phi_2 \} \\
A_3 &= \{ a_i + z_k + \sigma(v, p, u, q, w) & | & i \in \Sigma_3, k \in \Psi_3 \} \\
A_4 &= \{ a_i + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5}) & | & i \in \Sigma_4 \} \\
A_5 &= \{ y_j + z_k + \sigma(v, p, u, q, w) & | & j \in \Phi_5, k \in \Psi_5 \} \\
A_6 &= \{ y_j + \sigma(v, p, u, q, w) & | & j \in \Phi_6 \} \\
A_7 &= \{ z_k + \sigma(v, p, u, q, w) & | & k \in \Psi_7 \}.
\end{aligned}$$

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.

## 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 A.3, for a partition of the index set  $\{1, \dots, n\} = \Delta_0 \cup \Delta_1 \cup \Delta_2 \cup \dots \cup \Delta_m \cup \Omega_1 \cup \Omega_2$ , we may assume that  $W$  has a basis

$$\begin{aligned}
A_8 &= \{ p_i + \sigma(q) & | & i \in \Delta_1 \} \\
A_9 &= B_2 \cup \dots \cup B_m \\
A_{10} &= \{ p_i + \sigma(q_{\Omega_1}) & | & i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} &= \{ q_j & | & j \in \Omega_2 \}.
\end{aligned}$$

Here  $B_s = \{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) \mid j \in \Delta_s \}$ . We recall that  $\Delta_{[s+1:m]} = \Delta_{s+1} \cup \dots \cup \Delta_m \cup \Omega_1 \cup \Omega_2$ . 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 5.3.** A partial basis for  $\text{rowspan}(M_4)$  from Lemma 5.2 can be further extended as  $A_8 \cup A_9 \cup A_{10} \cup A_{11}$  where

$$\begin{aligned}
A_8 &= \{ p_i + \sigma(q) + \sigma(v, u, w) & | & i \in \Delta_1 \} \\
A_9 &= B_2 \cup \dots \cup B_m \\
A_{10} &= \{ p_i + \sigma(q_{\Omega_1}) + \sigma(v, u, w) & | & i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} &= \{ q_j + \sigma(v, u, w) & | & j \in \Omega_2 \}.
\end{aligned}$$

Here  $B_s = \left\{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) + \sigma(v, u, w) \mid j \in \Delta_s \right\}$ . 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 5.4.** *A partial basis for  $\text{rowspan}(M_4)$  from Lemmas 5.2 and 5.3 can be completed by adding  $A_{12} \cup A_{13}$  where (for some index set  $\Gamma$ )*

$$\begin{aligned} A_{12} &= \{ u_i + \sigma(v, w) \mid i \in \Gamma \} \\ A_{13} &= \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)). \end{aligned}$$

To summarize, basis for  $\text{rowspan}(M_4)$  can be assumed to be  $\cup_{s=1}^{13} A_s$  where

$$\begin{aligned} A_1 &= \{ a_i + y_j + z_k + \sigma(v, p, u, q, w) & \mid & i \in \Sigma_1, \quad j \in \Phi_1, \quad k \in \Psi_1 \} \\ A_2 &= \{ a_i + y_j + \sigma(v, p, u, q, w) & \mid & i \in \Sigma_2, \quad j \in \Phi_2 & \} \\ A_3 &= \{ a_i + z_k + \sigma(v, p, u, q, w) & \mid & i \in \Sigma_3, \quad k \in \Psi_3 \} \\ A_4 &= \{ a_i + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5}) & \mid & i \in \Sigma_4 & \} \\ A_5 &= \{ y_j + z_k + \sigma(v, p, u, q, w) & \mid & j \in \Phi_5, \quad k \in \Psi_5 \} \\ A_6 &= \{ y_j + \sigma(v, p, u, q, w) & \mid & j \in \Phi_6 & \} \\ A_7 &= \{ z_k + \sigma(v, p, u, q, w) & \mid & k \in \Psi_7 \} \\ \\ A_8 &= \{ p_i + \sigma(q) + \sigma(v, u, w) & \mid & i \in \Delta_1 & \} \\ A_9 &= B_2 \cup \dots \cup B_m \\ A_{10} &= \{ p_i + \sigma(q_{\Omega_1}) + \sigma(v, u, w) & \mid & i \in \Omega_1 \cup \Omega_2 & \} \\ A_{11} &= \{ q_j + \sigma(v, u, w) & \mid & j \in \Omega_2 & \} \end{aligned}$$

$$\begin{aligned} A_{12} &= \{ u_i + \sigma(v, w) \mid i \in \Gamma \} \\ A_{13} &= \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)). \end{aligned} \tag{19}$$

Here  $B_s = \left\{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) + \sigma(v, u, w) \mid j \in \Delta_s \right\}$ . The variables  $\{a_i, y_j, z_k\}$  appearing in  $A_1, \dots, A_7$ , the variables  $\{p_i, q_j\}$  appearing in  $A_8, \dots, A_{11}$  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).

## 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 (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.

$$\begin{aligned}
A_1 &= \{ a_i + y_j + z_k & | & i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \} \\
A_2 &= \{ a_i + y_j & + \sigma( & q, w) & | & i \in \Sigma_2, j \in \Phi_2 & \} \\
A_3 &= \{ a_i & + z_k & + \sigma( & p, u & ) & | & i \in \Sigma_3, & k \in \Psi_3 \} \\
A_4 &= \{ a_i & & + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5}) & | & i \in \Sigma_4 & \} \\
A_5 &= \{ & y_j + z_k & & | & j \in \Phi_5, k \in \Psi_5 \} \\
A_6 &= \{ & y_j & + \sigma( & q, w) & | & j \in \Phi_6 & \} \\
A_7 &= \{ & z_k & + \sigma( & p, u & ) & | & k \in \Psi_7 \} \\
A_8 &= & \{ p_i + \sigma(q) & + \sigma(v, w) & | & i \in \Delta_1 & \} \\
A_9 &= B_2 \cup \dots \cup B_m \\
A_{10} &= & \{ p_i + \sigma(q_{\Omega_1}) & + \sigma(v, w) & | & i \in \Omega_1 \cup \Omega_2 & \} \\
A_{11} &= & \{ & q_j + \sigma(v, u) & | & j \in \Omega_2 & \} \\
A_{12} &= \{ u_i + \sigma(v, w) & | & i \in \Gamma \} \\
A_{13} &= \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)).
\end{aligned}$$

Here  $B_s = \{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) + \sigma(v, u) \mid j \in \Delta_s \}$ .

## 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  $p_{\Delta_{[2:m]}}$  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  $\text{rowspan}(M_4)$  can be written as:

$$\begin{aligned}
A_1 &= \{ a_i + y_j + z_k & | & i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \} \\
A_2 &= \{ a_i + y_j & + \sigma( & q, w) & | & i \in \Sigma_2, j \in \Phi_2 & \} \\
A_3 &= \{ a_i & + z_k & + \sigma( & p, u & ) & | & i \in \Sigma_3, & k \in \Psi_3 \} \\
A_4 &= \{ a_i & & + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5}) & | & i \in \Sigma_4 & \} \\
A_5 &= \{ & y_j + z_k & & | & j \in \Phi_5, k \in \Psi_5 \} \\
A_6 &= \{ & y_j & + \sigma( & q, w) & | & j \in \Phi_6 & \} \\
A_{7a} &= \{ & z_k + p_s & & | & k \in \Psi_{7a}, s \in \Delta'_0 \} \\
A_{7b} &= \{ & z_k & + \sigma( & p_{\Delta_{[2:m]}} & , u & ) & | & k \in \Psi_{7b} \}
\end{aligned}$$

$$\begin{aligned}
A_8 &= \{ p_i + \sigma(q) + \sigma(v, w) \mid i \in \Delta_1 \} \\
A_9 &= B_2 \cup \dots \cup B_m \\
A_{10} &= \{ p_i + \sigma(q_{\Omega_1}) + \sigma(v, w) \mid i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} &= \{ q_j + \sigma(v, u) \mid j \in \Omega_2 \}
\end{aligned}$$

$$\begin{aligned}
A_{12} &= \{ u_i + \sigma(v, w) \mid i \in \Gamma \} \\
A_{13} &= \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)).
\end{aligned}$$

Here  $B_s = \{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) + \sigma(v, u) \mid j \in \Delta_s \}$ .

### Step 3

Finally, we consider  $A_{11}$  and its vectors  $\{ q_j + \sigma(v, u) \mid j \in \Omega_2 \}$ . By adding vectors from  $A_{12}$  if necessary, we can assume that the form  $\sigma(\cdot)$  does not depend on  $u_\Gamma$  (we may need re-absorption of  $w$  into  $q_j$ ). In other words,  $\sigma(\cdot)$  depends only on the remaining variables of  $u$  denoted as  $u_{\bar{\Gamma}}$ . By a change of basis on  $u_{\bar{\Gamma}}$  variables, we can write, for some  $\bar{\Gamma}_0 \subseteq \bar{\Gamma}$ ,

$$\text{Span}(A_{11}) = \text{Span}(\{ u_i + \sigma(q_{\Omega_2}, v) \mid i \in \bar{\Gamma}_0 \}) \oplus (\text{Span}(A_{11}) \cap \text{Span}(q_{\Omega_2}, v)).$$

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  $\Gamma \cup \bar{\Gamma}_0$ ). Thus we reach our **final form**:

$$\begin{aligned}
A_1 &= \{ a_i + y_j + z_k \mid i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \} \\
A_2 &= \{ a_i + y_j + \sigma(q, w) \mid i \in \Sigma_2, j \in \Phi_2 \} \\
A_3 &= \{ a_i + z_k + \sigma(p, u) \mid i \in \Sigma_3, k \in \Psi_3 \} \\
A_4 &= \{ a_i + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5}) \mid i \in \Sigma_4 \} \\
A_5 &= \{ y_j + z_k \mid j \in \Phi_5, k \in \Psi_5 \} \\
A_6 &= \{ y_j + \sigma(q, w) \mid j \in \Phi_6 \} \\
A_{7a} &= \{ z_k + p_s \mid k \in \Psi_{7a}, s \in \Delta'_0 \} \\
A_{7b} &= \{ z_k + \sigma(p_{\Delta_{[2:m]}}) + \sigma(v, u) \mid k \in \Psi_{7b} \}
\end{aligned}$$

$$\begin{aligned}
A_8 &= \{ p_i + \sigma(q) + \sigma(v, w) \mid i \in \Delta_1 \} \\
A_9 &= B_2 \cup \dots \cup B_m \\
A_{10} &= \{ p_i + \sigma(q_{\Omega_1}) + \sigma(v, w) \mid i \in \Omega_1 \cup \Omega_2 \} \\
A_{11} &\subseteq \text{Span}(q_{\Omega_2}, v).
\end{aligned}$$

Here  $B_s = \{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) + \sigma(v, u) \mid j \in \Delta_s \}$ .

$$\begin{aligned}
A_{12} &= \{ u_i + \sigma(q_{\Omega_2}, v, w) \mid i \in \Gamma \} \\
A_{13} &= \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)).
\end{aligned} \tag{20}$$

## 6 Four-wise Correlations: A (somewhat) Simplified Case

We now begin the proof of our main technical Theorem 2.15, i.e. to upper bound the expectation

$$\mathbb{E}_{x \in (\{0,1\}^k)^\ell} [f_{=i}(M_1x)f_{=i}(M_2x)f_{=i}(M_3x)f_{=i}(M_4x)]. \quad (21)$$

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 (19) (ignore Section 5.5 and modifications therein for now):

- All the linear forms  $\sigma(\cdot)$  are zero.
- $A_{10} = A_{11} = A_{13} = \emptyset$ .  $A_9$  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  $H_4 = g = \{g_1, \dots, g_{h_4}\}$  be the rows that appeared in all four matrices (and were removed).
- Let  $H_3 = r = \{r_1, \dots, r_{h_3}\}$  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  $x \in (\{0,1\}^k)^\ell$ , 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 the ease of notation, we drop the prime from the superscript and call the new variable  $w$  as well.

Thus we assume that:

$$\begin{aligned} M_1x &= g, r(1), v, p, u, y \\ M_2x &= g, r(2), v, q, w, z \\ M_3x &= g, r(3), a, p_1 + q_1, \dots, p_n + q_n, u, w \\ M_4x &= g, r(4), A_1, \dots, A_9, A_{12}. \end{aligned} \quad (22)$$

We recall that (in the present special case):

$$\begin{aligned} A_1 &= \{ a_i + y_j + z_k \mid i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \} \\ A_2 &= \{ a_i + y_j \mid i \in \Sigma_2, j \in \Phi_2 \} \\ A_3 &= \{ a_i + z_k \mid i \in \Sigma_3, k \in \Psi_3 \} \\ A_4 &= \{ a_i \mid i \in \Sigma_4 \} \\ A_5 &= \{ y_j + z_k \mid j \in \Phi_5, k \in \Psi_5 \} \\ A_6 &= \{ y_j \mid j \in \Phi_6 \} \\ A_7 &= \{ z_k \mid k \in \Psi_7 \} \\ A_8 &= \{ p_i \mid i \in \Delta_1 \} \\ A_9 &= \{ q_j \mid j \in \Delta_2 \} \end{aligned}$$

$$A_{12} = \{ u_i \mid i \in \Gamma \}.$$

**Lemma 6.1.** *The dimension  $d$  of  $\oplus_{j=1}^4(\text{rowspan}(M_j))$  is:*

$$\begin{aligned} d &= |g| + |r| + |v| + (|p| + |q|) + (|u|) + |w| + (|a| + |y| + |z|) \\ &= |g| + |r| + |v| + (2|\Delta_0| + 2|\Delta_1| + 2|\Delta_2|) + (|\Gamma| + |\bar{\Gamma}|) + |w| + \\ &\quad (3|\Sigma_1| + 2|\Sigma_2| + 2|\Sigma_3| + |\Sigma_4| + 2|\Phi_5| + |\Phi_6| + |\Psi_7|). \end{aligned}$$

*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  $(u_\Gamma, u_{\bar{\Gamma}})$ .  $\square$

We split inputs  $M_1x, \dots, M_4x$  in (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 the ease of notation,  $(a+y+z)_1$  denotes the triples  $\{a_i+y_j+z_k \in A_1\}$  (and similarly).

$L$	$J$	$K$
$M_1x =: \{g, r(1), u_\Gamma\}$	$\{v, p_{\Delta_1 \cup \Delta_2}, u_{\bar{\Gamma}}, y_{\Phi_6}\}$	$\{p_{\Delta_0}, y_{\Phi_1}, y_{\Phi_2}, y_{\Phi_5}\}$
$M_2x =: \{g, r(2)\}$	$\{v, q_{\Delta_1 \cup \Delta_2}, w, z_{\Psi_7}\}$	$\{q_{\Delta_0}, z_{\Psi_1}, z_{\Psi_3}, z_{\Psi_5}\}$
$M_3x =: \{g, r(3), u_\Gamma\}$	$\{a_{\Sigma_4}, (p+q)_{\Delta_1 \cup \Delta_2}, u_{\bar{\Gamma}}, w\}$	$\{a_{\Sigma_1}, a_{\Sigma_2}, a_{\Sigma_3}, (p+q)_{\Delta_0}\}$
$M_4x =: \{g, r(4), u_\Gamma\}$	$\{a_{\Sigma_4}, y_{\Phi_6}, z_{\Psi_7}, p_{\Delta_1}, q_{\Delta_2}\}$	$\{(a+y+z)_1, (a+y)_2, (a+z)_3, (y+z)_5\}$

Denoting the parts in the splits as  $(L_1, J_1, K_1), \dots, (L_4, J_4, K_4)$  respectively, consider the restrictions:

$$\lambda_{1,L_1,J_1}(K_1) = f_{=i}(L_1, J_1, K_1), \dots, \lambda_{4,L_4,J_4}(K_4) = f_{=i}(L_4, J_4, K_4).$$

Dropping the subscripts (but keeping in mind that they are always there), the goal is to upper bound

$$\mathbb{E} [\lambda_1(K_1)\lambda_2(K_2)\lambda_3(K_3)\lambda_4(K_4)], \tag{23}$$

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:

$$\begin{aligned} \lambda_1( & p_{\Delta_0}, & y_{\Phi_1}, & y_{\Phi_2}, & y_{\Phi_5}, & ) \\ \lambda_2( & q_{\Delta_0}, & z_{\Psi_1}, & z_{\Psi_3}, & z_{\Psi_5}, & ) \\ \lambda_3( & a_{\Sigma_1}, & a_{\Sigma_2}, & a_{\Sigma_3}, & (p+q)_{\Delta_0} & ) \\ \lambda_4( & a_{\Sigma_1} + y_{\Phi_1} + z_{\Psi_1}, & a_{\Sigma_2} + y_{\Phi_2}, & a_{\Sigma_3} + z_{\Psi_3}, & y_{\Phi_5} + z_{\Psi_5} & ). \end{aligned}$$

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 (23) equals (there is a product of four terms that are written one below the other for visual ease)

$$\mathbb{E} \left[ \begin{array}{c} \widehat{\lambda}_1(S, W, Y, B) \\ \widehat{\lambda}_2(S, W, Z, B) \\ \widehat{\lambda}_3(W, Y, Z, S) \\ \widehat{\lambda}_4(W, Y, Z, B). \end{array} \right] \quad (24)$$

To explain the reasoning, we note that the Fourier expansion will have a term (as part of a larger product term)

$$\cdots \chi_W(a_{\Sigma_1} + y_{\Phi_1} + z_{\Psi_1}) \chi_{W'}(y_{\Phi_1}) \chi_{W''}(z_{\Psi_1}) \chi_{W'''}(a_{\Sigma_1}) \cdots$$

and taking expectation over  $a_{\Sigma_1}, y_{\Phi_1}, z_{\Psi_1}$ , the term vanishes unless  $W = W' = W'' = W'''$ . Similar reasoning is applied above to “Fourier tuples”  $Y, Z, S, B$ .

For fixed  $L_1, \dots, L_4, W, Y, Z, S, B$ , we consider the expectation over  $J_1, \dots, J_4$  (or rather inputs in those sets). The point here is that all inputs in  $J_1, \dots, J_4$  appear twice:

- Exactly twice, these being  $\{v, u_{\overline{\Gamma}}, y_{\Phi_6}, w, z_{\Psi_7}, a_{\Sigma_4}\}$ .
- 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 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 A.4), we see that (24) is upper bounded by

$$\mathbb{E}_{g,r,u_{\Gamma}} \left[ \begin{array}{c} \sum_{W,Y,Z,S,B} \sqrt{\mathbb{E}_{J_1} [\widehat{\lambda}_{1,J_1}^2(S, W, Y, B)]} \sqrt{\mathbb{E}_{J_2} [\widehat{\lambda}_{2,J_2}^2(S, W, Z, B)]} \\ \sqrt{\mathbb{E}_{J_3} [\widehat{\lambda}_{3,J_3}^2(W, Y, Z, S)]} \sqrt{\mathbb{E}_{J_4} [\widehat{\lambda}_{4,J_4}^2(W, Y, Z, B)]} \end{array} \right].$$

Again applying Cauchy-Schwartz (note that the pairing is first-third and fourth-second factors) we get an upper bound  $\sqrt{\text{Term}_1} \cdot \sqrt{\text{Term}_2}$  where

$$\begin{aligned} \text{Term}_1 &= \mathbb{E}_{g,r,u_{\Gamma}} \left[ \sum_{W,Y,Z,S,B} \mathbb{E}_{J_1} [\widehat{\lambda}_{1,J_1}^2(S, W, Y, B)] \mathbb{E}_{J_3} [\widehat{\lambda}_{3,J_3}^2(W, Y, Z, S)] \right] \\ \text{Term}_2 &= \mathbb{E}_{g,r,u_{\Gamma}} \left[ \sum_{W,Y,Z,S,B} \mathbb{E}_{J_4} [\widehat{\lambda}_{4,J_4}^2(W, Y, Z, B)] \mathbb{E}_{J_2} [\widehat{\lambda}_{2,J_2}^2(S, W, Z, B)] \right]. \end{aligned}$$

We consider  $\text{Term}_1$ . Noting that  $W, Y, S$  appear in both  $\widehat{\lambda}_1(\cdot), \widehat{\lambda}_3(\cdot)$ ,  $B$  appears only in  $\widehat{\lambda}_1(\cdot)$ ,  $Z$  appears only in  $\widehat{\lambda}_3(\cdot)$ , 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:

$$\text{Term}_1 = \mathbb{E}_{g,r,u_\Gamma} \left[ \sum_{W,Y,S} \left( \mathbb{E}_{r(1) \setminus r(3), J_1} \left[ \sum_B \hat{\lambda}_{1,J_1}^2(S, W, Y, B) \right] \right) \left( \mathbb{E}_{J_3} \left[ \sum_Z \hat{\lambda}_{3,J_3}^2(W, Y, Z, S) \right] \right) \right].$$

Finally, using Lemma A.5, we have the upper bound:

$$\text{Term}_1 \leq \left( \max_{g,r(1) \cap r(3), u_\Gamma, r(1) \setminus r(3), J_1} \mathbb{E}_{W,Y,S} \left[ \sum_B \hat{\lambda}_{1,J_1}^2(S, W, Y, B) \right] \right) \left( \mathbb{E}_{g,r,u_\Gamma, J_3} \left[ \sum_{W,Y,Z,S} \hat{\lambda}_{3,J_3}^2(W, Y, Z, S) \right] \right).$$

The second factor is  $\mathbb{E}_{L_3, J_3, K_3} [\|\lambda_{3, L_3, J_3}(K_3)\|_2^2] = \|f_{=i}\|_2^2 \leq \frac{2^{i^2}}{2^{i\ell}} \eta$ . The first factor is bounded by, using Lemma 3.20,  $2^{6i^3} \frac{\varepsilon}{2^{d_1 - \ell}}$  where

$$\begin{aligned} d_1 &= (|J_1| + |r(1) \setminus r(3)|) + 2(|W| + |Y| + |S|) + |B| \\ &= |v| + |\Delta_1| + |\Delta_2| + |\bar{\Gamma}| + |\Phi_6| + |r(1) \setminus r(3)| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0| + |\Phi_5|. \end{aligned}$$

We similarly re-write  $\text{Term}_2$  as:

$$\text{Term}_2 = \mathbb{E}_{g,r,u_\Gamma} \left[ \sum_{W,Z,B} \left( \mathbb{E}_{J_4, u_\Gamma, r(4) \setminus r(2)} \left[ \sum_Y \hat{\lambda}_{4,J_4}^2(W, Y, Z, B) \right] \right) \left( \mathbb{E}_{J_2} \left[ \sum_S \hat{\lambda}_{2,J_2}^2(S, W, Z, B) \right] \right) \right].$$

Here  $\lambda_2(\cdot)$  does not depend on  $u_\Gamma$  and  $r(4) \setminus r(2)$ , so both are pushed inside. As before,

$$\text{Term}_2 \leq \left( \max_{g,r(4) \cap r(2), W,Z,B} \mathbb{E}_{J_4, u_\Gamma, r(4) \setminus r(2)} \left[ \sum_Y \hat{\lambda}_{4,J_4}^2(W, Y, Z, B) \right] \right) \left( \mathbb{E}_{g,r,u_\Gamma, J_2} \left[ \sum_{W,Z,B,S} \hat{\lambda}_{2,J_2}^2(S, W, Z, B) \right] \right).$$

The second factor is  $\mathbb{E}_{L_2, J_2, K_2} [\|\lambda_{2, L_2, J_2}(K_2)\|_2^2] = \|f_{=i}\|_2^2 \leq \frac{2^{i^2}}{2^{i\ell}} \eta$ . The first factor is bounded by, using Lemma 3.20,  $2^{6i^3} \frac{\varepsilon}{2^{d_2 - \ell}}$  where

$$\begin{aligned} d_2 &= (|J_4| + |r(4) \setminus r(2)| + |\Gamma|) + 2(|W| + |Z| + |B|) + |Y| \\ &= |\Sigma_4| + |\Phi_6| + |\Psi_7| + |\Delta_1| + |\Delta_2| + |r(4) \setminus r(2)| + |\Gamma| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Phi_5| + |\Sigma_2|. \end{aligned}$$

The proof of Theorem 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. One gets an upper bound of  $\frac{2^{7r^3}}{2^{d\ell}} \eta \varepsilon$  in Theorem 2.15.

**Lemma 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  $i$  ( $= |L_2 \cup J_2 \cup K_2|$ ):

$$\begin{aligned} d_1 &= |v| + |\Delta_1| + |\Delta_2| + |\bar{\Gamma}| + |\Phi_6| + |r(1) \setminus r(3)| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0| + |\Phi_5|. \\ d_2 &= |\Sigma_4| + |\Phi_6| + |\Psi_7| + |\Delta_1| + |\Delta_2| + |r(4) \setminus r(2)| + |\Gamma| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Phi_5| + |\Sigma_2|. \\ i &= |g| + |r(3)| + |\Gamma| + |\Sigma_4| + |\Delta_1| + |\Delta_2| + |\bar{\Gamma}| + |w| + |\Sigma_1| + |\Sigma_2| + |\Sigma_3| + |\Delta_0|. \\ i &= |g| + |r(2)| + |v| + |\Delta_1| + |\Delta_2| + |w| + |\Psi_7| + |\Delta_0| + |\Sigma_1| + |\Sigma_3| + |\Phi_5|. \end{aligned}$$

It can be verified that the overall sum is exactly  $2d$  where as in Lemma 6.1,

$$d = |g| + |r| + |v| + 2|\Delta_0| + 2|\Delta_1| + 2|\Delta_2| + |\Gamma| + |\bar{\Gamma}| + |w| + 3|\Sigma_1| + 2|\Sigma_2| + 2|\Sigma_3| + |\Sigma_4| + 2|\Phi_5| + |\Phi_6| + |\Psi_7|.$$

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,  $|r| = |r(3)| + |r(1) \setminus r(3)| = |r(2)| + |r(4) \setminus r(2)|$ .  $\square$

## 7 Four-wise Correlations: the General Case

We now begin the full proof of our main technical Theorem 2.15, i.e. to upper bound the expectation

$$\mathbb{E}_{x \in (\{0,1\}^k)^\ell} [f_{=i}(M_1x)f_{=i}(M_2x)f_{=i}(M_3x)f_{=i}(M_4x)]. \quad (25)$$

Given matrices  $M_1, M_2, M_3, M_4$ , we recall:

- Let  $H_4 = g = \{g_1, \dots, g_{h_4}\}$  be the rows that appeared in all four matrices (and were removed).
- Let  $H_3 = r = \{r_1, \dots, r_{h_3}\}$  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  $x \in (\{0,1\}^k)^\ell$ , 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 the ease of notation, we drop the prime from the superscript and call the new variable  $w$  as well.

Thus we assume that (given the basis for  $\text{rowspan}(M_4)$  by (20), written again below for convenience):

$$\begin{aligned} M_1x &= g, r(1), v, p, u, y \\ M_2x &= g, r(2), v, q, w, z \\ M_3x &= g, r(3), a, p_1 + q_1, \dots, p_n + q_n, u, w \\ M_4x &= g, r(4), A_1, \dots, A_6, A_{7a}, A_{7b}, A_8, \dots, A_{13}. \end{aligned} \quad (26)$$

**Lemma 7.1.** *The dimension  $d$  of  $\oplus_{j=1}^4(\text{rowspan}(M_j))$  is:*

$$\begin{aligned} d &= |g| + |r| + |v| + (|p| + |q|) + (|u|) + |w| + (|a| + |y| + |z|) \\ &= |g| + |r| + |v| + (2|\Delta'_0| + 2|\Delta''_0| + 2|\Delta_1| + 2|\Delta_2| + \dots + 2|\Delta_m| + 2|\Omega_1| + 2|\Omega_2|) + (|\Gamma| + |\bar{\Gamma}|) + |w| + \\ &\quad (3|\Sigma_1| + 2|\Sigma_2| + 2|\Sigma_3| + |\Sigma_4| + 2|\Phi_5| + |\Phi_6| + |\Psi_{7a}| + |\Psi_{7b}|). \end{aligned}$$

*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  $|\Delta'_0| + |\Delta''_0| + |\Delta_1| + |\Delta_2| + \dots + |\Delta_m| + |\Omega_1| + |\Omega_2|$  and  $\Delta_0 = \Delta'_0 \cup \Delta''_0$ . We have  $|\Sigma_1| = |\Phi_1| = |\Psi_1|$  and similar equalities. We emphasize that  $|\Delta'_0| = |\Psi_{7a}|$ .  $\square$

We recall for convenience that:

$$\begin{array}{l}
A_1 = \{ a_i + y_j + z_k \quad | \quad i \in \Sigma_1, j \in \Phi_1, k \in \Psi_1 \} \\
A_2 = \{ a_i + y_j \quad + \sigma(q, w) \quad | \quad i \in \Sigma_2, j \in \Phi_2 \quad \} \\
A_3 = \{ a_i \quad + z_k \quad + \sigma(p, u) \quad | \quad i \in \Sigma_3, \quad k \in \Psi_3 \} \\
A_4 = \{ a_i \quad + \sigma(v, p, u, q, w) + \sigma(y_{\Phi_1, \Phi_5}) \quad | \quad i \in \Sigma_4 \quad \} \\
A_5 = \{ \quad y_j + z_k \quad | \quad \quad \quad j \in \Phi_5, k \in \Psi_5 \} \\
A_6 = \{ \quad y_j \quad + \sigma(q, w) \quad | \quad \quad \quad j \in \Phi_6 \quad \} \\
A_{7a} = \{ \quad \quad z_k + p_s \quad | \quad \quad \quad k \in \Psi_{7a}, s \in \Delta'_0 \} \\
A_{7b} = \{ \quad \quad z_k \quad + \sigma(p_{\Delta_{[2:m]}}, u) \quad | \quad \quad \quad k \in \Psi_{7b} \}
\end{array}$$

$$\begin{array}{l}
A_8 = \{ p_i + \sigma(q) \quad + \sigma(v, w) \quad | \quad i \in \Delta_1 \quad \} \\
A_9 = B_2 \cup \dots \cup B_m \\
A_{10} = \{ p_i + \sigma(q_{\Omega_1}) \quad + \sigma(v, w) \quad | \quad i \in \Omega_1 \cup \Omega_2 \quad \} \\
A_{11} \subseteq \text{Span}(q_{\Omega_2}, v).
\end{array}$$

Here  $B_s = \{ q_j + \sigma(p_{\Delta_{[s+1:m]}}) + \sigma(v, u) \mid j \in \Delta_s \}$ .

$$\begin{array}{l}
A_{12} = \{ u_i + \sigma(q_{\Omega_2}, v, w) \mid i \in \Gamma \} \\
A_{13} = \text{basis}(\text{rowspan}(M_4) \cap \text{Span}(v, w)).
\end{array}$$

We split each input in (26) into two parts. The mix of Fourier analysis and Cauchy-Schwartz will not be very clean. The splits are as below.

$J$	$K$
$M_1x =: \{g, r(1), v, p_{\Delta_{[2:m]}}, u_{\Gamma}, u_{\bar{\Gamma}}\}$	$\{p_{\Delta_0 \cup \Delta_1}, y_{\Phi_1}, y_{\Phi_2}, y_{\Phi_5}, y_{\Phi_6}\}$
$M_2x =: \{g, r(2), v, q_{\Delta_{[2:m]}}, w, z_{\Psi_{7b}}\}$	$\{q_{\Delta_0 \cup \Delta_1}, z_{\Psi_1}, z_{\Psi_3}, z_{\Psi_5}, z_{\Psi_{7a}}\}$
$M_3x =: \{g, r(3), (p+q)_{\Delta_{[2:m]}}, u_{\Gamma}, u_{\bar{\Gamma}}, w\}$	$\{a, (p+q)_{\Delta_0 \cup \Delta_1}\}$
$M_4x =: \{g, r(4), z_{\Psi_{7b}}, A_9, A_{10}, A_{11}, A_{12}, A_{13}\}$	$\{(a+y+z)_1, (a+y)_2, (a+z)_3, a_{\Sigma_4}, (y+z)_5, y_{\Phi_6}, z_{\Psi_{7a}} + p_{\Delta'_0}, p_{\Delta_1}\}$ .

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 for ease. Denoting the parts in the splits as  $(J_1, K_1), \dots, (J_4, K_4)$  respectively, consider the restrictions:

$$\lambda_{1, J_1}(K_1) = f_{=i}(J_1, K_1), \dots, \lambda_{4, J_4}(K_4) = f_{=i}(J_4, K_4).$$

Dropping the subscripts (but keeping in mind that they are always there), the goal is to upper bound

$$\mathbb{E} [\lambda_1(K_1) \lambda_2(K_2) \lambda_3(K_3) \lambda_4(K_4)], \tag{27}$$

where for notational ease, 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:

$$\begin{aligned}
\lambda_1( & p_{\Delta'_0}, & p_{\Delta''_0}, & p_{\Delta_1}, & y_{\Phi_1}, & y_{\Phi_2}, & y_{\Phi_5}, & y_{\Phi_6} & ) \\
\lambda_2( & q_{\Delta'_0}, & q_{\Delta''_0}, & q_{\Delta_1}, & z_{\Psi_1}, & z_{\Psi_3}, & z_{\Psi_5}, & z_{\Psi_{7a}} & ) \\
\lambda_3( & a_{\Sigma_1}, & a_{\Sigma_2}, & a_{\Sigma_3}, & a_{\Sigma_4}, & (p+q)_{\Delta'_0}, & (p+q)_{\Delta''_0}, & (p+q)_{\Delta_1} & ) \\
\lambda_4( & a_{\Sigma_1} + y_{\Phi_1} + z_{\Psi_1}, & a_{\Sigma_2} + y_{\Phi_2}, & a_{\Sigma_3} + z_{\Psi_3}, & a_{\Sigma_4}, & y_{\Phi_5} + z_{\Psi_5}, & y_{\Phi_6}, & z_{\Psi_{7a}} + p_{\Delta'_0}, & p_{\Delta_1} & ).
\end{aligned}$$

Now writing the  $\lambda_s$  in the Fourier representation and taking expectation over their inputs, we see that the expectation in (27) equals, up to a caveat to be fixed shortly, (there is a product of four terms that are written one below the other for visual ease)

$$\mathbb{E}_{\substack{g,r,v,u_{\Gamma},u_{\bar{\Gamma}},w \\ p_{\Delta[2:m]},q_{\Delta[2:m]},z_{\Psi_{7b}}}} \left[ \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \text{sign} \cdot \begin{array}{l} \widehat{\lambda}_1( S+Q, D, X+N, W, Y, P, T ) \\ \widehat{\lambda}_2( S, D, X, W, Z, P, Q ) \\ \widehat{\lambda}_3( W, Y, Z, B, S, D, X ) \\ \widehat{\lambda}_4( W, Y, Z, B, P, T, Q, N ) \end{array} \right] \quad (28)$$

A remark: there are  $\sigma(\cdot)$  terms that were omitted from the notation. They have a two-fold effect. Firstly, there is a  $\text{sign} \in \{-1, 1\}$  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 (28), but actually as below:

$$\begin{aligned}
\widehat{\lambda}_1( & S+Q+ & D+ & X+N+ & W+ & Y, & P & T & ) \\
& \sigma(B, Z), & \sigma(B, Z), & \sigma(B, Z), & \sigma(B), & & +\sigma(B), & & ) \\
\widehat{\lambda}_2( & S+ & D+ & X+ & W, & Z, & P, & Q & ) \\
& \sigma(Y, B, T, N), & \sigma(Y, B, T, N), & \sigma(Y, B, T, N), & & & & & ) \\
\widehat{\lambda}_3( & W, & Y, & Z, & B, & S, & D, & X & ) \\
\widehat{\lambda}_4( & W, & Y, & Z, & B, & P, & T, & Q, & N & ).
\end{aligned} \quad (29)$$

Denoting the Fourier coefficients as  $\widehat{\lambda}_1(V_1), \widehat{\lambda}_2(V_2), \widehat{\lambda}_3(V_3), \widehat{\lambda}_4(V_4)$ , an upper bound on the desired expectation is:

$$\begin{aligned}
& \mathbb{E}_{\substack{g,r,v,u_\Gamma,u_{\bar{\Gamma}},w,z_{\Psi_{7b}} \\ p_{\Delta[2:m]},q_{\Delta[2:m]}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ |\widehat{\lambda}_1(V_1)| \cdot |\widehat{\lambda}_2(V_2)| \cdot |\widehat{\lambda}_3(V_3)| \cdot |\widehat{\lambda}_4(V_4)| \right] \\
&= \mathbb{E}_{\substack{g,r,v,u_\Gamma,u_{\bar{\Gamma}},w \\ p_{\Delta[2:m]},q_{\Delta[2:m]}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \mathbb{E}_{z_{\Psi_{7b}}} \left[ |\widehat{\lambda}_1(V_1)| \cdot |\widehat{\lambda}_2(V_2)| \cdot |\widehat{\lambda}_3(V_3)| \cdot |\widehat{\lambda}_4(V_4)| \right].
\end{aligned}$$

We note that  $z_{\Psi_{7a}}$  appears only in  $J_2, J_4$ . Using Cauchy-Schwartz, we get an upper bound

$$\mathbb{E}_{\substack{g,r,v,u_\Gamma,u_{\bar{\Gamma}},w \\ p_{\Delta[2:m]},q_{\Delta[2:m]}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ |\widehat{\lambda}_1(V_1)| \sqrt{\mathbb{E}_{z_{\Psi_{7b}}} [\widehat{\lambda}_2^2(V_2)]} |\widehat{\lambda}_3(V_3)| \sqrt{\mathbb{E}_{z_{\Psi_{7b}}} [\widehat{\lambda}_4^2(V_4)]} \right].$$

A point to note here is as follows: in  $\lambda_4$ , the variables  $z_{\Psi_{7b}}$  actually appear along with additional  $\sigma(p_{\Delta[2:m]}, u)$  terms. However the expectation over these additional variables is still not considered and is still at the ‘‘outer’’ level. Hence the Cauchy-Schwartz over  $z_{\Psi_{7b}}$  can be safely applied. Moreover, once Cauchy-Schwartz, i.e. expectation over  $z_{\Psi_{7b}}$ , is applied, we can ignore these  $\sigma(\cdot)$  terms henceforth.<sup>12</sup> We will use this trick repeatedly.

Next, we consider the variables  $(p_{\Delta_2}, q_{\Delta_2}), \dots, (p_{\Delta_m}, q_{\Delta_m})$ , 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 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-Schwartz 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

$$\mathbb{E}_{\substack{g,r,v,u_\Gamma,u_{\bar{\Gamma}},w \\ p_{\Delta[3:m]},q_{\Delta[3:m]}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ \sqrt{\mathbb{E}_{p_{\Delta_2}} [\widehat{\lambda}_1^2(V_1)]} \sqrt{\mathbb{E}_{z_{\Psi_{7b}},q_{\Delta_2}} [\widehat{\lambda}_2^2(V_2)]} \sqrt{\mathbb{E}_{p_{\Delta_2}} [\widehat{\lambda}_3^2(V_3)]} \sqrt{\mathbb{E}_{z_{\Psi_{7b}},q_{\Delta_2}} [\widehat{\lambda}_4^2(V_4)]} \right].$$

We apply the same argument iteratively to get an upper bound

$$\mathbb{E}_{\substack{g,r,v,u_\Gamma,u_{\bar{\Gamma}},w \\ p_{\Omega_1 \cup \Omega_2}, q_{\Omega_1 \cup \Omega_2}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ \sqrt{\mathbb{E}_{p_{\Delta_2, \dots, m}} [\widehat{\lambda}_1^2(V_1)]} \sqrt{\mathbb{E}_{z_{\Psi_{7b}}, q_{\Delta_2, \dots, m}} [\widehat{\lambda}_2^2(V_2)]} \sqrt{\mathbb{E}_{p_{\Delta_2, \dots, m}} [\widehat{\lambda}_3^2(V_3)]} \sqrt{\mathbb{E}_{z_{\Psi_{7b}}, q_{\Delta_2, \dots, m}} [\widehat{\lambda}_4^2(V_4)]} \right].$$

Next, we Cauchy-Schwartz over  $u_{\bar{\Gamma}}$ . 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

$$\mathbb{E}_{\substack{g,r,v,u_\Gamma,u_{\bar{\Gamma}},w \\ p_{\Omega_1 \cup \Omega_2}, q_{\Omega_1 \cup \Omega_2}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ \sqrt{\mathbb{E}_{u_{\bar{\Gamma}}} [\widehat{\lambda}_1^2(V_1)]} \sqrt{\mathbb{E}_{z_{\Psi_{7b}}, q_{\Delta_2, \dots, m}} [\widehat{\lambda}_2^2(V_2)]} \sqrt{\mathbb{E}_{u_{\bar{\Gamma}}} [\widehat{\lambda}_3^2(V_3)]} \sqrt{\mathbb{E}_{z_{\Psi_{7b}}, q_{\Delta_2, \dots, m}} [\widehat{\lambda}_4^2(V_4)]} \right].$$

<sup>12</sup> Formally, if one wishes to, by change of variables  $z_{\Psi_{7b}} \leftarrow z_{\Psi_{7b}} + \sigma(p_{\Delta[2:m]}, u)$ .

Finally, we apply Cauchy-Schwartz 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

$$\begin{aligned} \text{Term}_1 &= \mathbb{E}_{\substack{g,r,v,u_\Gamma,w \\ p_{\Omega_1 \cup \Omega_2}, q_{\Omega_1 \cup \Omega_2}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ \mathbb{E}_{\substack{u_\Gamma \\ p_{\Delta_2, \dots, m}}} \left[ \widehat{\lambda}_1^2(V_1) \right] \cdot \mathbb{E}_{\substack{u_\Gamma \\ p_{\Delta_2, \dots, m}}} \left[ \widehat{\lambda}_3^2(V_3) \right] \right] \\ \text{Term}_2 &= \mathbb{E}_{\substack{g,r,v,u_\Gamma,w \\ p_{\Omega_1 \cup \Omega_2}, q_{\Omega_1 \cup \Omega_2}}} \sum_{\substack{W,Y,Z,B,P \\ T,Q,N,S,D,X}} \left[ \mathbb{E}_{\substack{z_{\Psi_{7b}} \\ q_{\Delta_2, \dots, m}}} \left[ \widehat{\lambda}_4^2(V_4) \right] \cdot \mathbb{E}_{\substack{z_{\Psi_{7b}} \\ q_{\Delta_2, \dots, m}}} \left[ \widehat{\lambda}_2^2(V_2) \right] \right]. \end{aligned} \quad (30)$$

**Lemma 7.2.** *We have the upper bound  $\text{Term}_1 \leq 2^{7i^3} \frac{\eta \varepsilon}{2^{(d_1+i) \cdot \ell}}$  where*

$$d_1 = |r(1) \setminus r(3)| + |\Delta_{2, \dots, m}| + |\Omega_1| + |\Omega_2| + |v| + |\bar{\Gamma}| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta_0''| + |\Psi_{7a}| + |\Delta_1| + |\Phi_5| + |\Phi_6|.$$

*Proof.* Let us recall the definitions of  $V_1, V_3$ :

$$\begin{aligned} V_1 &= (S + Q + \sigma(B, Z), D + \sigma(B, Z), X + N + \sigma(B, Z), W + \sigma(B), P + \sigma(B), Y, T), \\ V_3 &= (W, Y, Z, B, S, D, X). \end{aligned}$$

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

$$V_1 = (Q, D + \sigma(B, Z), N, W + \sigma(B), P, Y, T).$$

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

$$\begin{aligned} \text{Term}_1 &= \mathbb{E}_{\substack{g, u_\Gamma \\ r(3)}} \sum_{W, Y, D} \left[ \left( \mathbb{E}_{\substack{v, r(1) \setminus r(3), p_{\Omega_1 \cup \Omega_2} \\ p_{\Delta_2, \dots, m}, u_\Gamma}} \left[ \sum_{Q, N, P, T} \widehat{\lambda}_1^2(W, Y, D, Q, N, P, T) \right] \right) \right. \\ &\quad \left. \left( \mathbb{E}_{\substack{q_{\Omega_1 \cup \Omega_2}, w \\ p_{\Delta_2, \dots, m}, u_\Gamma}} \left[ \sum_{B, Z, S, X} \widehat{\lambda}_3^2(W, Y, D, B, Z, S, X) \right] \right) \right]. \end{aligned}$$

The vigilant reader must have noticed that we have pushed several expectations ‘‘inside’’. This is justified as follows.  $\lambda_1$  does not depend on  $q_{\Omega_1 \cup \Omega_2}$  and  $w$ .  $\lambda_3$  does not depend on  $v, r(1) \setminus r(3)$ , and since it depends only on  $(p + q)_{\Omega_1 \cup \Omega_2}$ , it is ‘‘effectively’’ independent of  $p_{\Omega_1 \cup \Omega_2}$ . Using Lemma A.5,  $\text{Term}_1$  is bounded by

$$\left( \max_{\substack{g, r(1) \cap r(3), u_\Gamma \\ W, Y, D}} \mathbb{E}_{\substack{v, r(1) \setminus r(3), p_{\Omega_1 \cup \Omega_2} \\ p_{\Delta_2, \dots, m}, u_\Gamma}} \left[ \sum_{\substack{Q, N, \\ P, T}} \lambda_1^2(W, Y, D, Q, N, P, T) \right] \right) \left( \mathbb{E}_{\substack{W, Y, D, \\ B, Z, S, X}} \left[ \sum \widehat{\lambda}_3^2(W, Y, D, B, Z, S, X) \right] \right).$$

The second factor equals (as usual)  $\|f_{=i}\|_2^2 \leq \frac{2^{i^2}}{2^{i\ell}} \eta$ . The first factor is bounded, using Lemma 3.20, by  $2^{6i^3} \frac{\varepsilon}{2^{d_1 \cdot \ell}}$  where

$$\begin{aligned} d_1 &= |r(1) \setminus r(3)| + (|\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2|) + |v| + |\bar{\Gamma}| + (2|W| + 2|Y| + 2|D|) + |Q| + |N| + |P| + |T| \\ &= |r(1) \setminus r(3)| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |v| + |\bar{\Gamma}| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta''_0| + |\Psi_{7a}| + |\Delta_1| + |\Phi_5| + |\Phi_6|. \end{aligned}$$

□

**Lemma 7.3.** *We have the upper bound  $\text{Term}_2 \leq 2^{7i^3} \frac{\eta\varepsilon}{2^{d_2 \cdot \ell}}$  where*

$$\begin{aligned} d_2 &= |r(4) \setminus r(2)| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |\Gamma| + |\Psi_{7b}| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Phi_5| + 2|\Psi_{7a}| + \\ &\quad |\Sigma_2| + |\Sigma_4| + |\Phi_6| + |\Delta_1|. \end{aligned}$$

*Proof.* Let us recall the the definitions of  $V_2, V_4$ .

$$V_2 = (S + \sigma(Y, B, T, N), D + \sigma(Y, B, T, N), X + \sigma(Y, B, T, N), W, Z, P, Q),$$

$$V_4 = (W, Y, Z, B, P, T, Q, N).$$

Since  $S, D, X$  do not appear in  $V_4$ , 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

$$\begin{aligned} \text{Term}_2 &= \mathbb{E}_{\substack{g,v,r(2),w \\ q_{\Omega_1 \cup \Omega_2}}} \sum_{W,Z,P,Q} \left[ \left( \mathbb{E}_{\substack{r(4) \setminus r(2), u_{\Gamma}, p_{\Omega_1 \cup \Omega_2} \\ z_{\Psi_{7b}, q_{\Delta_{2,\dots,m}}}}} \left[ \sum_{Y,B,T,N} \widehat{\lambda}_4^2(W, Z, P, Q, Y, B, T, N) \right] \right) \right. \\ &\quad \left. \left( \mathbb{E}_{z_{\Psi_{7b}, q_{\Delta_{2,\dots,m}}}} \left[ \sum_{S,D,X} \widehat{\lambda}_2^2(W, Z, P, Q, S, D, X) \right] \right) \right]. \end{aligned}$$

We have pushed the expectation over  $r(4) \setminus r(2), u_{\Gamma}, p_{\Omega_1 \cup \Omega_2}$  inside as  $\lambda_2$  does not depend on them. Using Lemma A.5, we upper bound  $\text{Term}_2$  by

$$\left( \max_{\substack{g,v,r(4) \cap r(2), w \\ q_{\Omega_1 \cup \Omega_2}, W, Z, P, Q}} \mathbb{E}_{\substack{r(4) \setminus r(2), u_{\Gamma}, p_{\Delta_{\Omega_1 \cup \Omega_2}} \\ z_{\Psi_{7b}, q_{\Delta_{2,\dots,m}}}}} \left[ \sum_{\substack{Y,B \\ T,N}} \widehat{\lambda}_4^2 \left( \begin{matrix} W, Z, P, Q, \\ Y, B, T, N \end{matrix} \right) \right] \right) \left( \mathbb{E}_{\substack{W,Z,P,Q \\ S,D,X}} \left[ \sum \widehat{\lambda}_2^2 \left( \begin{matrix} W, Z, P, Q, \\ S, D, X \end{matrix} \right) \right] \right).$$

As before, the second factor equals  $\|f_{=i}\|_2^2 \leq \frac{2^{i^2}}{2^{i\ell}} \eta$ . The first factor is bounded, using Lemma 3.20, by  $2^{6i^3} \frac{\varepsilon}{2^{d_2 \cdot \ell}}$  where

$$\begin{aligned} d_2 &= |r(4) \setminus r(2)| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |\Gamma| + |\Psi_{7b}| + (2|W| + 2|Z| + 2|P| + 2|Q|) + \\ &\quad |Y| + |B| + |T| + |N| \\ &= |r(4) \setminus r(2)| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |\Gamma| + |\Psi_{7b}| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Phi_5| + 2|\Psi_{7a}| + \\ &\quad |\Sigma_2| + |\Sigma_4| + |\Phi_6| + |\Delta_1|. \end{aligned}$$

□

We get the overall upper bound  $\sqrt{\text{Term}_1}\sqrt{\text{Term}_2}$ , which is at most  $2^{\tilde{r}r^3 \frac{\eta\epsilon}{2d\ell}}$  (noting  $i \leq r$ ) provided  $\frac{1}{2}((d_1 + i) + (d_2 + i)) = d$ . This is proved below completing the proof of Theorem 2.15.

**Lemma 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  $M_2$ ):

$$\begin{aligned} d_1 &= |r(1) \setminus r(3)| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |v| + |\bar{\Gamma}| + 2|\Sigma_1| + 2|\Sigma_2| + 2|\Delta'_0| + |\Psi_{7a}| + |\Delta_1| + |\Phi_5| + |\Phi_6|. \\ d_2 &= |r(4) \setminus r(2)| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |\Gamma| + |\Psi_{7b}| + 2|\Sigma_1| + 2|\Sigma_3| + 2|\Phi_5| + 2|\Psi_{7a}| + |\Sigma_2| + |\Sigma_4| + |\Phi_6| + |\Delta_1|. \\ i &= |g| + |r(3)| + |\Sigma_1| + |\Sigma_2| + |\Sigma_3| + |\Sigma_4| + |\Delta'_0| + |\Delta''_0| + |\Delta_1| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |\Gamma| + |\bar{\Gamma}| + |w|. \\ i &= |g| + |r(2)| + |v| + |\Delta'_0| + |\Delta''_0| + |\Delta_1| + |\Delta_{2,\dots,m}| + |\Omega_1| + |\Omega_2| + |w| + |\Sigma_1| + |\Sigma_3| + |\Phi_5| + |\Psi_{7a}| + |\Psi_{7b}|. \end{aligned}$$

It can be verified that the overall sum is exactly  $2d$  where as in Lemma 7.1,

$$\begin{aligned} d &= |g| + |r| + |v| + (2|\Delta'_0| + 2|\Delta''_0| + 2|\Delta_1| + 2|\Delta_{2,\dots,m}| + 2|\Omega_1| + 2|\Omega_2|) + (|\Gamma| + |\bar{\Gamma}|) + |w| + (3|\Sigma_1| + 2|\Sigma_2| + 2|\Sigma_3| + |\Sigma_4| + 2|\Phi_5| + |\Phi_6| + |\Psi_{7a}| + |\Psi_{7b}|). \end{aligned}$$

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  $|\Psi_{7a}| = |\Delta'_0|$ .  $\square$

## 8 Acknowledgement

Our most sincere thanks to Boaz Barak, Irit Dinur, Yuval Filmus, Guy Kindler, Pravesh Kothari, Dana Moshkovitz, Prasad Raghavendra, Ran Raz, and David Steurer for several discussions and collaborations that led to this work.

## References

- [1] S. Arora and C. Lund. *Approximation Algorithms for NP-hard Problems*, editor : D. Hochbaum. PWS Publishing, 1996.
- [2] S. Arora, C. Lund, R. Motawani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.
- [3] S. Arora and S. Safra. Probabilistic checking of proofs : A new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.
- [4] Sanjeev Arora, Boaz Barak, and David Steurer. Subexponential algorithms for unique games and related problems. *J. ACM*, 62(5):42:1–42:25, 2015.

- [5] Sanjeev Arora, Subhash Khot, Alexandra Kolla, David Steurer, Madhur Tulsiani, and Nisheeth K. Vishnoi. Unique games on expanding constraint graphs are easy. In *Proc. ACM Symposium on the Theory of Computing*, pages 21–28, 2008.
- [6] Boaz Barak. The intermediate complexity conjecture. Blog post, <https://windowsontheory.org/2017/08/09/the-intermediate-complexity-conjecture/>.
- [7] Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 307–326, 2012.
- [8] Boaz Barak, Parikshit Gopalan, Johan Håstad, Raghu Meka, Prasad Raghavendra, and David Steurer. Making the long code shorter. *SIAM J. Comput.*, 44(5):1287–1324, 2015.
- [9] Boaz Barak, Pravesh Kothari, and David Steurer. Small-set expansion in shortcode graph and the 2-to-1 conjecture. *ECCC Report TR18-077*.
- [10] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *FOCS*, pages 472–481, 2011.
- [11] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability. *Electronic Colloquium on Computational Complexity, Technical Report TR95-024*, 1995.
- [12] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *Proc. 20th IEEE Conference on Computational Complexity*, pages 144–153, 2005.
- [13] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? *Electronic Colloquium on Computational Complexity (ECCC)*, 23:198, 2016.
- [14] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in Grassmann graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:94, 2017.
- [15] Irit Dinur, Elchanan Mossel, and Oded Regev. Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009.
- [16] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Ann. of Math. (2)*, 162(1):439–485, 2005.
- [17] U. Feige. A threshold of  $\ln n$  for approximating set cover. *Journal of the ACM*, 45(4):634–652, 1998.
- [18] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996.
- [19] U. Feige and L. Lovász. Two-prover one-round proof systems, their power and their problems. In *Proc. 24th Annual ACM Symposium on Theory of Computing*, pages 733–744, 1992.
- [20] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

- [21] Venkatesan Guruswami, Rajsekar Manokaran, and Prasad Raghavendra. Beating the random ordering is hard: Inapproximability of maximum acyclic subgraph. In *Proc. Annual IEEE Symposium on Foundations of Computer Science*, pages 573–582, 2008.
- [22] Venkatesan Guruswami and Ali Kemal Sinop. Improved inapproximability results for maximum  $k$ -colorable subgraph. *Theory of Computing*, 9:413–435, 2013.
- [23] J. Håstad. Clique is hard to approximate within  $n^{1-\epsilon}$ . *Acta Mathematica*, 182:105–142, 1999.
- [24] J. Håstad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.
- [25] Johan Håstad. On the Efficient Approximability of Constraint Satisfaction Problems. In *Surveys in Combinatorics*, volume 346, pages 201–222. Cambridge University Press, 2007.
- [26] Thomas Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proc. ACM Symposium on the Theory of Computing*, pages 411–419, 2007.
- [27] Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972.
- [28] S. Khot. Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry. In *Proc. the International Congress of Mathematicians*, 2010.
- [29] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 767–775, 2002.
- [30] Subhash Khot. On the unique games conjecture (invited survey). In *IEEE Conference on Computational Complexity*, pages 99–121, 2010.
- [31] Subhash Khot. Hardness of approximation. In *Proc. of the International Congress of Mathematicians*, 2014.
- [32] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM J. Comput.*, 37(1):319–357, 2007.
- [33] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Pseudorandom sets in Johnson graph have near-perfect expansion. *ECCC Report TR18-078*.
- [34] Subhash Khot, Dor Minzer, and Muli Safra. On independent sets, 2-to-2 games, and Grassmann graphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 576–589, 2017.
- [35] Subhash Khot and Dana Moshkovitz. NP-hardness of approximately solving linear equations over reals. *SIAM J. Comput.*, 42(3):752–791, 2013.
- [36] Subhash Khot and Dana Moshkovitz. Candidate hard unique game. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 63–76, 2016.

- [37] Subhash Khot and Ryan O’Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory of Computing*, 5(1):83–117, 2009.
- [38] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within  $2 - \epsilon$ . *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.
- [39] Subhash Khot and Muli Safra. A two prover one round game with strong soundness. In *FOCS*, pages 648–657, 2011.
- [40] Subhash Khot and Nisheeth K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into  $l_1$ . *J. ACM*, 62(1):8:1–8:39, 2015.
- [41] Alexandra Kolla, Konstantin Makarychev, and Yury Makarychev. How to play unique games against a semi-random adversary: Study of semi-random models of unique games. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 443–452, 2011.
- [42] P. Raghavendra and D. Steurer. Graph expansion and the unique games conjecture. In *Proc. 42nd ACM Symposium on Theory of Computing*, 2010.
- [43] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 245–254, 2008.
- [44] Anup Rao. Parallel repetition in projection games and a concentration bound. In *Proc. ACM Symposium on the Theory of Computing*, pages 1–10, 2008.
- [45] R. Raz. A parallel repetition theorem. *SIAM J. of Computing*, 27(3):763–803, 1998.
- [46] Grant Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 593–602, 2008.
- [47] L. Trevisan. Inapproximability of combinatorial optimization problems. *Optimisation Combinatoire 2*, (Vangelis Paschos, Editor), Hermes, 2005.
- [48] Luca Trevisan. On Khot’s unique games conjecture. *Bull. Amer. Math. Soc. (N.S.)*, 49(1):91–111, 2012.

## A Auxiliary Lemmas

**Lemma A.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:*

- $a_1, \dots, a_p, a'_1, \dots, a'_r$  are in  $A$  and are linearly independent.
- $b_1, \dots, b_q, b'_1, \dots, b'_r$  are in  $B$  and are linearly independent.
- $(a_1, \dots, a_p, b_1, \dots, b_q, a'_1 + b'_1, \dots, a'_r + b'_r)$  is a basis for  $C$ .

Moreover:

- If in addition,  $A \subseteq B \oplus C, B \subseteq A \oplus C,$ 
  - $a_1, \dots, a_p, a'_1, \dots, a'_r$  is already a basis for  $A.$
  - $b_1, \dots, b_q, b'_1, \dots, b'_r$  is already a basis for  $B.$
- Otherwise, the sets can (clearly) be extended further so that
  - $a_1, \dots, a_p, a'_1, \dots, a'_r, a''_1, \dots, a''_m$  is a basis for  $A.$
  - $b_1, \dots, b_q, b'_1, \dots, b'_r, b''_1, \dots, b''_n$  is a basis for  $B.$

*Proof.* Let  $(a_1, \dots, a_p)$  be a basis for  $A \cap C$  and  $(b_1, \dots, b_q)$  be a basis for  $B \cap C.$  Let  $c_1, \dots, c_r \in C$  be such that  $(a_1, \dots, a_p, b_1, \dots, b_q, c_1, \dots, c_r)$  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  $a_1, \dots, a_p, a'_1, \dots, a'_r$  are linearly independent. Suppose (on the contrary) that for some index sets  $\Phi \subseteq \{1, \dots, p\}$  and  $\Psi \subseteq \{1, \dots, r\},$  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  $\Sigma \subseteq \{1, \dots, q\}$  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  $a_1, \dots, a_p, b_1, \dots, b_q, a'_1 + b'_1, \dots, a'_r + b'_r$  is a basis for  $C$  and hence linearly independent.

Finally, we show that if  $A \subseteq B \oplus C,$  then  $a_1, \dots, a_p, a'_1, \dots, a'_r$  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  $a = \bigoplus_{j \in \Phi} a_j \bigoplus_{j \in \Psi} a'_j.$  □

**Lemma A.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  $\cup_{s=1}^7 A_s$  where

$$\begin{array}{l} A_1 = \{ a_i + y_j + z_k \quad | \quad i \in \Sigma_1, \quad j \in \Phi_1, \quad k \in \Psi_1 \} \\ A_2 = \{ a_i + y_j \quad | \quad i \in \Sigma_2, \quad j \in \Phi_2 \quad \quad \quad \} \\ A_3 = \{ a_i \quad + z_k \quad | \quad i \in \Sigma_3, \quad \quad \quad \quad \quad \quad \quad k \in \Psi_3 \} \\ A_4 = \{ a_i \quad \quad \quad + \sigma \quad | \quad i \in \Sigma_4 \quad \quad \quad \quad \quad \quad \quad \} \\ A_5 = \{ \quad \quad \quad y_j + z_k \quad | \quad j \in \Phi_5, \quad k \in \Psi_5 \} \\ A_6 = \{ \quad \quad \quad y_j \quad \quad \quad | \quad j \in \Phi_6 \quad \quad \quad \} \\ A_7 = \{ \quad \quad \quad \quad \quad \quad z_k \quad \quad \quad | \quad k \in \Psi_7 \} \end{array}$$

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  $Y$ .
- $\{z_k \mid k \in \Psi_1 \cup \Psi_3 \cup \Psi_5 \cup \Psi_7\}$  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  $S = \cup_{s=1}^3 A_s \cup_{s=5}^7 A_s$  as below. We note that we do not add vectors to  $A_4$  yet (this will be done after the process below ends):

**Initialize**  $A_1 = A_2 = A_3 = A_5 = A_6 = A_7 = \emptyset$ .  $S = \cup_{s=1}^3 A_s \cup_{s=5}^7 A_s = \emptyset$ .

**Initialize**  $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \Sigma_4 = \emptyset$ .  $\Phi = \Phi_1 \cup \Phi_2 \cup \Phi_5 \cup \Phi_6 = \emptyset$ .  $\Psi = \Psi_1 \cup \Psi_3 \cup \Psi_5 \cup \Psi_7 = \emptyset$ .

**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  $a \notin \text{Span}\{a_i \mid i \in \Sigma\}$ ,  $y \notin \text{Span}\{y_j \mid j \in \Phi\}$ ,  $z \notin \text{Span}\{z_k \mid k \in \Psi\}$ , 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.
- ..... 5 more similar cases .....

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  $\text{Span}(S) \subseteq W$  be the span of all the vectors chosen so far. A small modification of the process above ensures that  $(W \cap (Y \oplus Z)) \subseteq \text{Span}(S)$ . 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 A.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  $A_4$ -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  $a \notin \text{Span}\{a_i \mid i \in \Sigma\}$ . This is because, otherwise we can cancel out  $a$  by adding back appropriate vectors in  $S$ . This would result in a vector in  $W \cap (Y \oplus Z) \subseteq \text{Span}(S)$ , a contradiction.

- It must be that  $y \in \text{Span}\{y_j | j \in \Phi\}$  as well as  $z \in \text{Span}\{z_k | k \in \Psi\}$ . 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  $\{y_j | j \in \Phi_2 \cup \Phi_6\}$ ” by adding back appropriate vectors in  $S$ .

□

**Lemma A.3.** *Suppose  $P, Q$  are independent spaces,  $\dim(P) = \dim(Q) = n$ , and  $W \subseteq P \oplus Q$ . Suppose moreover that  $p_1, \dots, p_n$  and  $q_1, \dots, q_n$  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)*

$$(p_1, \dots, p_n) \leftarrow M(p_1, \dots, p_n), \quad (q_1, \dots, q_n) \leftarrow M(q_1, \dots, q_n),$$

there is a partition of the index set  $\{1, \dots, n\} = \Delta_0 \cup \Delta_1 \cup \Delta_2 \cup \dots \cup \Delta_m \cup \Omega_1 \cup \Omega_2$  and a basis for  $W$  of the form:

$$\begin{aligned} & \{p_i + \sigma(q) \mid i \in \Delta_1\} \cup B_2 \cup \dots \cup B_m \cup C_1 \cup C_2, \\ & B_k = \left\{ q_j + \sigma(p_{\Delta_{[k+1:m]}}) \mid j \in \Delta_k \right\} \end{aligned}$$

where  $\Delta_{[k+1:m]} = \Delta_{k+1} \cup \dots \cup \Delta_m \cup \Omega_1 \cup \Omega_2$ ,

$$C_1 = \{p_i + \sigma(q_{\Omega_1}) \mid i \in \Omega_1 \cup \Omega_2\},$$

$$C_2 = \{q_j \mid j \in \Omega_2\}.$$

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.

$$\begin{aligned} W_P &= \{p \in P \mid \exists q \in Q, p + q \in W\}, \\ W_Q &= \{q \in Q \mid \exists p \in P, p + q \in W\}. \end{aligned}$$

It is clearly possible to choose matched bases  $(p_1, \dots, p_d, p_{d+1}, \dots, p_n)$  and  $(q_1, \dots, q_d, q_{d+1}, \dots, q_n)$  for  $P$  and  $Q$  respectively such that

$$\begin{aligned} W_P &= \text{Span of } p_1, \dots, p_s, & p_{t+1}, \dots, p_d \\ W_Q &= \text{Span of } q_1, \dots, q_s, & q_{s+1}, \dots, q_t \end{aligned}$$

The “unused” indices  $\{d+1, \dots, n\}$  are placed in  $\Delta_0$ . We choose arbitrary linear forms  $\sigma_i(q)$  so that

$$p_{t+1} + \sigma_{t+1}(q), \dots, p_d + \sigma_d(q) \in W.$$

These vectors are added to a partial basis for  $W$  and the indices  $\{t+1, \dots, d\}$  are added to  $\Delta_1$ . Letting  $W' = W \cap \text{Span}(p_1, \dots, p_t, q_1, \dots, q_t)$ , clearly

$$W = W' \oplus \text{Span}(p_{t+1} + \sigma_{t+1}(q), \dots, p_d + \sigma_d(q)).$$

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  $\{1, \dots, t\}$  and we can continue iteratively. This process makes progress unless  $p_{t+1}, \dots, p_d$  are absent, i.e. if

$$\begin{aligned} W_P &= \text{Span of } p_1, \dots, p_s, \\ W_Q &= \text{Span of } q_1, \dots, q_s, \quad q_{s+1}, \dots, q_t. \end{aligned}$$

In this case, the iterative process is stopped, and we begin a new iterative process. We set  $m = 2$  and choose

$$q_{s+1} + \sigma(p_{\{1, \dots, s\}}), \dots, q_t + \sigma(p_{\{1, \dots, s\}}) \in W.$$

These vectors are added to a partial basis for  $W$  letting  $\Delta_m = \{s + 1, \dots, t\}$ . We increment  $m$  by one and iterate the process on  $W' = W \cap \text{Span}(p_1, \dots, p_s, q_1, \dots, q_s)$ . We note that  $W'_Q = \text{Span}(q_1, \dots, q_s)$  while  $W'_P \subseteq \text{Span}(p_1, \dots, p_s)$  (it may shrink to a proper subspace). After appropriate change of matched basis,  $W'_P = \text{Span}(p_1, \dots, p_r)$  for  $r \leq s$ . This process makes progress unless we have

$$\begin{aligned} W_P &= \text{Span of } p_1, \dots, p_s, \\ W_Q &= \text{Span of } q_1, \dots, q_s. \end{aligned}$$

At this point, a basis for  $W$  is completed by first taking elements

$$p_1 + \sigma_1(q_{\{1, \dots, s\}}), \dots, p_s + \sigma_s(q_{\{1, \dots, s\}}) \in W,$$

and adding to it  $q_{r+1}, \dots, q_s \in W \cap Q$ . We can eliminate dependency of the former on the latter by elimination. The proof is completed by setting  $\Omega_1 = \{1, \dots, r\}$  and  $\Omega_2 = \{r + 1, \dots, s\}$ .  $\square$

**Lemma A.4.** *Let  $X_1, \dots, X_n$  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  $r \in \{1, \dots, n\}$ .*
- *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*

$$\mathbb{E}_{X_1, \dots, X_n} \left[ \prod_{i=1}^m |\lambda_i(Y_{i1}, \dots, Y_{is_i})| \right] \leq \prod_{i=1}^m \sqrt{\mathbb{E}_{Y_{i1}, \dots, Y_{is_i}} [\lambda_i^2(Y_{i1}, \dots, Y_{is_i})]}.$$

*Proof.* By induction. For  $n = 1$ , the only scenario and its proof is Cauchy-Schwartz:

$$\mathbb{E}_{X_1} [|\lambda_1(X_1)\lambda_2(X_1)|] \leq \sqrt{\mathbb{E}_{X_1} [\lambda_1^2(X_1)]} \sqrt{\mathbb{E}_{X_1} [\lambda_2^2(X_1)]}.$$

Otherwise, we assume w.l.o.g. that  $n \geq 2$  and  $Y_{11} = Y_{21} = X_n$ . Applying Cauchy-Schwartz on  $X_n$ ,

$$\mathbb{E}_{X_1, \dots, X_n} \left[ \prod_{i=1}^m |\lambda_i(Y_{i1}, \dots, Y_{is_i})| \right] \leq \mathbb{E}_{X_1, \dots, X_{n-1}} \left[ \sqrt{\mathbb{E}_{X_n} [\lambda_1^2(X_n, \cdot)]} \sqrt{\mathbb{E}_{X_n} [\lambda_2^2(X_n, \cdot)]} \prod_{i=3}^m |\lambda_i(Y_{i1}, \dots, Y_{is_i})| \right].$$

We get a further desired upper bound by induction hypothesis applied to functions  $\lambda'_1, \lambda'_2, \lambda_3, \dots, \lambda_m$  where

$$\lambda'_1(Y_{12}, \dots, Y_{1s_1}) = \mathbb{E}_{X_n} [\lambda_1^2(X_n, Y_{12}, \dots, Y_{1s_1})], \quad \lambda'_2(Y_{22}, \dots, Y_{2s_2}) = \mathbb{E}_{X_n} [\lambda_2^2(X_n, Y_{12}, \dots, Y_{1s_1})].$$

□

**Lemma A.5.** *Suppose  $x \in \{0, 1\}^k$  is a uniformly distributed input and  $A, B \subseteq \{1, \dots, k\}$  such that  $A \cup B = \{1, \dots, k\}$ . 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

$$\mathbb{E}_x [\lambda(x_A)\psi(x_B)] = \mathbb{E}_{x_{A \cap B}} \left[ \mathbb{E}_{x_{A \setminus B}} [\lambda(x_A)] \cdot \mathbb{E}_{x_{B \setminus A}} [\psi(x_B)] \right] \leq \beta \cdot \mathbb{E}_{x_{A \cap B}} \left[ \mathbb{E}_{x_{B \setminus A}} [\psi(x_B)] \right] = \beta \cdot \mathbb{E}_{x_B} [\psi(x_B)].$$

□