

# Sparse Graph Counting and Kelley–Meka Bounds for Binary Systems

Yuval Filmus

Technion

Haifa, Israel

yuvalfi@cs.technion.ac.il

Hamed Hatami

McGill University

Montreal, Canada

hatami@cs.mcgill.ca

Kaave Hosseini

University of Rochester

Rochester, USA

kaave.hosseini@rochester.edu

Esty Kelman

Boston University &amp; MIT

Boston &amp; Cambridge, USA

ekelman@mit.edu

**Abstract**—In a recent breakthrough, Kelley and Meka (FOCS 2023) obtained a strong upper bound on the density of sets of integers without non-trivial three-term arithmetic progressions. In this work, we extend their result, establishing similar bounds for all linear patterns defined by binary systems of linear forms, where “binary” indicates that every linear form depends on exactly two variables. Prior to our work, no strong bounds were known for such systems even in the finite field model setting.

A key ingredient in our proof is a *graph counting lemma*. The classical graph counting lemma, developed by Thomason (Random Graphs 1985) and Chung, Graham, and Wilson (Combinatorica 1989), is a fundamental tool in combinatorics. For a fixed graph  $H$ , it states that the number of copies of  $H$  in a pseudorandom graph  $G$  is similar to the number of copies of  $H$  in a purely random graph with the same edge density as  $G$ . However, this lemma is only non-trivial when  $G$  is a dense graph. In this work, we prove a graph counting lemma that is also effective when  $G$  is sparse. Moreover, our lemma is well-suited for density increment arguments in additive number theory.

As an immediate application, we obtain a strong bound for the Turán problem in abelian Cayley sum graphs: let  $\Gamma$  be a finite abelian group with odd order. If a Cayley sum graph on  $\Gamma$  does not contain any  $r$ -clique as a subgraph, it must have at most  $2^{-\Omega_r(\log^{1/16} |\Gamma|)} \cdot |\Gamma|^2$  edges.

These results hinge on the technology developed by Kelley and Meka and the follow-up work by Kelley, Lovett, and Meka (STOC 2024).

**Index Terms**—sparse graph counting lemma, Kelley–Meka bounds, Density increment, additive combinatorics, pseudorandomness and derandomization

## I. INTRODUCTION

The *graph counting lemma*, due to Thomason [1] and Chung, Graham, and Wilson [2], is a fundamental result in graph theory, with many applications in extremal combinatorics, additive number theory, discrete geometry, and theoretical computer science. Informally, it states that the number of embeddings of a fixed graph  $H$  into a random-looking (pseudorandom) graph  $G$  is similar to the number of embeddings of  $H$  into a genuine random graph with the same edge density as  $G$ .

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 802020-ERC-HARMONIC. This work was partially supported by the Goergen Institute for Data Science at the University of Rochester. E. K. was supported in part by the NSF TRIPODS program (award DMS-2022448). This project was partially carried out at the Simons Institute for the Theory of Computing, UC Berkeley, while the authors were participating in the program Analysis and TCS in Summer 2023.

For certain problems, particularly in additive number theory, it is possible to use the counting lemma to obtain a density increment argument. Roughly speaking, if the number of embeddings of  $H$  into  $G$  deviates from that of a random graph, then by the counting lemma,  $G$  is not pseudorandom and, therefore, contains some structure. In certain cases, it is possible to exploit this structure to restrict the problem to a significantly denser subproblem and repeat the argument. As the density cannot exceed one, this process must stop after a bounded number of iterations.

Many of the strongest known bounds in additive number theory and extremal combinatorics rely on the density increment approach. This method often yields much better bounds compared to other common techniques, such as the regularity method.

The classical graph counting lemma has a shortcoming: it is only effective for dense graphs. Indeed, while the counting lemma and the regularity method provide an elegant theory for understanding the subgraph densities in dense graphs [3], obtaining an analogous theory for the sparse setting has proven challenging. Over the past few decades, considerable effort has been devoted towards developing sparse analogs of the counting and regularity lemmas (see for example [4]–[11]).

The conventional strategy for extending these results to sparse graphs is to introduce an additional requirement: that the graph  $G$ , even though sparse, is a relatively dense subgraph of a much more pseudorandom host graph  $\Gamma$ . For example, in an important paper, Conlon, Fox, and Zhao [10] proved an effective sparse counting lemma under this additional requirement. Their result generalizes an earlier theorem of Kohayakawa, Rödl, Schacht, and Skokan [9] that only applied to counting triangles. However, the additional requirement that  $G$  must live in a much more pseudorandom graph  $\Gamma$  makes such sparse counting lemmas unsuitable for certain applications of the original counting lemma.

In this paper, we prove a counting lemma for sparse graphs (Theorem I.4) that circumvents this extra assumption on  $G$ . Our theorem states that if the number of embeddings of  $H$  into  $G$  deviates from what is expected, then either  $G$  contains many vertices of small degree, or there exist large subsets  $S, T \subseteq V(G)$  such that the density of the induced bipartite graph between them is notably *higher* than the density of  $G$ . Due to its guarantee of increased density, our counting lemma

is tailored for density increment arguments.

We use our new sparse counting lemma alongside various tools from additive number theory, such as the Bohr set machinery, almost periodicity, and dependent random choice, to extend Kelley and Meka's strong bound on 3-progression-free sets to a broad and important class of more general linear patterns. We will postpone the detailed discussion of this result to Section I-B.

### A. The sparse graph counting lemma

We identify undirected graphs as symmetric functions  $A: \mathcal{X} \times \mathcal{X} \rightarrow \{0, 1\}$  through the correspondence with the characteristic function of the edges. The *density* of  $A$  is

$$\mathbb{E}[A] := \mathbb{E}_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}}[A(\mathbf{x}_1, \mathbf{x}_2)],$$

where  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$  means that  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are chosen uniformly and independently from  $\mathcal{X}$ .

Let  $H = ([k], E_H)$  be a fixed undirected graph with vertex set  $[k] := \{1, \dots, k\}$  and  $m = |E_H|$  edges. We define the *density* of  $H$  in  $A$  as

$$t_H(A) := \mathbb{E}_{\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{X}} \left[ \prod_{\{i, j\} \in E_H} A(\mathbf{x}_i, \mathbf{x}_j) \right].$$

When  $A$  is the Erdős–Rényi random graph with parameter  $\alpha$ , we have  $t_H(A) \approx \alpha^m$ , with high probability. For an arbitrary  $A$ , the graph counting lemma of [1], [2] guarantees  $t_H(A) \approx \alpha^m$  if  $A$  is pseudorandom in a certain structural sense. We state this result in the contrapositive form.

**Theorem I.1** (Graph Counting Lemma [1], [2]). *For every fixed graph  $H$  with  $m$  edges and  $\varepsilon \in (0, 1)$ , there exists  $\delta = \delta(\varepsilon, m) > 0$  such that the following is true. If  $A$  is a graph on  $\mathcal{X}$  with density  $\alpha$  that satisfies*

$$|t_H(A) - \alpha^m| \geq \varepsilon,$$

*there must exist  $S, T \subseteq \mathcal{X}$  with densities  $\frac{|S|}{|\mathcal{X}|}, \frac{|T|}{|\mathcal{X}|} \geq \Omega_{m, \varepsilon}(1)$  such that*

$$|\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T}[A(\mathbf{x}, \mathbf{y})] - \alpha| \geq \delta.$$

In other words, if  $t_H(A)$  is significantly different from what is expected from a random graph of density  $\alpha$ , then on some large combinatorial rectangle  $S \times T$ , the density of  $A$  significantly deviates from  $\alpha$ .

In the sparse setting, where  $\alpha = o(1)$ , we have  $|t_H(A) - \alpha^m| = o(1)$ , and therefore, to obtain a useful counting lemma for sparse graphs, we need to refine the assumption of the counting lemma to

$$|t_H(A) - \alpha^m| \geq \varepsilon \alpha^m. \quad (1)$$

The following well-known example shows that (1) does not necessarily imply the existence of  $S, T \subseteq \mathcal{X}$  with density  $\Omega_{m, \varepsilon}(1)$  that satisfy

$$|\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T}[A(\mathbf{x}, \mathbf{y})] - \alpha| \geq \Omega(\alpha). \quad (2)$$

**Example I.2.** Let  $H$  be the triangle graph, and let  $A$  be the Erdős–Rényi random graph on  $n$  vertices with parameter

$\alpha = n^{-2/3}$ . With high probability, we have  $t_H(A) \approx \alpha^3 = n^{-2}$ , and therefore, there are at most  $O(n)$  triangles in  $A$ . By removing at most  $O(n)$  edges from  $A$ , we can obtain a triangle-free  $A'$ . Since  $A'$  is triangle-free and its density is very close to  $\alpha$ , it satisfies the assumption (1). Moreover,

$$|\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T}[A'(\mathbf{x}, \mathbf{y})] - \alpha| = o(\alpha)$$

for all subsets  $S, T \subseteq [n]$  of density at least, say,  $n^{-1/7}$ . Therefore,  $A'$  violates the desired conclusion (2).

To circumvent such examples, Conlon, Fox and Zhao [10] assume further that  $A$  is a relatively dense subgraph of a host graph  $\Gamma$  that satisfies much stronger pseudo-randomness conditions. Their counting lemma states that under this extra requirement, (1) indeed implies (2) for some  $S, T \subseteq \mathcal{X}$  with density  $\Omega_{m, \varepsilon}(1)$ . We prove a counting lemma that does not require such an extra pseudorandomness condition on  $A$ .

**Theorem I.3.** *Let  $H$  be a graph with  $m$  edges, and let  $\varepsilon \in (0, 1)$ . For every graph  $A$  with vertex set  $\mathcal{X}$  and density  $\alpha$ , the following holds.*

(i) *If  $t_H(A) \geq (1 + \varepsilon)^m \alpha^m$ , there exist  $S, T \subseteq \mathcal{X}$  with  $\frac{|S||T|}{|\mathcal{X}|^2} = \Omega(\alpha^{m+1})$  such that*

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T}[A(\mathbf{x}, \mathbf{y})] \geq \left(1 + \frac{\varepsilon}{2}\right) \alpha.$$

(ii) *If  $t_H(A) \leq (1 - \varepsilon)^m \alpha^m$ , there exist  $S, T \subseteq \mathcal{X}$  with  $\frac{|S||T|}{|\mathcal{X}|^2} = \Omega(\varepsilon(1 - \varepsilon)^{m-1} \alpha^m)$  such that*

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T}[A(\mathbf{x}, \mathbf{y})] \leq \left(1 - \frac{\varepsilon}{2}\right) \alpha.$$

Theorem I.3 is a special case of a slightly more general result (Theorem II.2) that we state and prove later. Theorem I.3 is not our main counting lemma, and as explained later, it is not useful for strong density increment applications. However, it can be viewed as an extension of Theorem I.1 to the sparse setting and provides a nice complement to Example I.2.

Kelley, Lovett, Meka [12] proved the first part of Theorem I.3 using a smooth analog of the dependent random choice method in the special case when  $H$  is a complete bipartite graph. In Theorem I.3, we generalize their result to all graphs using a similar proof.

a) *A density increment counting lemma.:* Theorem I.3 is not useful for obtaining density increment arguments since, in case (ii), it only provides a combinatorial rectangle with *decreased* density.

Let us consider the original counting lemma (Theorem I.1). By tracking the parameters in the proof of this theorem, one can easily show that (1) implies the existence of  $S, T \subseteq \mathcal{X}$  with densities  $\Omega_{m, \varepsilon}(1)$  such that

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T}[A(\mathbf{x}, \mathbf{y})] \geq \alpha + \Omega(\alpha^m).$$

While this result provides a density increase, the magnitude of this increase is insignificant when  $\alpha$  is small. In many applications, having an increase of  $\Omega(\alpha)$  is crucial. For example, a density increment that multiplies the density by a factor of  $1 + \Omega(1)$  can be repeated at most  $O(\log(2/\alpha))$  times. In

contrast, a weaker density increment by a factor of  $1 + \Omega(\alpha^c)$  for  $c > 0$  needs to be repeated at least  $\Omega(\alpha^{-c} \log(2/\alpha))$  times before it reaches a contradiction. Since each iteration decreases the size of the domain by a constant factor, this large number of iterations would only be possible if  $\alpha \geq \Omega(1/\log |\mathcal{X}|)$ .

Our main graph counting lemma provides a significant density increment when  $A$  does not contain many vertices with small degrees.

**Theorem I.4** (Main graph counting lemma). *Let  $H$  be a graph with  $m$  edges, and let  $\varepsilon \in (0, 1)$ . There exists  $\delta = \delta(\varepsilon, m) > 0$  such that for every graph  $A$  of density  $\alpha$  on a vertex set  $\mathcal{X}$ , the following holds. If*

$$|t_H(A) - \alpha^m| \geq \varepsilon \alpha^m,$$

*then, either  $A$  has  $2^{-O_{m,\varepsilon}(\log(2/\alpha))} |\mathcal{X}|$  vertices with degrees at most  $(1 - \delta)\alpha |\mathcal{X}|$ , or there exist  $S, T \subseteq \mathcal{X}$  of densities at least  $2^{-O_{m,\varepsilon}(\log^2(2/\alpha))}$  satisfying*

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S \times T} [A(\mathbf{x}, \mathbf{y})] \geq (1 + \delta)\alpha.$$

Theorem I.4 states that if  $A$  is a graph with  $|t_H(A) - \alpha^m| \geq \varepsilon \alpha^m$ , then either it has many vertices with small degrees, or the density of  $A$  on some large rectangle is greater than  $(1 + \delta)\alpha$ . The significance of this theorem lies in the fact that even when  $t_H(A)$  is *smaller* than, say,  $\alpha^m/2$ , it provides a rectangle with *increased* density. This feature makes the theorem suitable for density increment arguments, especially in additive number theory, where one naturally deals with regular graphs.

Theorem I.4 is a special case of a more general result (Theorem II.1) which we state and prove later.

#### B. Kelley–Meka bounds for binary systems of linear forms

To study the number of occurrences of a linear pattern (e.g., 3-progressions) in a subset of the interval  $\{1, \dots, M\}$ , it suffices to embed  $\{1, \dots, M\}$  in  $\mathbb{Z}_N$  for a prime  $N = O(M)$  chosen sufficiently large to avoid wraparound. Consequently, rather than working with the interval, one can focus on subsets of finite abelian groups  $G$ . The two important cases are  $G = \mathbb{Z}_N$ , where  $N$  is a large prime and asymptotics are as  $N$  tends to infinity, and  $G = \mathbb{F}_q^n$ , where  $q$  is a fixed prime and asymptotics are as  $n$  tends to infinity. The former is important for applications in number theory, and the latter is the setting that is most relevant to applications in combinatorics and computer science.

In 1953, Roth [13] proved that a set of integers  $A \subseteq \mathbb{Z}_N$  without non-trivial 3-progressions must have density  $O(1/\log \log N)$ . The quest to determine the optimal bound in Roth’s theorem stands as a central problem in additive number theory, which has been studied extensively over the past seven decades (notably [14]–[19]). For a long time,  $1/\log N$  was perceived as a barrier for all the available techniques. In [18], Bloom and Sisask overcame this barrier and established the bound  $O(1/\log^{1+\varepsilon} N)$  for some small  $\varepsilon > 0$ . More recently, in a remarkable breakthrough, Kelley and Meka [19] took a huge leap forward and improved the bound to  $2^{-\Omega(\log^{1/12} N)}$ . Subsequently, this bound was refined

to  $2^{-\Omega(\log^{1/9} N)}$  by Bloom and Sisask [20]. On the other hand, Behrend’s classical construction [21] shows that there are sets with density  $2^{-O(\log^{1/2} N)}$  that are free of non-trivial 3-progressions.

Similarly to all previous approaches, [19] uses a density increment argument. Note that the density of 3-progressions in a set  $A \subseteq \mathbb{Z}_N$  is given by

$$t_{3\text{AP}}(A) := \mathbb{E}_{\mathbf{x}, \mathbf{y} \in \mathbb{Z}_N} [A(\mathbf{x})A(\mathbf{x} + \mathbf{y})A(\mathbf{x} + 2\mathbf{y})], \quad (3)$$

where we identified  $A$  with its indicator function. If  $A$  is a random set of density  $\alpha$ , then we expect  $t_{3\text{AP}}(A) \approx \alpha^3$ . Kelley and Meka [19] show that if a set  $A$  with density  $\alpha$  satisfies

$$|t_{3\text{AP}}(A) - \alpha^3| \geq \frac{\alpha^3}{2},$$

then restricting to some large low-rank Bohr set  $B$  increases the relative density of  $A$  by a multiplicative factor of  $1 + \Omega(1)$ . Repeating this argument iteratively for at most  $O(\log(2/\alpha))$  times shows that if  $A$  does not contain any non-trivial 3-progressions, then  $\alpha \leq 2^{-\Omega(\log^{1/12} N)}$ .

Most linear patterns, such as 4-progressions, are fundamentally beyond the scope of such a Bohr set counting lemma: there are sets  $A$  with density  $\alpha = \Omega(1)$  that satisfy  $|t_{4\text{AP}}(A) - \alpha^4| = \Omega(1)$ , but the relative density of  $A$  in every large low-rank Bohr set remains extremely close to  $\alpha$ , see [22].

*Remark 1.* In 1975, Szemerédi [23] extended Roth’s  $o(1)$  bound using a completely new approach to arithmetic progressions of arbitrary length. However, in contrast to Roth’s Fourier-analytic proof, all the various known proofs of Szemerédi’s theorem give much weaker bounds. Indeed, it was considered a major breakthrough when Gowers [22] proved an upper bound of  $1/(\log \log N)^{2^{-2^{k+9}}}$  on the density of sets of integers without  $k$ -term arithmetic progressions. Very recently, Leng, Sah, and Sawhney posted a preprint [24] improving Gowers’ bound to an impressive bound of  $O(2^{-(\log \log N)^{c_k}})$ , where  $c_k > 0$  is a constant depending on  $k$ .

Nonetheless, there is an important class of linear patterns that lie somewhere between 3-progressions and 4-progressions in terms of their “complexity”. These linear patterns, defined by *binary systems of linear forms*, stem from the graph-theoretic approach of Ruzsa and Szemerédi [25] to Roth’s theorem. Ruzsa and Szemerédi discovered a versatile proof of Roth’s theorem by establishing and applying the so-called triangle removal lemma, which is a consequence of Szemerédi’s regularity lemma. Their proof uses the expression

$$t_{3\text{AP}}(A) = \mathbb{E}_{\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{Z}_N} [A(2\mathbf{x} - 2\mathbf{y})A(\mathbf{x} - \mathbf{z})A(2\mathbf{y} - 2\mathbf{z})] \quad (4)$$

to represent the density of 3-progressions in  $A$  as the density of triangles in a tripartite graph constructed from  $A$ , and then applies these graph theoretic tools.

In [26], Erdős, Frankl, and Rödl proved the graph removal lemma, which generalizes the triangle removal lemma to all graphs. This generalization extends the graph-theoretic approach to a large class of linear patterns, namely those defined by *binary systems of linear forms*. Essentially, these

linear patterns have a graph-like structure, similar to how 3-progressions have a triangle-like structure. We will formally define these systems in Section I-D, but for now, it suffices to mention that the term *binary* indicates that each linear form is supported on exactly two variables. Note that  $(2x - 2y, x - z, 2y - 2z)$ , which captures 3-progressions, satisfies this property. As another notable example of a binary system, consider

$$t_{K_4}(A) := \mathbb{E}_{\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w} \in \mathbb{Z}_N} [A(\mathbf{x} + \mathbf{y})A(\mathbf{x} + \mathbf{z})A(\mathbf{x} + \mathbf{w})A(\mathbf{y} + \mathbf{z})A(\mathbf{y} + \mathbf{w})A(\mathbf{z} + \mathbf{w})], \quad (5)$$

which corresponds to the density of  $K_4$  (the complete graph on 4 vertices) in the Cayley sum graph  $\text{Cay}(\mathbb{Z}_N, A)$ , whose edges are  $\{x, y\}$  for  $x + y \in A$ .

The graph-theoretic approach relies on Szemerédi's regularity lemma, and therefore results in an extremely weak bound compared to Roth's original bound. However, one can use a common technique in additive number theory and extremal combinatorics, based on a few careful applications of the Cauchy–Schwarz inequality, to extend Roth's original Fourier-analytic proof and upper bound to all binary systems.

Unfortunately, the aforementioned Cauchy–Schwarz approach incurs too much loss to extend Kelley and Meka's upper bound to such a broad class of linear patterns. In fact, a priori, it is unclear whether Kelley and Meka's techniques could be applied to, for instance,  $t_{K_4}(A)$ . Let us briefly review the key steps of their proof:

- 1) Their proof starts from the assumption  $|t_{3\text{AP}}(A) - \alpha^3| \geq \varepsilon\alpha^3$ , and notes that

$$t_{3\text{AP}}(A) - \alpha^3 = \langle (A - \alpha) * (A - \alpha), C \rangle,$$

where  $C = \{2a \mid a \in A\}$ .

- 2) (Hölder step): Since the density of  $C$  is  $\alpha$ , one can use Hölder's inequality to show that the right-hand side is at most  $O(\alpha \|(A - \alpha) * (A - \alpha)\|_p)$  for  $p \approx \log(2/\alpha)$ .
- 3) (Positivity/Significant Increment): Then, in one of the key steps of the proof, the proof uses the fact that the Fourier coefficients of the convolution  $(A - \alpha) * (A - \alpha)$  are positive to deduce from  $\|(A - \alpha) * (A - \alpha)\|_p = \Omega(\alpha^2)$  that

$$\|A * A\|_p = (1 + \Omega(1))\alpha^2.$$

- 4) (Sifting/Dependent random choice): This key step involves the application of the dependent random choice technique to the lower bound on  $\|A * A\|_p$ .
- 5) (Almost periodicity): At this point, one can combine the result of the previous step with the available almost periodicity result of [27] to achieve the desired density increment.

Note that several steps in the above proof rely on the properties of convolution, and while one can express  $t_{3\text{AP}}(A)$  by a simple formula involving convolution, this is not true for other more complex binary systems such as  $t_{K_4}(A)$ . Is it still true that  $|t_{K_4}(A) - \alpha^6| = \Omega(\alpha^6)$  implies a significant density increment in a large low-rank Bohr set? If  $|t_{K_4}(A) - \alpha^6| =$

$\Omega(\alpha^6)$ , then our counting lemma (Theorem I.4) provides two sets  $S$  and  $T$  with densities  $2^{-O(\log^2(2/\alpha))}$  such that

$$(1 + \Omega(1))\alpha_S\alpha_T\alpha \leq \mathbb{E}_{\mathbf{x}, \mathbf{y} \in \mathbb{Z}_N} [S(\mathbf{x})T(\mathbf{y})A(\mathbf{x} + \mathbf{y})], \quad (6)$$

where  $\alpha_S$  and  $\alpha_T$  are the densities of  $S$  and  $T$ , respectively.

This is promising, as  $\mathbb{E}[S(\mathbf{x})T(\mathbf{y})A(\mathbf{x} + \mathbf{y})]$  is a much simpler expression than  $t_{K_4}(A)$ , and moreover, we have shown in (6) that it is significantly larger than its “expected” value  $\alpha_S\alpha_T\alpha$ . By establishing a sifting step for general finite abelian groups that is applicable to (6), we prove the following theorem.

**Theorem I.5** (Kelley–Meka-type bounds over finite abelian groups). *Let  $\mathcal{L}$  be a binary system of linear forms over a finite abelian group  $G$ , and suppose  $|G|$  is coprime with all the coefficients in the linear forms. If  $A \subseteq G$  is free of translations of non-degenerate instances of  $\mathcal{L}$ , then*

- (i)  $|A| \leq |G| \cdot 2^{-\Omega_{\mathcal{L}}(\log^{1/16}|G|)}$ .
- (ii) *If, in addition, the underlying graph of  $\mathcal{L}$  is 2-degenerate, then*

$$|A| \leq |G| \cdot 2^{-\Omega_{\mathcal{L}}(\log^{1/9}|G|)}.$$

Here, an undirected graph  $H$  is called *k-degenerate* if every subgraph of  $H$  has a vertex of degree at most  $k$ .

*Remark 2.* It is worth noting that even in the case of the group  $G = \mathbb{F}_q^n$ , despite the success of the polynomial method in proving strong bounds for Roth's theorem [28], [29], the bound of Theorem I.5 is new. In 2017, [28], [29] used the polynomial method to prove that any subset of  $G = \mathbb{F}_q^n$  without 3-progressions must be of size at most  $c^n$ , where  $c = c(q) < q$  is a fixed constant. This bound is stronger than Kelley and Meka's bound, and indeed, the significance of Kelley and Meka's result lies in the integer case, where there is no known analog of the polynomial method. However, even in the setting of  $\mathbb{F}_q^n$ , the polynomial method appears to be limited in dealing with general binary systems. In a recent article [30], Gijswijt extended the polynomial method to a larger class of systems of linear forms. However, even this class does not include the binary systems whose underlying graphs are dense graphs such as  $K_4$  or larger cliques. For cases such as (5), the strongest known previous bounds were of the form  $q^{n-\Omega_q(\log n)}$ . Theorem I.5 improves these bounds to  $q^{n-\Omega_q(n^{1/16})}$ .

*Remark 3.* In the case of 3 - progressions, currently the strongest known bound for  $G = \mathbb{Z}_N$  is  $|A| \leq |G| \cdot 2^{-\Omega(\log^{1/9}|G|)}$ , due to Bloom and Sisask [20], which is stronger than what we obtain for general binary systems in Theorem I.5 (i). To briefly explain this disparity, note that we need to apply our counting lemma to obtain (6), and the two resulting sets,  $S$  and  $T$ , are of densities  $2^{-O(\log^2(2/\alpha))}$ , which is small compared to  $\alpha$ . In contrast, in the case of 3-progressions, one can directly work with  $\|A * A\|_p$ , which leads to a stronger bound. Nevertheless, in Theorem I.5 (ii) we show that if the underlying graph of the binary system is 2-degenerate (e.g., the triangle graph for 3-progressions), we can obtain the same upper bound  $|A| \leq |G| \cdot 2^{-\Omega(\log^{1/9}|G|)}$ .

### C. Turán's theorem for Cayley sum graphs over abelian groups

Here, we present a quick application of Theorem I.5 to Turán's theorem, a classical result in extremal graph theory regarding the largest edge density of a graph that does not contain any copies of  $K_r$  (the complete graph on  $r$  vertices), for some given  $r$ . Turán proved that this density is at most  $1 - \frac{1}{r-1} + o(1)$ , which is tight, as witnessed by a complete  $(r-1)$ -partite graph where the parts have an almost equal number of vertices.

Recall that given a finite abelian group  $\Gamma$  and a set  $A \subseteq \Gamma$ , the Cayley sum graph  $\text{Cay}(\Gamma, A)$  is the graph with vertex set  $\Gamma$  and edge set  $\{(x, y) \mid x + y \in A\}$ . Theorem I.5 (i) implies a strong upper bound on the edge density of  $G$  in Turán's theorem if  $G$  is further assumed to be a Cayley sum graph of a finite abelian group  $\Gamma$  with  $2 \nmid |\Gamma|$ .

**Corollary I.6.** *Let  $r \geq 2$  be a fixed positive integer. Let  $\Gamma$  be a finite abelian group such that  $2 \nmid |\Gamma|$ , and let  $A \subseteq \Gamma$ . If  $\text{Cay}(\Gamma, A)$  does not contain any copies of  $K_r$ , then its edge density is at most  $2^{-\Omega_r(\log^{1/16} |\Gamma|)}$ .*

*Proof.* The system of linear forms associated with  $r$ -clique is  $\mathcal{L} = \{x_i + x_j : i \neq j, i, j \in [r]\}$ . Next, we show that if  $2 \nmid |\Gamma|$ , then  $\mathcal{L}$  is translation invariant (see Definition 1 on page 5), and therefore, we can apply Theorem I.5 to conclude the corollary. We need to show that every shift of an  $r$ -clique is also an  $r$ -clique. This follows from the identity  $x_i + x_j + c = (x_i + c/2) + (x_j + c/2)$  for all  $i \neq j$ , which is valid if  $2 \nmid |\Gamma|$ .  $\square$

Note that the requirement  $2 \nmid |\Gamma|$  is crucial. For example, let  $\Gamma = \mathbb{Z}_{2N}$  and let  $A$  be the set of all odd numbers in  $\mathbb{Z}_{2N}$ , namely,  $A = \{2i + 1 \bmod 2N : i \in [N]\} \subseteq \Gamma$ . The Cayley sum graph  $\text{Cay}(\Gamma, A)$  does not contain any triangles because for any three elements  $x, y, z \in \mathbb{Z}_{2N}$ , at least one of  $x+y, x+z, y+z$  is even, hence not in  $A$ . However,  $\text{Cay}(\Gamma, A)$  has edge density  $\approx \frac{1}{2}$ .

*Layout of the paper:* In the remainder of the introduction, we introduce some notation and terminology on systems of linear forms. In Section II, we discuss the graph counting lemmas and their proofs. In Section III, we discuss the Kelley–Meka-type bounds for binary linear systems. In Section IV, we discuss some concluding remarks regarding the limitations of these methods in relation to other systems of linear forms.

### D. Notation and terminology

The function  $\log$  is the base 2 logarithm. For a positive integer  $m$ , we denote  $[m] := \{1, \dots, m\}$ . We denote random variables with bold font.

Let  $\mathcal{X}$  be a finite set. Given a function  $f: \mathcal{X} \rightarrow \mathbb{R}$  and  $p \in [1, \infty)$ , let  $\|f\|_p = (\mathbb{E}[|f(\mathbf{x})|^p])^{1/p}$ , and let  $\|f\|_\infty = \max_x |f(x)|$ .

We use the standard Bachmann–Landau asymptotic notation:  $O(\cdot), \Omega(\cdot), \Theta(\cdot), o(\cdot)$ , and  $\omega(\cdot)$ . Sometimes, we will use these notations in combination with subscripts to indicate that the hidden constants may depend on the parameters in the

subscript. For example,  $O_m(1)$  means bounded by a constant  $c$  that may depend on  $m$ . We sometimes write  $f \lesssim g$  to indicate  $f = O(g)$ .

*a) Systems of linear forms.:* A *linear form* in  $d$  variables is a vector  $L = (\lambda_1, \dots, \lambda_d) \in \mathbb{Z}^d$ . Given an abelian group  $G$ , the linear form  $L$  defines a linear map  $G^d \rightarrow G$  by  $L(x_1, \dots, x_d) := \sum_i \lambda_i x_i$ . A *system of  $m$  linear forms in  $d$  variables* is a tuple  $\mathcal{L} = (L_1, \dots, L_m)$  of linear forms. A tuple  $(a_1, \dots, a_m) \in G^m$  is called an *instance* of  $\mathcal{L}$  if there exists  $x \in G^d$  with  $\mathcal{L}(x) := (L_1(x), \dots, L_m(x)) = (a_1, \dots, a_m)$ . It is called a *non-degenerate instance* if additionally  $a_1, \dots, a_m$  are all distinct.

We often require that  $|G|$  be coprime with all the coefficients of the linear forms that define  $\mathcal{L}$ . Otherwise, it would be possible to have sets that do not contain instances of even a single linear form. For example, let  $L(x) = 2x$  and  $G = \mathbb{Z}_m$  for some even number  $m$ . Then the set  $A = \{a \in G : a \bmod 2 \equiv 1\}$  does not contain any instance of  $L(x)$  for  $x \in G$ .

For an integer  $\lambda$  with  $(\lambda, |G|) = 1$ , the map  $x \mapsto \lambda x$  is a  $G$ -automorphism. Let  $\lambda^{-1}$  be a positive integer such that  $\lambda^{-1}(\lambda x) = x$  for all  $x \in G$ . This is guaranteed to exist because  $(\lambda, |G|) = 1$  implies the existence of an integer  $\lambda^{-1}$  with  $\lambda^{-1}\lambda \equiv 1 \pmod{|G|}$ .

Let  $\mathcal{L} = (L_1, \dots, L_m)$  be a system of linear forms in  $d$  variables and let  $A \subseteq G$ . Let  $\mathbf{x}_1, \dots, \mathbf{x}_d$  be independent random variables taking values in  $G$  uniformly at random. The probability that  $\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_d) \in A^m$  is given by the expectation

$$t_{\mathcal{L}}(A) := \mathbb{E} \left[ \prod_{i=1}^m A(L_i(\mathbf{x}_1, \dots, \mathbf{x}_d)) \right].$$

*b) Translation-invariance.:* When studying linear patterns, it is natural to consider systems whose instances are invariant under translations.

**Definition 1** (Translation-Invariance). A system

$$\mathcal{L} = (L_1, \dots, L_m)$$

in  $d$  variables is *translation-invariant* over a finite abelian group  $G$  if for every instance  $(a_1, \dots, a_m)$  of  $\mathcal{L}$  and every  $c$ , the tuple  $(c + a_1, \dots, c + a_m)$  is also an instance of  $\mathcal{L}$ .

For example,  $(a, a + b, a + 2b)$ , which represents 3-progressions, is translation-invariant over every finite abelian group. As another example,  $(x-y, y-z, 2x-2z)$  is translation-invariant over  $\mathbb{Z}_3$ , but it is not translation-invariant over  $\mathbb{Z}_5$ .

Given a system  $\mathcal{L} = (L_1, \dots, L_m)$  of linear forms (not necessarily translation-invariant) in  $d$  variables  $x_1, \dots, x_d$ , we are interested in sets  $A \subseteq G$  that are free of *translations* of non-degenerate instances of  $\mathcal{L}$ . That is, there are no  $c \in G$  and non-degenerate instance  $(a_1, \dots, a_m)$  of  $\mathcal{L}$  such that  $(c + a_1, \dots, c + a_m) \in A^m$ . Equivalently, one can introduce an additional variable  $x_{d+1}$ , and require that  $A$  is free of non-

degenerate instances of the *translation-invariant* system  $\mathcal{L}'$ , where

$$\begin{aligned}\mathcal{L}'(x_1, \dots, x_{d+1}) := \\ (x_{d+1} + L_1(x_1, \dots, x_d), \dots, x_{d+1} + L_m(x_1, \dots, x_d)).\end{aligned}$$

**Definition 2** (Binary Systems of Linear forms). A system of linear forms  $\mathcal{L} = (L_1, \dots, L_m)$  in the  $d$  variables  $x_1, \dots, x_d$  is *binary* if every linear form in  $\mathcal{L}$  is supported on exactly two variables, and moreover no two linear forms in  $\mathcal{L}$  are supported on the same two variables.

In other words, every  $L_i$  is of the form  $\lambda_i x_{a_i} + \eta_i x_{b_i}$  where  $a_i < b_i$  and  $a_i, b_i \in \{1, \dots, d\}$ , and moreover,  $\{a_i, b_i\} \neq \{a_j, b_j\}$  if  $i \neq j$ . Such a binary system naturally defines an oriented acyclic graph  $H$  with vertex set  $\{1, \dots, d\}$  and an edge  $(a_i, b_i)$  for the linear form  $L_i = \lambda_i x_{a_i} + \eta_i x_{b_i}$ . We say that  $H$  is the *underlying graph* of  $\mathcal{L}$ . In the case of 3-progressions, the underlying graph  $H$  is an oriented triangle, as described below.

**Example I.7** (3-progressions as a binary system of linear forms). Let  $N > 2$  be an odd number. One can apply the change of variables  $a = 2x - 2y$  and  $b = 2y - x - z$  to  $(a, a + b, a + 2b)$  and capture 3-progressions in  $\mathbb{Z}_N$  with the system of linear forms  $\mathcal{L} = (2x - 2y, x - z, 2y - 2z)$ . The underlying graph  $H$  of  $\mathcal{L}$  is the oriented triangle  $\{(1, 2), (1, 3), (2, 3)\}$ .

## II. GRAPH COUNTING LEMMA

We will adopt a more general framework that encompasses oriented graphs. Recall that a directed graph is an *oriented graph* if it contains no opposite pairs of directed edges, or equivalently, it does not contain any 2-cycles. Let  $H = (V_H, E_H)$  be a fixed oriented graph with vertex set  $V_H = \{1, \dots, k\}$ . Let  $\mathcal{X}$  be a finite set. For every  $A \subseteq \mathcal{X} \times \mathcal{X}$ , define

$$t_H(A) := \mathbb{E} \left[ \prod_{uv \in E_H} A(\mathbf{x}_u, \mathbf{x}_v) \right],$$

where  $\mathbf{x}_v$  for  $v \in V_H$  are independent random variables taking values in  $\mathcal{X}$  uniformly at random. One can interpret  $A$  as an oriented graph with vertex set  $\mathcal{X}$  and edges  $(u, v) \in A$ . Then  $t_H(A)$  is the probability that a random map from the vertices of  $H$  to the vertices of  $A$  is a graph morphism, i.e., it maps every edge of  $H$  to an edge in  $A$ . It is worth noting that, more generally, one can define  $t_H(f)$  for a function  $f: \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ . Our counting lemma is based on the following notion of pseudo-randomness.

**Definition 3** (Spreadness for graphs). Let  $r \geq 1$  and  $\delta > 0$ , and let  $\mathcal{X}$  and  $\mathcal{Y}$  be finite sets. A set  $A \subseteq \mathcal{X} \times \mathcal{Y}$  with density  $\alpha$  is  $(\delta, r)$ -spread if, for all subsets  $S$  and  $T$  of density at least  $2^{-r}$ , we have

$$\mathbb{E}_{\mathbf{x} \sim S, \mathbf{y} \sim T} [A(\mathbf{x}, \mathbf{y})] \leq (1 + \delta)\alpha.$$

The following counting lemma states that if  $A$  is spread and does not have many vertices with small degrees, then  $t_H(A) \approx$

$\alpha^m$  for all oriented graphs  $H$  on  $m$  edges. We state the result in a more general setting.

**Theorem II.1** (Main graph counting lemma). Let  $H = ([k], E)$  be an oriented graph with  $m$  edges such that  $(u, v) \in E$  implies  $u < v$ , and let  $\alpha > 0$ . For every  $0 < \varepsilon < 1$ , there exists  $\delta = \delta(\varepsilon, m) > 0$  such that the following holds. Let  $\mathcal{X}_1, \dots, \mathcal{X}_k$  be finite sets, and for every edge  $(u, v) \in E$ , suppose  $A_{uv} \subseteq \mathcal{X}_u \times \mathcal{X}_v$  has density  $\alpha_{uv} \geq \alpha$ . If

$$\left| \mathbb{E} \prod_{(u, v) \in E} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) - \prod_{(u, v) \in E} \alpha_{uv} \right| \geq \varepsilon \prod_{(u, v) \in E} \alpha_{uv},$$

then for some  $(u, v) \in E$ , at least one of the following cases holds:

- 1) There exist  $S \subseteq \mathcal{X}_u$  and  $T \subseteq \mathcal{X}_v$  of densities at least  $2^{-O_{m, \varepsilon}(\log^2(2/\alpha))}$  such that

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S \times T} [A_{uv}(\mathbf{x}, \mathbf{y})] \geq (1 + \delta)\alpha_{uv}$$

- 2) There exists  $S \subseteq \mathcal{X}_u$  of density at least  $2^{-O_{m, \varepsilon}(\log(2/\alpha))}$  such that

$$\mathbb{E}_{\mathbf{x} \in \mathcal{X}_v} [A_{uv}(z, \mathbf{x})] \leq (1 - \delta)\alpha_{uv} \quad \forall z \in S$$

In other words, if  $|t_H(A) - \alpha^m| \geq \varepsilon \alpha^m$ , then either  $A$  contains a large subgraph that is significantly denser, or it contains many vertices whose degrees are significantly smaller than the average.

Our next theorem shows that when the goal is to identify a rectangle where the density deviates from  $\alpha$ , we do not need to require any assumption on the degrees of the vertices, and additionally, we can ensure a stronger lower bound on the density of the rectangle.

**Theorem II.2.** Let  $H = ([k], E)$  be an oriented graph with  $m$  edges such that  $(u, v) \in E$  implies  $u < v$ , and let  $\alpha > 0$ . Let  $\mathcal{X}_1, \dots, \mathcal{X}_k$  be finite sets, and for every edge  $(u, v) \in E$ , suppose  $A_{uv} \subseteq \mathcal{X}_u \times \mathcal{X}_v$  has density  $\alpha_{uv} \geq \alpha$ .

- (i) If

$$\mathbb{E} \prod_{(u, v) \in E} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \geq (1 + \varepsilon)^m \prod_{(u, v) \in E} \alpha_{uv},$$

then for some  $(u, v) \in E$ , there exist  $S \times T \subseteq \mathcal{X}_u \times \mathcal{X}_v$  with  $\frac{|S||T|}{|\mathcal{X}_u||\mathcal{X}_v|} = \Omega(\varepsilon \alpha^{m+1})$  such that

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T} [A_{uv}(\mathbf{x}, \mathbf{y})] \geq \left(1 + \frac{\varepsilon}{2}\right) \alpha_{uv}.$$

- (ii) If

$$\mathbb{E} \prod_{(u, v) \in E} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \leq (1 - \varepsilon)^m \prod_{(u, v) \in E} \alpha_{uv},$$

then for some  $(u, v) \in E$ , there exist  $S \times T \subseteq \mathcal{X}_u \times \mathcal{X}_v$  with  $\frac{|S||T|}{|\mathcal{X}_u||\mathcal{X}_v|} = \Omega(\varepsilon(1 - \varepsilon)^{m-1} \alpha^m)$  such that

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T} [A_{uv}(\mathbf{x}, \mathbf{y})] \leq \left(1 - \frac{\varepsilon}{2}\right) \alpha_{uv}.$$

This implies the following immediate corollary for bipartite graphs which we use in the proof of Theorem II.1.

**Corollary II.3.** Let  $H = (L, R, E)$  be a bipartite graph with  $m$  edges with left vertices  $L = [k]$  and right vertices  $R = [p]$ , and let  $0 < \varepsilon < 1$ . For every bipartite graph  $A: \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$  and density  $\alpha$ , the following holds. Suppose

$$\mathbb{E}_{\substack{\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{X} \\ \mathbf{y}_1, \dots, \mathbf{y}_p \in \mathcal{Y}}} \left[ \prod_{\{i, j\} \in E} A(\mathbf{x}_i, \mathbf{y}_j) \right] \geq (1 + \varepsilon)^m \alpha^m.$$

Then there exist  $S \subseteq \mathcal{X}$  and  $T \subseteq \mathcal{Y}$  with  $\frac{|S||T|}{|\mathcal{X}||\mathcal{Y}|} = \Omega(\varepsilon\alpha^{m+1})$  such that

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim S \times T} [A(\mathbf{x}, \mathbf{y})] \geq \left(1 + \frac{\varepsilon}{2}\right) \alpha.$$

As mentioned in the introduction, the case where  $H$  is a complete bipartite graph was proven in [12].

#### A. Proof of Theorem II.2

To prove Theorem II.2, we need the following lemma, which is a generalization of [12, Claim 4.6]. Let  $\mathcal{X}, \mathcal{Y}$  be finite sets, define

$$\mathcal{R} := \{S \times T : S \subseteq \mathcal{X}, T \subseteq \mathcal{Y}\},$$

and let  $\text{conv}(\mathcal{R})$  denote the convex hull of  $\mathcal{R}$ .

**Lemma II.4.** Let  $\Delta \geq 1$  and  $\gamma, \varepsilon \in (0, 1]$ . Let  $D: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$  and  $F \in \text{conv}(\mathcal{R})$ , and suppose  $\|D\|_{\infty} \leq \Delta$  and  $\|F\|_1 \geq \gamma$ .

(i) If  $\left\langle \frac{F}{\|F\|_1}, D \right\rangle \geq 1 + \varepsilon$ , there exists a rectangle  $S \times T$  with  $\frac{|S||T|}{|\mathcal{X}||\mathcal{Y}|} \geq \frac{\varepsilon\gamma}{4\Delta}$  and

$$\mathbb{E}_{(x, y) \in S \times T} D(x, y) \geq 1 + \frac{\varepsilon}{2}.$$

(ii) If  $\left\langle \frac{F}{\|F\|_1}, D \right\rangle \leq 1 - \varepsilon$ , there exists a rectangle  $S \times T$  with  $\frac{|S||T|}{|\mathcal{X}||\mathcal{Y}|} \geq \frac{\varepsilon\gamma}{4\Delta}$  and

$$\mathbb{E}_{(x, y) \in S \times T} D(x, y) \leq 1 - \frac{\varepsilon}{2}.$$

*Proof.* Let  $F = \sum_i c_i R_i$  for rectangles  $R_i \in \mathcal{R}$  and  $c_i \geq 0$  with  $\sum_i c_i = 1$ . Let  $\tau = \frac{\varepsilon\gamma}{4\Delta}$ , and set  $F := \sum_i c'_i R_i$  with

$$c'_i = \begin{cases} c_i & \text{if } \|R_i\|_1 \geq \tau, \\ 0 & \text{if } \|R_i\|_1 < \tau. \end{cases}$$

Note  $\|F - F'\|_1 \leq \tau$ .

**Part (i):** In this case, we have

$$\begin{aligned} \frac{\langle F', D \rangle}{\|F'\|_1} &\geq \frac{\langle F', D \rangle}{\|F\|_1} = \frac{\langle F, D \rangle}{\|F\|_1} + \frac{\langle F' - F, D \rangle}{\|F\|_1} \geq \\ &1 + \varepsilon - \frac{\|F' - F\|_1 \|D\|_{\infty}}{\|F\|_1} \geq 1 + \varepsilon - \frac{\tau\Delta}{\gamma} \geq 1 + \frac{3\varepsilon}{4}. \end{aligned}$$

The definition of  $F'$  implies there is a choice of  $R = R_i$  with  $\|R\|_1 \geq \tau$  and  $\left\langle \frac{R}{\|R\|_1}, D \right\rangle \geq 1 + \frac{3\varepsilon}{4}$ .

**Part (ii):** We have

$$\begin{aligned} \frac{\langle F', D \rangle}{\|F\|_1} &= \frac{\langle F, D \rangle}{\|F\|_1} + \frac{\langle F' - F, D \rangle}{\|F\|_1} \leq \\ &1 - \varepsilon + \frac{\|F' - F\|_1 \|D\|_{\infty}}{\|F\|_1} \leq 1 - \varepsilon + \frac{\tau\Delta}{\gamma} \leq 1 - \frac{3\varepsilon}{4}. \end{aligned}$$

Furthermore,

$$\|F'\|_1 \geq \|F\|_1 - \tau \geq \|F\|_1 - \frac{\varepsilon\|F\|_1}{4} = \left(1 - \frac{\varepsilon}{4}\right) \|F\|_1.$$

We get

$$\frac{\langle F', D \rangle}{\|F'\|_1} \leq \frac{1 - (3\varepsilon/4)}{1 - (\varepsilon/4)} \leq 1 - \frac{\varepsilon}{2}.$$

The definition of  $F'$  implies there is a choice of  $R = R_i$  with  $\|R\|_1 \geq \tau$  and  $\left\langle \frac{R}{\|R\|_1}, D \right\rangle \leq 1 - \frac{\varepsilon}{2}$ .  $\square$

We can now prove Theorem II.2.

*Proof of Theorem II.2..* We will prove the theorem by induction on  $m$ . The base case  $m = 1$  is trivial.

Pick any edge  $(u_0, v_0) \in E$  and let  $E' := E - \{(u_0, v_0)\}$  and  $H' := (V, E')$ . Let  $V' = V \setminus \{u_0, v_0\}$  and let  $E''$  be the edges with both endpoints in  $V'$ . For  $(x, y) \in \mathcal{X}_{u_0} \times \mathcal{X}_{v_0}$ , define

$$F(x, y) := \mathbb{E}_{\mathbf{x}_w \in \mathcal{X}_w} \left[ \prod_{(u, v) \in E''} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right. \\ \left. \prod_{(u_0, v) \in E'} A_{u_0 v}(x, \mathbf{x}_v) \prod_{(u, v_0) \in E'} A_{u v_0}(\mathbf{x}_u, y) \right].$$

Note that  $F$  is an average of rectangles: each of the factors may depend on  $x$  or  $y$  but not both. Therefore,  $F \in \text{conv}(\mathcal{R})$ .

**Part (i):** By applying the induction hypothesis to  $H'$ , we may assume

$$\|F\|_1 = \mathbb{E} \prod_{(u, v) \in E'} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \leq (1 + \varepsilon)^{m-1} \prod_{(u, v) \in E'} \alpha_{uv},$$

as otherwise, we are done. Since

$$\langle F, A_{u_0 v_0} \rangle = \mathbb{E} \prod_{(u, v) \in E} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \geq (1 + \varepsilon)^m \prod_{uv \in E} \alpha_{uv},$$

by setting  $D = \frac{A_{u_0 v_0}}{\|A_{u_0 v_0}\|_1}$ , we have

$$\left\langle \frac{F}{\|F\|_1}, D \right\rangle \geq \frac{(1 + \varepsilon)^m \prod_{uv \in E} \alpha_{uv}}{\alpha_{u_0 v_0} \|F\|_1} \geq 1 + \varepsilon.$$

Since  $\|D\|_{\infty} \leq \frac{1}{\|A_{u_0 v_0}\|_1} \leq \frac{1}{\alpha}$  and by our assumption  $\|F\|_1 \geq \mathbb{E} \prod_{(u, v) \in E} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \geq \alpha^m$ , Lemma II.4 provides a rectangle  $S \times T \subseteq \mathcal{X}_{u_0} \times \mathcal{X}_{v_0}$  with  $\frac{|S||T|}{|\mathcal{X}_{u_0}||\mathcal{X}_{v_0}|} = \Omega(\varepsilon\alpha^{m+1})$  and

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S \times T} A_{u_0 v_0}(\mathbf{x}, \mathbf{y}) \geq (1 + \frac{\varepsilon}{2}) \alpha_{u_0 v_0}.$$

**Part (ii):** By applying the induction hypothesis to  $H'$ , we may assume

$$\|F\|_1 = \mathbb{E} \prod_{(u, v) \in E'} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) \geq (1 - \varepsilon)^{m-1} \prod_{(u, v) \in E'} \alpha_{uv},$$

as otherwise, we are done. In this case, we have

$$\left\langle \frac{F}{\|F\|_1}, D \right\rangle \leq \frac{(1-\varepsilon)^m \prod_{(u,v) \in E} \alpha_{uv}}{(1-\varepsilon)^{m-1} \prod_{(u,v) \in E} \alpha_{uv}} \leq 1 - \varepsilon.$$

Since  $\|D\|_\infty \leq \frac{1}{\alpha}$  and  $\|F\|_1 \geq (1-\varepsilon)^{m-1} \alpha^{m-1}$ , Lemma II.4 provides a rectangle  $S \times T$  with  $\frac{|S||T|}{|\mathcal{X}_{u_0}||\mathcal{X}_{v_0}|} = \Omega(\varepsilon(1-\varepsilon)^{m-1} \alpha^m)$  and  $\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S \times T} A_{u_0 v_0}(\mathbf{x}, \mathbf{y}) \leq (1 - \frac{\varepsilon}{2}) \alpha_{u_0 v_0}$ .  $\square$

### B. Proof of Theorem II.1

The proof of Theorem II.1 is more involved and requires some preparation.

a) *Grid norms*.: One of the striking results in the work of Chung, Graham and Wilson [2] states that if an undirected graph with edge density  $\alpha$  has approximately the same number of 4-cycles as the Erdős–Rényi random graph  $G(n, \alpha)$ , then it shares several structural, spectral and statistical properties with  $G(n, \alpha)$ .

The pseudo-randomness theorem of [2] naturally generalizes to binary relations  $A \subseteq \mathcal{X} \times \mathcal{Y}$ . A set  $A \subseteq \mathcal{X} \times \mathcal{Y}$  exhibits any of several pseudo-random properties if  $t_{K_{2,2}}^b(A) = \alpha^4 + o(1)$  where

$$t_{K_{2,2}}^b(A) := \mathbb{E}_{\substack{\mathbf{x} \sim \mathcal{X}^k \\ \mathbf{y} \sim \mathcal{Y}^p}} \left[ \prod_{i=1}^k \prod_{j=1}^p A(\mathbf{x}_i, \mathbf{y}_j) \right].$$

One of the key ideas used in the work of Kelley and Meka, and more explicitly in the follow-up work of Kelley, Lovett, and Meka [12], is to consider a certain generalization of  $t_{K_{2,2}}(A)$ .

**Definition 4** (Grid semi-norms). For a function  $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  and  $k, p \in \mathbb{N}$ , let

$$U_{k,p}(f) := \mathbb{E}_{\substack{\mathbf{x} \sim \mathcal{X}^k \\ \mathbf{y} \sim \mathcal{Y}^p}} \left[ \prod_{i=1}^k \prod_{j=1}^p f(\mathbf{x}_i, \mathbf{y}_j) \right].$$

Also define  $\|f\|_{U(k,p)} := |U_{k,p}(f)|^{1/pk}$  and  $\|f\|_{|U(k,p)|} := U_{k,p}(|f|)^{1/pk}$ .

**Remark 4.** Note that if  $A: \mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$ , then  $t_{K_{2,2}}^b(A) = U_{k,p}(A)$ .

Lemma II.5 below is the analogue of the so-called Gowers–Cauchy–Schwarz inequality for  $U(k, p)$  norms.

**Lemma II.5. (Gowers – Cauchy – Schwarz inequality)** Let  $p, k \in \mathbb{N}$ , and for  $i \in [k]$  and  $j \in [p]$ , let  $f_{ij}: \mathcal{X}_i \times \mathcal{Y}_j \rightarrow \mathbb{R}$ . Suppose that either  $k, p$  are both even numbers, or the functions  $f_{ij}$  take only non-negative values. Then

$$\mathbb{E}_{\mathbf{x}_1 \sim \mathcal{X}_1, \dots, \mathbf{x}_k \sim \mathcal{X}_k} \mathbb{E}_{\mathbf{y}_1 \sim \mathcal{Y}_1, \dots, \mathbf{y}_p \sim \mathcal{Y}_p} \left[ \prod_{i=1}^k \prod_{j=1}^p f_{ij}(\mathbf{x}_i, \mathbf{y}_j) \right] \leq \prod_{i=1}^k \prod_{j=1}^p \|f_{ij}\|_{U(k,p)}. \quad (7)$$

*Proof.* Recall the generalized Hölder inequality:

$$\mathbb{E}_{\mathbf{z}} \left| \prod_{j=1}^p g_j(\mathbf{z}) \right| \leq \prod_{j=1}^p (\mathbb{E}_{\mathbf{z}} |g_j(\mathbf{z})|^p)^{1/p}.$$

Choosing  $z = (x_1, \dots, x_k)$  and

$$g_j(x_1, \dots, x_k) = \mathbb{E}_{\mathbf{y}_j \sim \mathcal{Y}_j} \prod_i f_{ij}(x_i, \mathbf{y}_j),$$

we obtain

L.H.S. of (7)  $\leq$

$$\prod_{j=1}^p \left( \mathbb{E}_{\mathbf{x}_1 \sim \mathcal{X}_1, \dots, \mathbf{x}_k \sim \mathcal{X}_k} \left( \mathbb{E}_{\mathbf{y}_j \sim \mathcal{Y}_j} \prod_{i=1}^k f_{ij}(\mathbf{x}_i, \mathbf{y}_j) \right)^p \right)^{1/p}. \quad (8)$$

Next, note that for every fixed  $j$ , we can apply the generalized Hölder inequality to the variable  $\mathbf{y} \sim \mathcal{Y}_j^p$  and obtain

$$\begin{aligned} & \mathbb{E}_{\mathbf{x}_1 \sim \mathcal{X}_1, \dots, \mathbf{x}_k \sim \mathcal{X}_k} \left( \mathbb{E}_{\mathbf{y} \sim \mathcal{Y}_j^p} \prod_{i=1}^k f_{ij}(\mathbf{x}_i, \mathbf{y}) \right)^p \\ &= \mathbb{E}_{\mathbf{y} \sim \mathcal{Y}_j^p} \mathbb{E}_{\mathbf{x}_1 \sim \mathcal{X}_1, \dots, \mathbf{x}_k \sim \mathcal{X}_k} \prod_{i=1}^k \prod_{t=1}^p f_{ij}(\mathbf{x}_i, \mathbf{y}_t) \\ &= \mathbb{E}_{\mathbf{y} \sim \mathcal{Y}_j^p} \prod_{i=1}^k \left( \mathbb{E}_{\mathbf{x}_i \sim \mathcal{X}_i} \prod_{t=1}^p f_{ij}(\mathbf{x}_i, \mathbf{y}_t) \right) \\ &\leq \prod_{i=1}^k \left( \mathbb{E}_{\mathbf{y} \sim \mathcal{Y}_j^p} \left( \mathbb{E}_{\mathbf{x} \sim \mathcal{X}_i} \prod_{t=1}^p f_{ij}(\mathbf{x}, \mathbf{y}_t) \right)^k \right)^{1/k} \\ &= \prod_{i=1}^k \|f_{ij}\|_{U(k,p)}^p. \end{aligned}$$

Substituting these in (8), we obtain the desired inequality.  $\square$

Applying Lemma II.5 to the expansion of  $U_{k,p}(f+g)$  easily implies that  $\|\cdot\|_{|U(k,p)|}$  is a norm, and that when  $p$  and  $k$  are both even,  $\|\cdot\|_{U(k,p)}$  is a semi-norm.

**Corollary II.6** ([12, Claim 4.2]). *If  $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$  then  $\|f\|_{U(k,p)}$  is monotone in both  $k$  and  $p$ .*

*Proof.* We will prove monotonicity for  $k$ ; the same argument can be applied to  $p$ . Let  $k' \geq k$ , and define  $f_{ij}: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_{\geq 0}$  for  $1 \leq i \leq k'$  and  $1 \leq j \leq p$  as

$$f_{ij} = \begin{cases} f & i \leq k, \\ 1 & i > k. \end{cases}$$

Applying (7) to these functions, we have

$$\begin{aligned} \|f\|_{U(k,p)}^{kp} &= \mathbb{E}_{\mathbf{x}_1, \dots, \mathbf{x}_k} \mathbb{E}_{\mathbf{y}_1, \dots, \mathbf{y}_p} \left[ \prod_{i=1}^{k'} \prod_{j=1}^p f_{ij}(\mathbf{x}_i, \mathbf{y}_j) \right] \leq \\ &\quad \|f\|_{U(k',p)}^{kp} \|1\|_{U(k',p)}^{k'p-kp} = \|f\|_{U(k',p)}^{kp}, \end{aligned}$$

as desired.  $\square$

We also need the following simple lemma.

**Lemma II.7.** *Let  $p \geq 1$ , and let  $f: \mathcal{X} \rightarrow [0, 1]$  satisfy  $\mathbb{E}[f] = \alpha$  and  $\|f - \alpha\|_p \geq \varepsilon\alpha$ . Then one of the following two cases holds:*

- (i)  $\mathbb{E}_{\mathbf{x}} \left[ \mathbf{1}_{[f(\mathbf{x}) > \alpha(1 + \frac{\varepsilon}{4})]} \right] \geq \varepsilon^p \alpha^p / 4$ .
- (ii)  $\mathbb{E}_{\mathbf{x}} \left[ \mathbf{1}_{[f(\mathbf{x}) < \alpha(1 - \frac{\varepsilon}{4})]} \right] \geq \varepsilon^p / 4$ .

*Proof.* From the assumption,

$$\mathbb{E}_{\mathbf{x}} |f(\mathbf{x}) - \alpha|^p = \|f - \alpha\|_p^p \geq \varepsilon^p \alpha^p.$$

We have

$$\mathbb{E} \left[ |f(\mathbf{x}) - \alpha|^p \mathbf{1}_{[|f(\mathbf{x}) - \alpha| \leq \frac{\varepsilon\alpha}{4}]} \right] \leq (\varepsilon/4)^p \alpha^p \leq \frac{\varepsilon^p}{4} \alpha^p,$$

and therefore,

$$\begin{aligned} \frac{3\varepsilon^p}{4} \alpha^p &\leq \mathbb{E} \left[ |f(\mathbf{x}) - \alpha|^p \mathbf{1}_{[f(\mathbf{x}) > \alpha(1 + \frac{\varepsilon}{4})]} \right] \\ &\quad + \mathbb{E} \left[ |f(\mathbf{x}) - \alpha|^p \mathbf{1}_{[f(\mathbf{x}) < \alpha(1 - \frac{\varepsilon}{4})]} \right] \\ &\leq \mathbb{E} \left[ \mathbf{1}_{[f(\mathbf{x}) > \alpha(1 + \frac{\varepsilon}{4})]} \right] + \alpha^p \mathbb{E} \left[ \mathbf{1}_{[f(\mathbf{x}) < \alpha(1 - \frac{\varepsilon}{4})]} \right], \end{aligned}$$

which implies that one of the two cases (i) or (ii) must hold.  $\square$

*b) Main technical lemma.:* Next, we prove the following lemma, which constitutes the technical core of the proof of Theorem II.1.

**Lemma II.8** (Main technical lemma). *Let  $H = ([k], E)$  be an oriented graph with  $m$  edges such that  $(u, v) \in E$  implies  $u < v$ . For every  $0 < \varepsilon < 1$ , there exists  $\delta = \delta(\varepsilon, H) \geq \varepsilon^2 2^{-4m-10}$  such that the following holds. Let  $d$  be the maximum indegree of a vertex in  $H$  and let  $\alpha > 0$ . Let  $\mathcal{X}_1, \dots, \mathcal{X}_k$  be finite sets, and for every edge  $(u, v) \in E$ , let  $f_{uv}: \mathcal{X}_u \times \mathcal{X}_v \rightarrow \{0, 1\}$  with  $\mathbb{E}[f_{uv}] = \alpha_{uv} \geq \alpha$  be such that*

$$\left| \mathbb{E} \prod_{(u,v) \in E} f_{uv}(\mathbf{x}_u, \mathbf{x}_v) - \prod_{(u,v) \in E} \alpha_{uv} \right| \geq \varepsilon \prod_{(u,v) \in E} \alpha_{uv}. \quad (9)$$

*There exists  $(u, v) \in E$  and  $p \lesssim \frac{m}{\varepsilon} \log(2/\alpha)$  such that at least one of the following holds:*

$$(i) \quad \|f_{uv}\|_{U(2(d-1), p)} \geq (1 + \delta)\alpha_{uv}. \quad (10)$$

$$(ii) \quad \left\| \mathbb{E}_{\mathbf{x}_v \sim \mathcal{X}_v} f_{uv}(\cdot, \mathbf{x}_v) - \alpha_{uv} \right\|_p \geq \delta\alpha_{uv}. \quad (11)$$

*Proof.* We can assume that  $H$  contains no isolated vertices. We can also assume that  $H$  contains no isolated edges. Indeed, suppose that  $H$  contains an isolated edge  $e$ , and let  $H'$  be obtained from  $H$  by removing  $e$ . If (9) holds for  $H$  then it also holds for  $H'$ . In this way, we can remove all isolated edges from  $H$ .

We will prove the theorem by induction on the quantity  $\kappa(H) := 2|E| - d_1(H)$ , where  $d_1(H)$  is the number of vertices of degree 1 in  $H$ . More precisely, the induction hypothesis states that the theorem holds with  $\delta(\varepsilon, H) = \varepsilon^2 2^{-2\kappa(H)-10}$ .

To verify the base case  $\kappa(H) = 0$ , note that in this case,  $H$  is empty, and so the L.H.S. of (9) is always 0, and the statement is trivially true.

We normalize the functions  $f_{uv}: \mathcal{X}_u \times \mathcal{X}_v \rightarrow \{0, 1\}$  by defining  $F_{uv}: \mathcal{X}_u \times \mathcal{X}_v \rightarrow \{0, \alpha_{uv}^{-1}\}$  as  $F_{uv} := \alpha_{uv}^{-1} \cdot f_{uv}$ . We can rephrase the assumption, (9), as

$$\left| \mathbb{E} \prod_{uv \in E} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) - 1 \right| \geq \varepsilon. \quad (12)$$

Consider the vertex  $s = k$ , which is a sink in the graph since  $H$  is topologically ordered. Let  $u_1, \dots, u_\ell$  be the neighbors of  $s$ , where  $1 \leq \ell \leq d$ . Let  $H' = (V \setminus \{s\}, E')$  where  $E' \subsetneq E$  is the set of the edges that do not involve  $s$ .

Since  $H$  has no isolated edges,  $d_1(H') \geq d_1(H) - \ell$ , and so

$$\begin{aligned} \kappa(H') &= 2|E'| - d_1(H') \leq 2(|E| - \ell) - (d_1(H) - \ell) \\ &= \kappa(H) - \ell. \end{aligned}$$

By applying the induction hypothesis to  $H'$ , we may assume

$$\left| \mathbb{E} \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) - 1 \right| \leq \frac{\varepsilon}{2}, \quad (13)$$

as otherwise, we are done with

$$\begin{aligned} \delta &= \delta(\varepsilon/2, H') = (\varepsilon/2)^2 2^{-2\kappa(H')-10} \geq \\ &\quad \varepsilon^2 2^{-2\kappa(H)-10} = \delta(\varepsilon, H), \end{aligned}$$

as desired. In particular, we may assume

$$\mathbb{E} \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \leq 2.$$

Therefore, using  $F_{uv} \in \{0, \alpha_{uv}^{-1}\}$ , we have for every  $q \geq 1$ ,

$$\begin{aligned} \mathbb{E} \left[ \left( \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right)^q \right] &= \\ \mathbb{E} \left[ \prod_{uv \in E'} \alpha_{uv}^{-(q-1)} \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] &\leq \\ 2 \cdot \prod_{uv \in E'} \alpha_{uv}^{-(q-1)} &\leq 2 \cdot \alpha^{m(1-q)}. \quad (14) \end{aligned}$$

We will single out the edge  $u_1 s$  and set  $E'' := E \setminus \{u_1 s\}$ .

*c) Case 1:  $\ell = 1$ .* In this case,  $E'' = E'$ . Combining (13) and (12), we have

$$\begin{aligned} \left| \mathbb{E} \left[ (F_{u_1 s}(\mathbf{x}_{u_1}, \mathbf{x}_s) - 1) \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] \right| &= \\ \left| \mathbb{E} \prod_{uv \in E} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) - \mathbb{E} \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right| &\geq \frac{\varepsilon}{2}. \end{aligned}$$

Consequently, we may apply Hölder's inequality over the variables  $\mathbf{x} := \{\mathbf{x}_u\}_{u \in V \setminus \{s\}}$  with the even integer  $p = 2m \lceil \log(2\alpha^{-1}) \rceil \geq 2m \log(2\alpha^{-1})$  and  $\frac{1}{p} + \frac{1}{q} = 1$ , obtaining

$$\begin{aligned} \frac{\varepsilon}{2} &\leq \left| \mathbb{E} \left[ (F_{u_1s}(\mathbf{x}_{u_1}, \mathbf{x}_s) - 1) \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] \right| \\ &\leq \mathbb{E}_{\mathbf{x}} \left[ \left( \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right)^q \right]^{1/q} \\ &\quad \times \mathbb{E}_{\mathbf{x}} [(\mathbb{E}_{\mathbf{x}_s} (F_{u_1s}(\mathbf{x}_{u_1}, \mathbf{x}_s) - 1))^p]^{1/p} \\ &\leq 3 \cdot \mathbb{E}_{\mathbf{x}} [(\mathbb{E}_{\mathbf{x}_s} (F_{u_1s}(\mathbf{x}_{u_1}, \mathbf{x}_s) - 1))^p]^{1/p}, \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} \left( 2 \cdot \alpha^{m(1-q)} \right)^{1/q} &\leq 2 \cdot \alpha^{m(\frac{1}{q}-1)} = 2 \cdot \alpha^{-\frac{m}{p}} \leq \\ &2 \cdot \alpha^{-\frac{m}{2m \log(2\alpha^{-1})}} \leq 3. \quad (15) \end{aligned}$$

We conclude that

$$\|\mathbb{E}_{\mathbf{y}} F_{u_1s}(\cdot, \mathbf{y}) - 1\|_p \geq \varepsilon/6 \geq \delta,$$

which verifies Equation (11).

*d) Case 2:  $\ell \geq 2$ .*: Consider the graph  $H''$  obtained from  $H$  by “detaching” the edge  $u_1s$  from the node  $s$ . More formally, to obtain  $H''$ , we add a new vertex  $s'$  to  $H$ , remove the edge  $u_1s$ , and add the edge  $u_1s'$  instead. Note that  $\kappa(H'') \leq \kappa(H) - 1$ . By applying the induction hypothesis to  $H''$  with parameter  $\varepsilon/2$ , we may assume that

$$\begin{aligned} \left| \mathbb{E}_{\mathbf{x}} \left[ \left( \prod_{uv \in E''} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right) \cdot (\mathbb{E}_{\mathbf{z} \sim \mathcal{X}_s} F_{u_1s}(\mathbf{x}_{u_1}, \mathbf{z})) \right] - 1 \right| \\ \leq \frac{\varepsilon}{2}, \quad (16) \end{aligned}$$

as otherwise, the assertion of the theorem follows from the induction hypothesis with parameter  $\delta = \delta(\varepsilon/2, H'') \geq \delta(\varepsilon, H)$ , as desired. Set  $J(x, y) := F_{u_1s}(x, y) - \mathbb{E}_{\mathbf{z} \sim \mathcal{X}_s} F_{u_1s}(x, \mathbf{z})$ . We have

$$\begin{aligned} \mathbb{E} \left[ J(\mathbf{x}_{u_1}, \mathbf{x}_s) \prod_{uv \in E''} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] &= \\ &\mathbb{E} \prod_{uv \in E''} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) - \\ &\mathbb{E}_{\mathbf{x}} \left[ \left( \prod_{uv \in E''} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right) \cdot \mathbb{E}_{\mathbf{z} \sim \mathcal{X}_s} F_{u_1s}(\mathbf{x}_{u_1}, \mathbf{z}) \right]. \end{aligned}$$

Combining the above equation with (12) and (16), we obtain

$$\left| \mathbb{E} \left[ J(\mathbf{x}_{u_1}, \mathbf{x}_s) \prod_{uv \in E''} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] \right| \geq \frac{\varepsilon}{2}.$$

We may apply Hölder's inequality over the variables  $\mathbf{x} := \{\mathbf{x}_u\}_{u \in V \setminus \{s\}}$  with the even integer  $p = 2m \lceil \log(2\alpha^{-1}) \rceil \geq 2m \log(2\alpha^{-1})$  and  $\frac{1}{p} + \frac{1}{q} = 1$ , obtaining

$$\begin{aligned} &\left| \mathbb{E} \left[ J(\mathbf{x}_{u_1}, \mathbf{x}_s) \prod_{uv \in E''} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] \right| \\ &= \left| \mathbb{E}_{\mathbf{x}} \left[ \left( \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right) \right. \right. \\ &\quad \cdot \mathbb{E}_{\mathbf{x}_s} \left[ J(\mathbf{x}_{u_1}, \mathbf{x}_s) \cdot \prod_{uv \in E'' \setminus E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right] \left. \right] \left. \right| \\ &\leq \mathbb{E}_{\mathbf{x}} \left[ \left( \prod_{uv \in E'} F_{uv}(\mathbf{x}_u, \mathbf{x}_v) \right)^q \right]^{1/q} \\ &\quad \times \mathbb{E}_{\mathbf{x}} \left[ \left( \mathbb{E}_{\mathbf{x}_s} J(\mathbf{x}_{u_1}, \mathbf{x}_s) \cdot \prod_{i=2}^{\ell} F_{u_is}(\mathbf{x}_{u_i}, \mathbf{x}_s) \right)^p \right]^{1/p} \\ &\stackrel{(*)}{\leq} 3 \cdot \mathbb{E}_{\mathbf{x}} \left[ \left( \mathbb{E}_{\mathbf{x}_s} J(\mathbf{x}_{u_1}, \mathbf{x}_s) \cdot \prod_{i=2}^{\ell} F_{u_is}(\mathbf{x}_{u_i}, \mathbf{x}_s) \right)^p \right]^{1/p}, \end{aligned}$$

where  $(*)$  follows from (14) and (15). We conclude that

$$\mathbb{E}_{\mathbf{x}} \left[ \left( \mathbb{E}_{\mathbf{x}_s} J(\mathbf{x}_{u_1}, \mathbf{x}_s) \prod_{i=2}^{\ell} F_{u_is}(\mathbf{x}_{u_i}, \mathbf{x}_s) \right)^p \right]^{1/p} \geq \frac{\varepsilon}{6}. \quad (17)$$

On the other hand, by Cauchy–Schwarz and then (7), we have

$$\begin{aligned} &\mathbb{E}_{\mathbf{x}} \left[ \left( \mathbb{E}_{\mathbf{x}_s} J(\mathbf{x}_{u_1}, \mathbf{x}_s) \prod_{i=2}^{\ell} F_{u_is}(\mathbf{x}_{u_i}, \mathbf{x}_s) \right)^p \right] \\ &= \mathbb{E}_{\mathbf{y}_1, \dots, \mathbf{y}_p} \left[ \mathbb{E}_{\mathbf{x}_{u_1}} \prod_{j=1}^p J(\mathbf{x}_{u_1}, \mathbf{y}_j) \right. \\ &\quad \cdot \mathbb{E}_{\mathbf{x}_{u_2}, \dots, \mathbf{x}_{u_\ell}} \prod_{i=2}^{\ell} \prod_{j=1}^p F_{u_is}(\mathbf{x}_{u_i}, \mathbf{y}_j) \left. \right] \\ &\leq \|J\|_{U(2,p)}^p \\ &\quad \cdot \left[ \mathbb{E}_{\mathbf{y}_1, \dots, \mathbf{y}_p} \left( \mathbb{E}_{\mathbf{x}_2, \dots, \mathbf{x}_\ell} \prod_{i=2}^{\ell} \prod_{j=1}^p F_{u_is}(\mathbf{x}_i, \mathbf{y}_j) \right)^2 \right]^{1/2} \\ &\leq \|J\|_{U(2,p)}^p \prod_{i=2}^{\ell} \|F_{u_is}\|_{U(2(\ell-1),p)}^p. \end{aligned}$$

Taking the  $p$ -th root of this upper bound and combining it with (17), we get

$$\|J\|_{U(2,p)} \prod_{i=2}^{\ell} \|F_{u_is}\|_{U(2(\ell-1),p)} \geq \frac{\varepsilon}{6}.$$

If there exists  $i \in \{2, \dots, \ell\}$  with  $\|F_{u_is}\|_{U(2(\ell-1),p)} \geq 2$ , we are done by monotonicity of grid norms (Corollary II.6) with  $2(d-1) \geq 2(\ell-1)$ , and otherwise,

$$\|J\|_{U(2,p)} \geq \frac{\varepsilon}{6 \cdot 2^{\ell-1}}.$$

Note that  $\mathbb{E}_{\mathbf{z} \sim \mathcal{X}_s} J(x, \mathbf{z}) = 0$  for every  $x \in \mathcal{X}_{u_1}$ . Therefore, we can apply Lemma II.9 (stated and proved below) to  $J$  and obtain that for  $p' = 2p \lceil 1/\varepsilon \rceil$ , we have

$$\|1 + J\|_{U(2,p')} \geq 1 + \frac{\varepsilon^2}{45 \cdot 2^{2\ell}} \geq 1 + 2\delta.$$

The last inequality is true because

$$\kappa(H) = \sum_{v \in V: \deg(v) > 1} \deg(v) \geq \deg(s) = \ell.$$

Since  $J(x, y) = F_{u_1s}(x, y) - D(x, y)$  with  $D(x, y) := \mathbb{E}_{\mathbf{z} \sim \mathcal{X}_s} F_{u_1s}(x, \mathbf{z})$ , applying the triangle inequality, we have

$$\|F_{u_1s}\|_{U(2,p')} + \|D - 1\|_{U(2,p')} \geq 1 + 2\delta.$$

Now either

$$\|F_{u_1s}\|_{U(2,p')} \geq 1 + \delta,$$

in which case, we are done, or

$$\|D - 1\|_{U(2,p')} \geq \delta.$$

Since  $D(x, y) - 1$  does not depend on  $y$ , we have  $\|D - 1\|_{U(2,p')} = \|D - 1\|_2$ , and thus in this latter case, we get

$$\|D - 1\|_2 \geq \delta,$$

which verifies Equation (11).  $\square$

It remains to prove Lemma II.9 below.

**Lemma II.9.** *If  $g: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$  satisfies  $\|g\|_{U(2,p)}^2 \geq \varepsilon$  for some even  $p$  and some  $\varepsilon \in (0, 1/2]$  and*

$$\mathbb{E}_{\mathbf{z}} g(x, \mathbf{z}) = 0 \quad \forall x \in \mathcal{X},$$

*then for every even  $p' \geq \frac{2p}{\varepsilon}$ , we have  $\|1 + g\|_{U(2,p')} \geq 1 + \varepsilon/5$ .*

The proof will make use of the following result of Kelley and Meka.

**Proposition II.10** ([19, Proposition 5.7]). *Suppose  $f: \Omega \rightarrow \mathbb{R}$  is such that*

- $\mathbb{E}(f - 1)^k \geq 0$  for all odd  $k \in \mathbb{N}$ , and
- $\|f - 1\|_{k_0} \geq \varepsilon$  for some even  $k_0 \geq 2$  and some  $\varepsilon \in [0, 1/2]$ .

*Then, for any integer  $k' \geq 2k_0/\varepsilon$ ,*

$$\|f\|_{k'} \geq 1 + \frac{\varepsilon}{2}.$$

*Proof of Lemma II.9.* Let  $g \circ g: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$  be defined as  $(g \circ g)(x_1, x_2) := \mathbb{E}_{\mathbf{y}} g(x_1, \mathbf{y})g(x_2, \mathbf{y})$ . Note that

$$\|g\|_{U(2,p)}^2 = \|g \circ g\|_p.$$

Moreover, for every  $q \in \mathbb{N}$ ,

$$\mathbb{E}(g \circ g)^q = \mathbb{E}_{\mathbf{y}_1, \dots, \mathbf{y}_q} \left( \mathbb{E}_{\mathbf{x}} \prod_{i=1}^q g(\mathbf{x}, \mathbf{y}_i) \right)^2 \geq 0.$$

It follows from Proposition II.10 that for every integer  $p' \geq \frac{2p}{\varepsilon}$ ,

$$\|g \circ g + 1\|_{p'} \geq 1 + \frac{\varepsilon}{2}.$$

Moreover,  $\mathbb{E}_{\mathbf{y}} g(\cdot, \mathbf{y}) \equiv 0$  implies that  $(g + 1) \circ (g + 1) = (g \circ g) + 1$ , and thus

$$\|1 + g\|_{U(2,p')} = \|(g + 1) \circ (g + 1)\|_{p'}^{1/2} \geq \sqrt{1 + \frac{\varepsilon}{2}} \geq 1 + \frac{\varepsilon}{5}. \square$$

We are now ready to deduce Theorem II.1 from Lemma II.8 (main technical lemma). We repeat the statement of Theorem II.1 for convenience.

**Theorem II.1** (Main graph counting lemma). *Let  $H = ([k], E)$  be an oriented graph with  $m$  edges such that  $(u, v) \in E$  implies  $u < v$ , and let  $\alpha > 0$ . For every  $0 < \varepsilon < 1$ , there exists  $\delta = \delta(\varepsilon, m) > 0$  such that the following holds. Let  $\mathcal{X}_1, \dots, \mathcal{X}_k$  be finite sets, and for every edge  $(u, v) \in E$ , suppose  $A_{uv} \subseteq \mathcal{X}_u \times \mathcal{X}_v$  has density  $\alpha_{uv} \geq \alpha$ . If*

$$\left| \mathbb{E} \prod_{(u,v) \in E} A_{uv}(\mathbf{x}_u, \mathbf{x}_v) - \prod_{(u,v) \in E} \alpha_{uv} \right| \geq \varepsilon \prod_{(u,v) \in E} \alpha_{uv},$$

*then for some  $(u, v) \in E$ , at least one of the following cases holds:*

- 1) *There exist  $S \subseteq \mathcal{X}_u$  and  $T \subseteq \mathcal{X}_v$  of densities at least  $2^{-O_{m,\varepsilon}(\log^2(2/\alpha))}$  such that*

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S \times T} [A_{uv}(\mathbf{x}, \mathbf{y})] \geq (1 + \delta) \alpha_{uv}$$

- 2) *There exists  $S \subseteq \mathcal{X}_u$  of density at least  $2^{-O_{m,\varepsilon}(\log(2/\alpha))}$  such that*

$$\mathbb{E}_{\mathbf{x} \in \mathcal{X}_v} [A_{uv}(z, \mathbf{x})] \leq (1 - \delta) \alpha_{uv} \quad \forall z \in S$$

*Proof.* By applying Lemma II.8 to  $f_{uv} := A_{uv}$  and using the monotonicity of the grid norms, for some  $p \leq \frac{8m}{\varepsilon} \log \frac{2}{\alpha}$ ,  $\delta = \varepsilon^2 2^{-4m-10}$  and  $(u, v) \in E$ , at least one of the following holds:

(i)

$$\|A_{uv}\|_{U(2k,p)} \geq (1 + \delta) \alpha_{uv};$$

(ii)

$$\|\mathbb{E}_{\mathbf{y}} [A_{uv}(\cdot, \mathbf{y})] - \alpha_{uv}\|_p \geq \delta \alpha_{uv}.$$

**Case (i):** Note that by Remark 4,

$$\|A_{uv}\|_{U(2k,p)} = t_{K_{2k,p}}^b(A_{uv}).$$

In this case, we can apply Corollary II.3 with  $H = K_{2k,p}$ , which implies the existence of  $S \times T \subseteq \mathcal{X}_u \times \mathcal{X}_v$  with

$$\frac{|S||T|}{|\mathcal{X}_u||\mathcal{X}_v|} \geq \Omega(\delta \alpha^{2pk+1}) \geq 2^{-O\left(\frac{m^2}{\varepsilon^2} \log^2(2/\alpha)\right)},$$

that satisfies Item 1 of Theorem II.1.

**Case (ii):** By applying Lemma II.7 to the function  $f: \mathcal{X}_u \rightarrow [0, 1]$  given by  $f(x) := \mathbb{E}_{\mathbf{y}} [A_{uv}(x, \mathbf{y})]$ , we conclude that one of the following two cases holds:

$$\text{a) } \mathbb{E} \left[ \mathbf{1}_{[f(x) > \alpha_{uv}(1 + \frac{\delta}{4})]} \right] \geq \frac{(\delta \alpha_{uv})^p}{4} = 2^{-O\left(\frac{m^2}{\varepsilon^2} \log^2(2/\alpha)\right)}.$$

$$\text{b) } \mathbb{E} \left[ \mathbf{1}_{[f(x) < \alpha_{uv}(1 - \frac{\delta}{4})]} \right] = \delta^p / 4 \geq \alpha^{O(m^2/\varepsilon^2)}.$$

In case (a), let  $S := \{x \in \mathcal{X}_u : \mathbb{E}_{\mathbf{y}} [A_{uv}(x, \mathbf{y})] > \alpha_{uv}(1 + \delta/4)\}$  and  $T = \mathcal{X}_v$  and conclude Item 1 of Theorem II.1:

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S \times T} A_{uv}(\mathbf{x}, \mathbf{y}) \geq (1 + \delta/4) \alpha_{uv}.$$

In case (b), let  $S = \{x \in \mathcal{X}_u : \mathbb{E}_{\mathbf{y}}[A_{uv}(x, \mathbf{y})] < \alpha_{uv}(1 - \delta/4)\}$  and conclude Item 2 of Theorem II.1:

$$\mathbb{E}_{\mathbf{x} \in \mathcal{X}_v}[A_{uv}(z, \mathbf{x})] \leq (1 - \delta/4)\alpha_{uv} \quad \text{for all } z \in S. \quad \square$$

### III. KELLEY–MEKA-TYPE BOUNDS FOR BINARY SYSTEMS

In this section, we establish Kelley–Meka-type bounds for the density of sets that do not contain translations of non-degenerate instances of binary systems of linear forms.

**Theorem I.5** (Kelley–Meka-type bounds over finite abelian groups). *Let  $\mathcal{L}$  be a binary system of linear forms over a finite abelian group  $G$ , and suppose  $|G|$  is coprime with all the coefficients in the linear forms. If  $A \subseteq G$  is free of translations of non-degenerate instances of  $\mathcal{L}$ , then*

- (i)  $|A| \leq |G| \cdot 2^{-\Omega_{\mathcal{L}}(\log^{1/16} |G|)}$ .
- (ii) *If, in addition, the underlying graph of  $\mathcal{L}$  is 2-degenerate, then*

$$|A| \leq |G| \cdot 2^{-\Omega_{\mathcal{L}}(\log^{1/9} |G|)}.$$

The rest of this section is devoted to the proof of Theorem I.5.

#### A. Preliminaries

This section discusses the necessary notation and terminology from additive number theory.

*a) Set addition.:* Let  $G$  be a finite abelian group. For an integer  $k \geq 0$  and  $x \in G$ , let

$$kx = \underbrace{x + x + \cdots + x}_{k \text{ times}} \quad \text{and} \quad (-k)x = -kx.$$

We denote the sumset of two sets  $A, B \subseteq G$  as

$$A + B := \{a + b : a \in A, b \in B\}.$$

Given a natural number  $k \in \mathbb{N}$ , we denote the iterated sumset of  $A$  as

$$kA := \{a_1 + \cdots + a_k : a_1, \dots, a_k \in A\},$$

and we let

$$k \cdot A := \{ka : a \in A\}.$$

*b) Bohr sets.:* We start with some standard notation and some basic facts about Bohr sets. Let  $G$  be a finite abelian group and let  $\widehat{G}$  denote the character group of  $G$ .

Given a set  $B \subseteq G$ , we define  $\mu_B$  to be the uniform probability measure on  $B$ . The relative density of  $A \subseteq G$  in  $B$  is denoted by  $\mu_B(A) = \frac{|A \cap B|}{|B|}$ . The normalized indicator function of  $B$  is defined as  $\varphi_B(\cdot) = \frac{|G|}{|B|}B(\cdot)$  so that  $\|\varphi_B\|_1 = 1$ .

Given two functions  $f, g: G \rightarrow \mathbb{R}$ , define the convolution and the cross-correlation as

$f * g(x) = \mathbb{E}_{\mathbf{y}} f(\mathbf{y})g(x - \mathbf{y})$  and  $f \star g(x) = \mathbb{E}_{\mathbf{y}} f(\mathbf{y})g(x + \mathbf{y})$ , respectively. Note that

$$\langle f * g, h \rangle = \langle f, g \star h \rangle.$$

We recall the definition of a Bohr set.

**Definition 5** (Bohr sets). For a non-empty  $\Gamma \subseteq \widehat{G}$  and  $\tau \in [0, 2]$ , define the corresponding *Bohr set*  $B = \text{Bohr}(\Gamma, \tau)$  as

$$\text{Bohr}(\Gamma, \tau) := \{x \in G : |\chi(x) - 1| \leq \tau \quad \forall \chi \in \Gamma\}.$$

We call  $\Gamma$  the *frequency set* of  $B$ , and  $\tau$  the *width*. We define the *rank* of  $B$ , denoted by  $\text{rk}(B)$ , to be the size of  $\Gamma$ .

When we speak of a Bohr set, we implicitly refer to the triple  $(\Gamma, \tau, \text{Bohr}(\Gamma, \tau))$ , since the set  $\text{Bohr}(\Gamma, \tau)$  alone does not uniquely determine the frequency set nor the width.

We note some easy facts about Bohr sets. We have

$$\text{Bohr}(\Gamma, \tau) \cap \text{Bohr}(\Gamma', \tau) = \text{Bohr}(\Gamma \cup \Gamma', \tau), \quad (18)$$

and

$$\text{Bohr}(\Gamma, \tau) + \text{Bohr}(\Gamma, \tau') \subseteq \text{Bohr}(\Gamma, \tau + \tau').$$

In particular, for any  $\lambda \in \mathbb{N}$ , we have the following inclusion:

$$\lambda \cdot \text{Bohr}(\Gamma, \tau) \subseteq \lambda \text{Bohr}(\Gamma, \tau) \subseteq \text{Bohr}(\Gamma, \lambda\tau).$$

Since  $\chi(\lambda x) = \chi(x)^\lambda$ , we also have

$$\lambda \cdot \text{Bohr}(\Gamma, \tau) = \text{Bohr}(\lambda \cdot \Gamma, \tau). \quad (19)$$

**Definition 6** (Dilate of a Bohr set). If  $B = \text{Bohr}(\Gamma, \tau)$  is a Bohr set, and  $\rho \geq 0$ , then  $B_\rho = \text{Bohr}(\Gamma, \rho\tau)$  is the *dilate* of  $B$  with parameter  $\rho$ .

While Bohr sets, in general, lack approximate group-like properties, Bourgain [16] observed that certain Bohr sets are approximately closed under addition in a certain weak sense.

**Definition 7** (Regularity). A Bohr set  $B$  of rank  $d$  is *regular* if for all  $\kappa$  with  $|\kappa| \leq \frac{1}{100d}$ , we have

$$(1 - 100d|\kappa|)|B| \leq |B_{1+\kappa}| \leq (1 + 100d|\kappa|)|B|.$$

Note that by (19), if  $(\lambda, |G|) = 1$  and  $\text{Bohr}(\Gamma, \tau)$  is a regular Bohr set of rank  $d$ , then  $\lambda \cdot \text{Bohr}(\Gamma, \tau)$  is also a regular Bohr set of rank  $d$  and has the same size as  $\text{Bohr}(\Gamma, \tau)$ . We will use this fact frequently in the proofs.

The following two lemmas show that a low-rank Bohr set  $B$  can be made into a regular Bohr set  $B_\rho$  without decreasing its size by much.

**Lemma III.1** ([16]; See [31, Lemma 4.25]). *For any Bohr set  $B$ , there exists  $\rho \in [\frac{1}{2}, 1]$  such that  $B_\rho$  is regular.*

**Lemma III.2** ([31, Lemma 4.20]). *If  $\rho \in (0, 1)$  and  $B$  is a Bohr set of rank  $d$  then  $|B_\rho| \geq \left(\frac{\rho}{4}\right)^d |B|$ .*

One of the nuances of working with Bohr sets is that even if  $B$  is a regular Bohr set (e.g., a large interval centred at 0), then for randomly chosen  $\mathbf{x}, \mathbf{x}' \in B$ , the distribution of  $\mathbf{x} + \mathbf{x}'$  is very different from the uniform distribution on  $B$ . A common strategy to address this issue is to work with two Bohr sets. We will have an additional regular Bohr set  $B' = B_\rho$  for some sufficiently small  $\rho$  (e.g., a smaller interval centred at 0). Then, for every  $y_0 \in B'$ , when  $\mathbf{x} \in B$  is chosen randomly and uniformly, the distribution of  $\mathbf{x} + y_0$  is close to the uniform distribution on  $B$ .

**Lemma III.3** ([32, Lemma 4.5]). *If  $B$  is a regular Bohr set of rank  $d$  and  $x \in B_\rho$  with  $\rho \in (0, 1)$ , then*

$$\|\varphi_B - \varphi_{B+x}\|_1 \leq 200\rho d.$$

*Proof.* The assertion of the lemma is equivalent to  $|B\Delta(B+x)| \leq 200\rho d|B|$ . Note  $B+x \subseteq B_{1+\rho}$ . If  $\rho \geq \frac{1}{100d}$ , the lemma is obvious. Otherwise, by the regularity of  $B$ , we have  $|B_{1+\rho}| \leq (1+100d\rho)|B|$ , which implies the lemma.  $\square$

We also need the following generalization of [18, Lemma 12.1].

**Lemma III.4.** *There exists a constant  $c > 0$  such that the following is true. Let  $B$  be a regular Bohr set of rank  $d$ , and  $A \subseteq B$  with  $\mu_B(A) = \alpha$ . Let  $\gamma > 0$ , and suppose  $\mathcal{B}$  is a finite family of Bohr sets where each  $B' \in \mathcal{B}$  is a subset of  $B_\rho$ , where  $\rho \leq \frac{c\alpha\gamma}{d|\mathcal{B}|}$ . Then, there exists  $x \in G$  such that one of the following two cases holds for the shift  $A' = A + x$ .*

- 1) *For every  $B' \in \mathcal{B}$ , we have  $|\mu_{B'}(A') - \alpha| \leq \gamma\alpha$ .*
- 2) *There exists  $B' \in \mathcal{B}$  such that  $\mu_{B'}(A') \geq \left(1 + \frac{\gamma}{2|\mathcal{B}|}\right)\alpha$ .*

*Proof.* If  $c > 0$  is small enough, then by Lemma III.3, for every  $B' \in \mathcal{B}$ ,

$$\begin{aligned} |\langle A * \varphi_{B'}, \varphi_B \rangle - \langle A, \varphi_B \rangle| &= \\ |\langle A, \varphi_{B'} * \varphi_B \rangle - \langle A, \varphi_B \rangle| &\leq \\ \|\varphi_{B'} * \varphi_B - \varphi_B\|_1 &\leq \frac{\alpha\gamma}{2|\mathcal{B}|}. \end{aligned}$$

In particular, since  $\langle A, \varphi_B \rangle = \alpha$ , we have

$$\langle A * \varphi_{B'}, \varphi_B \rangle \geq \left(1 - \frac{\gamma}{2|\mathcal{B}|}\right)\alpha,$$

which, by averaging over all  $B' \in \mathcal{B}$ , implies

$$\begin{aligned} \left(1 - \frac{\gamma}{2|\mathcal{B}|}\right)\alpha &\leq \mathbb{E}_{\mathbf{x} \in B} \mathbb{E}_{B' \in \mathcal{B}} A * \varphi_{B'}(\mathbf{x}) = \\ \mathbb{E}_{\mathbf{x} \in B} \mathbb{E}_{B' \in \mathcal{B}} \mu_{B'}(\mathbf{x} - A) &= \\ \mathbb{E}_{\mathbf{x} \in B} \mathbb{E}_{B' \in \mathcal{B}} \mu_{B'}(A - \mathbf{x}). \end{aligned} \quad (20)$$

Here, the last equality follows from the fact that  $B' = -B'$  for any Bohr set  $B'$ . Let  $x \in B$  be such that

$$\mathbb{E}_{B' \in \mathcal{B}} \mu_{B'}(A - x) \geq \left(1 - \frac{\gamma}{2|\mathcal{B}|}\right)\alpha. \quad (21)$$

Suppose there is  $B'' \in \mathcal{B}$  such that  $\mu_{B''}(A - x) \leq (1 - \gamma)\alpha$ . Combining this with (21), implies that there is  $B' \in \mathcal{B}$  such that  $\mu_{B'}(A - x) \geq \left(1 + \frac{\gamma}{2|\mathcal{B}| - 2}\right)\alpha$  which implies item (2).  $\square$

We finally give the definition of an algebraically spread set with respect to Bohr sets.

**Definition 8** (Algebraic Spreadness for Bohr Sets). Let  $B \subseteq G$  be a regular Bohr set and  $A \subseteq B$  with  $\mu_B(A) = \alpha$ . We say that  $A$  is a  $(\delta, d', r)$ -algebraically spread subset of  $B$  if for every  $x_0 \in G$  and every regular Bohr set  $B'$  with  $\text{rk}(B') \leq \text{rk}(B) + d'$  and  $|B'| \geq 2^{-r}|B|$ , we have  $\mu_{B'}(A + x_0) \leq (1 + \delta)\alpha$ .

### B. Counting lemma over abelian groups

In this section, we prove an algebraic counting lemma that will provide the key density increment step in the proof of Theorem I.5. Roughly speaking, this result shows that an algebraically spread set must contain many instances of a given binary system of linear forms.

**Theorem III.5** (Algebraic counting lemma). *Let*

$$\mathcal{L} = \{L_1, \dots, L_m\}$$

*be a binary system of linear forms over variables  $x_1, \dots, x_k$ . For every  $\varepsilon > 0$ , there is a  $\delta = \delta(\varepsilon, \mathcal{L}) > 0$  such that the following is true.*

*Suppose  $G$  is a finite abelian group such that  $|G|$  is coprime with all the coefficients in  $\mathcal{L}$ . Let  $A \subseteq B$  have density  $\mu_B(A) \geq \alpha$  where  $B \subseteq G$  is a regular Bohr set of rank  $d = O(1/\alpha)$ . If  $A$  is  $(\delta, d', r)$ -algebraically spread, where  $r = O_\delta(d \log^2(2/\alpha) + \log^{14}(2/\alpha))$  and  $d' = O_\delta(\log^{12}(2/\alpha))$  are sufficiently large, then there exist Bohr sets  $B^{(1)}, \dots, B^{(k)}$  and a shift  $A' = A + x_0$  such that the following statements hold.*

- *For every  $i \in [k]$ , we have*

$$|B^{(i)}| \geq 2^{-O_{\mathcal{L}, \varepsilon}(d \log^2(2/\alpha))}|B|;$$

- *We have*

$$\mathbb{E}_{\mathbf{x}_1 \in B^{(1)}, \dots, \mathbf{x}_k \in B^{(k)}} \prod_{i=1}^m A'(L_i(\mathbf{x}_1, \dots, \mathbf{x}_k)) \geq (1 - \varepsilon)\alpha^m. \quad (22)$$

To prove Theorem III.5, we will need an almost-periodicity theorem from [20]. The following theorem is stated with  $U_1 = U_2$  in [20], but the same proof works for arbitrary  $U_1$  and  $U_2$ .

**Theorem III.6** (Almost Periodicity [20, Lemma 8]). *Let  $\varepsilon \in (0, \frac{1}{10})$ , and let  $B, B', B'' \subseteq G$  be regular Bohr sets of rank  $d$ . Suppose that  $U_1, U_2 \subseteq B$ ,  $A_1 \subseteq B'$ , and  $A_2 \subseteq B''$  with densities  $\mu_B(U_1), \mu_B(U_2) \geq \tau$ ,  $\mu_{B'}(A_1) = \alpha_1$  and  $\mu_{B''}(A_2) = \alpha_2$ . Let  $\Gamma \subseteq G$  be a set with  $|\Gamma| \leq 2|B''|$  such that*

- 1)  $\mathbb{E}_{\mathbf{x}_1 \in A_1, \mathbf{x}_2 \in A_2} \Gamma(\mathbf{x}_1 - \mathbf{x}_2) \geq 1 - \varepsilon$
- 2) *For all  $x \in \Gamma$ ,*

$$\mathbb{E}_{\mathbf{z} \in B} U_1(x + \mathbf{z}) U_2(\mathbf{z}) \geq (1 + 2\varepsilon) \mu_B(U_1) \mu_B(U_2)$$

*Let  $C = \log(2/\tau) + \log(d) + \log \log(2/\alpha_1) + \log \log(2/\alpha_2)$ . There exists  $x_0 \in G$  and a regular Bohr set  $B^* \subseteq B'$  with*

$$\text{rk}(B^*) \leq d + O_\varepsilon(\log^2(2/\tau) \log(2/\alpha_1) \log(2/\alpha_2))$$

*and*

$$|B^*| \geq 2^{-O_\varepsilon(r)}|B'|$$

*where*

$$r = C \cdot (d + \log^2(2/\tau) \log(2/\alpha_1) \log(2/\alpha_2))$$

*such that*

$$\mu_{B^*}(U_1 + x_0) \geq \left(1 + \frac{\varepsilon}{4}\right) \mu_B(U_1).$$

We will also need the following theorem, whose proof uses one of the key ideas in the work of Kelley and Meka [19], namely, the dependent random choice technique.

**Theorem III.7** (Dependent Random Choice). *Let  $G$  be a finite abelian group, let  $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y} \subseteq G$ , and  $\tau > 0$ . Suppose  $A \subseteq \mathcal{X}_1$  and  $B \subseteq \mathcal{X}_2$  satisfy*

$$\mathbb{E}_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{X}_1 \times \mathcal{X}_2} (\mathbb{E}_{\mathbf{y} \in \mathcal{Y}} A(\mathbf{x}_1 + \mathbf{y}) B(\mathbf{x}_2 + \mathbf{y}))^p \geq (1 + 2\varepsilon)^p \tau^p, \quad (23)$$

where  $\varepsilon > 0$  and  $p \geq \Omega(\log(2/\varepsilon)/\varepsilon)$ . Then there exist  $A_1 \subseteq \mathcal{X}_1$  and  $A_2 \subseteq \mathcal{X}_2$  with  $\mu_{\mathcal{X}_1}(A_1) \mu_{\mathcal{X}_2}(A_2) \geq \tau^p$  such that

$$\Pr_{(\mathbf{x}_1, \mathbf{x}_2) \in A_1 \times A_2} [\mathbb{E}_{\mathbf{y} \in \mathcal{Y}} A(\mathbf{x}_1 + \mathbf{y}) B(\mathbf{x}_2 + \mathbf{y}) > (1 + \varepsilon) \tau] \geq 1 - \frac{\varepsilon}{100}. \quad (24)$$

*Proof.* Our assumption says

$$\mathbb{E}_{\substack{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \\ \mathbf{y}_1, \dots, \mathbf{y}_p \in \mathcal{Y}}} \prod_{j=1}^p A(\mathbf{x}_1 + \mathbf{y}_j) B(\mathbf{x}_2 + \mathbf{y}_j) \geq (1 + 2\varepsilon)^p \tau^p.$$

Denote  $\vec{x} = (x_1, x_2)$  and  $\vec{y} = (y_1, \dots, y_p)$ , and let

$$\mathcal{F}(\vec{x}, \vec{y}) := \prod_{j=1}^p A(x_1 + y_j) B(x_2 + y_j).$$

Define

$$A'_{\vec{y}} := \{x \in \mathcal{X}_1 : x + y_1, \dots, x + y_k \in A\} = \bigcap_{i=1}^k (A - y_i),$$

and denote  $\alpha_{\vec{y}} := \mu_{\mathcal{X}_1}(A'_{\vec{y}})$ . Similarly, define

$$B'_{\vec{y}} := \{x \in \mathcal{X}_2 : x + y_1, \dots, x + y_k \in B\} = \bigcap_{i=1}^k (B - y_i),$$

and denote  $\beta_{\vec{y}} := \mu_{\mathcal{X}_2}(B'_{\vec{y}})$ . We have

$$\begin{aligned} (1 + 2\varepsilon)^p \tau^p &\leq \mathbb{E}_{\vec{x}, \vec{y}} \mathcal{F}(\vec{x}, \vec{y}) \\ &= \mathbb{E}_{\vec{y}} \left[ \left( \mathbb{E}_{\mathbf{x}_1 \in \mathcal{X}_1} \prod_{j=1}^p A(\mathbf{x}_1 + \mathbf{y}_j) \right) \right. \\ &\quad \left. \cdot \left( \mathbb{E}_{\mathbf{x}_2 \in \mathcal{X}_2} \prod_{j=1}^p B(\mathbf{x}_2 + \mathbf{y}_j) \right) \right] \\ &= \mathbb{E}_{\vec{y}} [\alpha_{\vec{y}} \beta_{\vec{y}}]. \end{aligned}$$

We will only be interested in those  $\vec{y}$  where  $A'_{\vec{y}}$  and  $B'_{\vec{y}}$  are not very small. Let

$$\Lambda = \{\vec{y} \in \mathcal{Y}^p : \alpha_{\vec{y}} \beta_{\vec{y}} > \tau^p\}.$$

We have

$$\begin{aligned} \mathbb{E}_{\vec{y}} [\Lambda(\vec{y}) \mathbb{E}_{\vec{x}} [\mathcal{F}(\vec{x}, \vec{y})]] &= \mathbb{E}_{\vec{y}} \Lambda(\vec{y}) \alpha_{\vec{y}} \beta_{\vec{y}} \geq \\ (1 + 2\varepsilon)^p \tau^p - \tau^p &= ((1 + 2\varepsilon)^p - 1) \tau^p. \end{aligned} \quad (25)$$

Let  $S \subseteq \mathcal{X}_1 \times \mathcal{X}_2$  defined such that  $(x_1, x_2) \in S$  iff  $\mathbb{E}_{\mathbf{y} \in \mathcal{Y}} A(\mathbf{x}_1 + \mathbf{y}) B(\mathbf{x}_2 + \mathbf{y}) \leq (1 + \varepsilon) \tau$ . Note that

$$\begin{aligned} \mathbb{E}_{\vec{x}, \vec{y}} \mathcal{F}(\vec{x}, \vec{y}) &= \\ \mathbb{E}_{(\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{X}_1 \times \mathcal{X}_2} (\mathbb{E}_{\mathbf{y} \in \mathcal{Y}} A(\mathbf{x}_1 + \mathbf{y}) B(\mathbf{x}_2 + \mathbf{y}))^p, \end{aligned}$$

and

$$\mathbb{E}_{\vec{x}, \vec{y}} \mathcal{F}(\vec{x}, \vec{y}) S(\vec{x}) \leq (1 + \varepsilon)^p \tau^p. \quad (26)$$

Putting (25) and (26) together gives

$$\frac{\mathbb{E}_{\vec{y}} \mathbb{E}_{\vec{x}} \mathcal{F}(\vec{x}, \vec{y}) S(\vec{x})}{\mathbb{E}_{\vec{y}} \Lambda(\vec{y}) \mathbb{E}_{\vec{x}} \mathcal{F}(\vec{x}, \vec{y})} \leq \frac{(1 + \varepsilon)^p}{(1 + 2\varepsilon)^p - 1}.$$

In particular, for  $p = \Omega(\log(2/\varepsilon)/\varepsilon)$ , there exists a choice of  $\vec{y} \in \Lambda$  such that

$$\frac{\mathbb{E}_{\vec{x}} \mathcal{F}(\vec{x}, \vec{y}) S(\vec{x})}{\mathbb{E}_{\vec{x}} \mathcal{F}(\vec{x}, \vec{y})} \leq \frac{(1 + \varepsilon)^p}{(1 + 2\varepsilon)^p - 1} \leq \frac{\varepsilon}{100}.$$

Now note that  $\mathcal{F}(\vec{x}, \vec{y}) = 1$  iff  $(x_1, x_2) \in A'_{\vec{y}} \times B'_{\vec{y}}$ , and thus

$$\Pr_{(\mathbf{x}_1, \mathbf{x}_2) \in A'_{\vec{y}} \times B'_{\vec{y}}} [(\mathbf{x}_1, \mathbf{x}_2) \in S] = \frac{\mathbb{E}_{\vec{x}} \mathcal{F}(\vec{x}, \vec{y}) S(\vec{x})}{\mathbb{E}_{\vec{x}} \mathcal{F}(\vec{x}, \vec{y})} \leq \frac{\varepsilon}{100}.$$

We set  $A_1 := A'_{\vec{y}}$  and  $A_2 := B'_{\vec{y}}$ . Since  $\vec{y} \in \Lambda$ , we have  $\mu_{\mathcal{X}_1}(A_1) \mu_{\mathcal{X}_2}(A_2) \geq \tau^p$ , as desired.  $\square$

*1) Proof of Theorem III.5:* Suppose the underlying graph of  $\mathcal{L}$  is an oriented graph  $H = ([k], E)$  such that  $(i, j) \in E$  implies  $i < j$ . For convenience, we re-index  $\mathcal{L} = \{L_{ij} : (i, j) \in E\}$  where  $L_{ij}(x_1, \dots, x_k) = \lambda_{ij} x_i + \eta_{ij} x_j$  with  $i < j$ . By our assumption, all the coefficients are coprime with  $|G|$ . Let  $\gamma = \Omega_{m, \varepsilon}(1)$  be a small constant to be determined. We emphasize that  $\gamma$  does not depend on the particular choice of the coefficients in  $\mathcal{L}$ . Let

$$K := \prod_{(i, j) \in E} |\eta_{ij}| |\lambda_{ij}| \quad \text{and} \quad \rho := \frac{\gamma \exp(-\gamma^{-1} \log^2(2/\alpha))}{10^3 d K^{k+1}}. \quad (27)$$

By Lemma III.1, there exist  $\frac{\rho^k}{2^k K^{k^2}} \leq \rho_1 \leq \dots \leq \rho_k \leq \frac{\rho}{K}$  such that the following properties hold.

- (P1)  $B_{\rho_i}$ 's are regular for all  $i$ ;
- (P2)  $\frac{\rho \rho_{i+1}}{2^k K^k} \leq \rho_i \leq \frac{\rho \rho_{i+1}}{K^k}$  for all  $i = 1, \dots, k-1$ . In particular,  $\frac{\rho}{2^k K^{k^2}} \leq \rho_1$ .

Let  $B^{(i)} = K^{k-i} \cdot B_{\rho_i}$  for all  $i$ . Note that the conditions on  $\rho$  and  $\rho_i$  guarantee the following.

- (P3) for all  $(i, j) \in E$ , we have

$$\lambda_{ij} \cdot B^{(i)} = \lambda_{ij} K^{k-i} \cdot B_{\rho_i} \subseteq B_{\rho_k K} \subseteq B_\rho \text{ and}$$

$$\eta_{ij} \cdot B^{(j)} = \eta_{ij} K^{k-j} \cdot B_{\rho_j} \subseteq B_{\rho_k K} \subseteq B_\rho;$$

- (P4) For every  $(i, j) \in E$ , since  $i < j$  and  $K$  is divisible by  $\eta_{ij}$ , we have  $\frac{\lambda_{ij} K^{j-i}}{\eta_{ij}} \cdot B_{\rho_i} \subseteq B_{\rho \cdot \rho_j}$ . In particular, for every  $(i, j) \in E$ , we have

$$\lambda_{ij} \cdot B^{(i)} \subseteq B'_\rho \quad \text{with} \quad B' := \eta_{ij} \cdot B^{(j)};$$

**(P5)** By Lemma III.2, for all  $i$ , we have

$$|B^{(i)}| \geq (\rho_i/4)^d |B| \geq \left( \frac{\rho^k}{4 \times 2^k K^{k^2}} \right)^d |B| \geq 2^{-O_{\mathcal{L}, \varepsilon}(d \log^2(2/\alpha))} |B|,$$

where the last inequality uses the assumption  $d = O(1/\alpha)$ .

Consider the collection  $\mathcal{B} = \{\lambda_{ij} \cdot B^{(i)} : (i, j) \in E\} \cup \{\eta_{ij} \cdot B^{(j)} : (i, j) \in E\}$  of Bohr sets and observe that  $|\mathcal{B}| \leq 2m$ . Note that by **(P3)**, all these Bohr sets are subsets of  $B_\rho$ , and moreover, if  $\gamma \leq O(1/m)$  is sufficiently small, we can ensure  $\rho < \frac{c\alpha\gamma}{d|\mathcal{B}|}$ , where  $c$  is the constant from Lemma III.4. Therefore, we can apply Lemma III.4 with  $A, \mathcal{B}, d, \gamma, \alpha_* = \mu_B(A) \geq \alpha$  to guarantee that there is a shift  $A' = A + x_0$  such that one of the following two cases holds:

**(P6)** For every  $B' \in \mathcal{B}$ ,

$$|\mu_{B'}(A') - \alpha_*| \leq \gamma\alpha_*; \quad (28)$$

**(P6')** There is  $B' \in \mathcal{B}$  such that  $\mu_{B'}(A') \geq \left(1 + \frac{\gamma}{2|B|}\right) \alpha_* \geq \left(1 + \frac{\gamma}{4m}\right) \alpha_*$ .

Since  $A$  is  $(\delta, d', r)$ -algebraically spread, assuming  $\delta < \frac{\gamma}{8m}$  and  $r = \Omega_{m, \varepsilon}(d \log^2(2/\alpha))$  is sufficiently small, **(P6')** cannot hold, and therefore, (28) holds. Assume to the contrary of the theorem that (22) does not hold. In particular,

$$\left| \mathbb{E}_{\mathbf{x}_1 \in B^{(1)}, \dots, \mathbf{x}_k \in B^{(k)}} \prod_{i=1}^m A'(L_i(\mathbf{x}_1, \dots, \mathbf{x}_k)) - \alpha_*^m \right| > \varepsilon \alpha_*^m. \quad (29)$$

We will show that  $A$  is not  $(\delta, d', r)$ -algebraically spread, reaching a contradiction. We divide the proof into two stages. In the first part, we find some large subsets  $S, T$  such that

$$\mathbb{E}_{\mathbf{x} \in S, \mathbf{y} \in T} A'(\mathbf{x} + \mathbf{y}) \geq (1 + 100\delta)\alpha_*.$$

In the second stage, we obtain a Bohr set  $B^*$  of rank  $d + d'$  and relative density at least  $2^{-r}$  such that

$$\mu_{B^*}(A') \geq (1 + \delta)\alpha_*$$

which gives the desired contradiction.

a) *Stage 1: Density increment on a rectangle.*: For every  $(i, j) \in E$ , define  $G_{ij} \subseteq B^{(i)} \times B^{(j)}$  as  $G_{ij}(x, y) := A'(\lambda_{ij}x + \eta_{ij}y)$  and let  $\alpha_{ij}$  denote the density of  $G_{ij}$  in  $B^{(i)} \times B^{(j)}$ .

**Claim III.8.** For any  $(i, j) \in E$  and every  $x \in B^{(i)}$ , we have

$$|\mathbb{E}_{\mathbf{y} \in B^{(j)}} G_{ij}(x, \mathbf{y}) - \alpha_*| \leq 2\gamma\alpha_*. \quad (30)$$

In particular,

$$|\alpha_{ij} - \alpha_*| \leq 2\gamma\alpha_*,$$

and

$$\left| \prod_{(i, j) \in E} \alpha_{ij} - \alpha_*^m \right| = O(m\gamma\alpha_*^m). \quad (31)$$

*Proof.* Denote  $B' = \eta_{ij} \cdot B^{(j)}$  and recall that by **(P4)**,  $\lambda_{ij}x \in \lambda_{ij} \cdot B^{(i)} \subseteq B'$ . Therefore, by Lemma III.3 and the choice of  $\rho$ , we have

$$\|\varphi_{B'+\lambda_{ij}x} - \varphi_{B'}\|_1 \leq 200\rho d \leq \gamma\alpha.$$

In particular,

$$\begin{aligned} |\mathbb{E}_{\mathbf{y} \in B'} A'(\lambda_{ij}x + \mathbf{y}) - \mu_{B'}(A')| &= \\ |\langle \varphi_{B'+\lambda_{ij}x} - \varphi_{B'}, A' \rangle| &\leq \gamma\alpha. \end{aligned}$$

By (28), we have  $|\mu_{B'}(A') - \alpha_*| \leq \gamma\alpha_*$ , and therefore, we can conclude

$$\begin{aligned} |\mathbb{E}_{\mathbf{y} \in B^{(j)}} G_{ij}(x, \mathbf{y}) - \alpha_*| &= \\ |\mathbb{E}_{\mathbf{y} \in B'} A'(\lambda_{ij}x + \mathbf{y}) - \alpha_*| &\leq 2\gamma\alpha_*, \end{aligned}$$

which verifies (30). Averaging this inequality over  $x \in B^{(i)}$  yields  $|\alpha_{ij} - \alpha_*| \leq 2\gamma\alpha_*$ . Finally, (31) follows from

$$(1 - 2\gamma)^m \alpha_*^m \leq \prod_{(i, j) \in E} \alpha_{ij} \leq (1 + 2\gamma)^m \alpha_*^m. \quad \square$$

By (29) and (31), if  $\gamma = O(\varepsilon/m)$  is sufficiently small, we get

$$\begin{aligned} &\left| \mathbb{E}_{\mathbf{x}_1 \in B^{(1)}, \dots, \mathbf{x}_k \in B^{(k)}} \prod_{(i, j) \in E} G_{ij}(\mathbf{x}_i, \mathbf{x}_j) - \prod_{(i, j) \in E} \alpha_{ij} \right| \\ &\geq \frac{\varepsilon}{2} \prod_{(i, j) \in E} \alpha_{ij}. \end{aligned}$$

Now, we are ready to apply our graph counting lemma (i.e., Theorem II.1). There exists a constant  $\delta' = \delta'(\varepsilon, m) > 0$  such that for some  $(i, j) \in E$ , at least one of the following cases holds.

(i) There exist  $S' \subseteq B^{(i)}$  and  $T' \subseteq B^{(j)}$  with  $|S'| |T'| \geq 2^{-O_{m, \varepsilon}(\log^2(2/\alpha))} |B^{(i)}| |B^{(j)}|$  such that

$$\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S' \times T'} [G_{ij}(\mathbf{x}, \mathbf{y})] \geq (1 + \delta')\alpha_{ij}; \quad (32)$$

(ii) There exists a set  $S \subseteq B^{(i)}$  of density at least  $2^{-O_{m, \varepsilon}(\log(2/\alpha))}$  such that

$$\mathbb{E}_{\mathbf{x} \in B^{(j)}} [G_{ij}(z, \mathbf{x})] \leq (1 - \delta')\alpha_{ij} \quad \forall z \in S;$$

If  $\gamma = O(\delta')$  is sufficiently small, by Claim III.8, for every  $z \in B^{(i)}$ , we have

$$|\mathbb{E}_{\mathbf{x} \in B^{(j)}} [G_{ij}(z, \mathbf{x})] - \alpha_{ij}| < \delta' \alpha_{ij},$$

and (ii) cannot hold. Therefore, (i) must be true. In particular, if  $\gamma \leq \delta'/8$ ,

$$\begin{aligned} \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \in S' \times T'} [G_{ij}(\mathbf{x}, \mathbf{y})] &\geq (1 + \delta')\alpha_{ij} \geq \\ (1 + \delta')(1 - 2\gamma)\alpha_* &\geq \left(1 + \frac{\delta'}{2}\right) \alpha_*, \quad (33) \end{aligned}$$

Denoting

$S := \lambda_{ij} \cdot S'$ ,  $T := \eta_{ij} \cdot T'$ ,  $B' := \eta_{ij} \cdot B^{(j)}$ ,  $B'' := \lambda_{ij} \cdot B^{(i)}$ , we have the following:

- $B', B'' \subseteq B$  are regular Bohr sets of rank  $d$  that by **(P5)** satisfy

$$|B'|, |B''| \geq 2^{-O_{\mathcal{L}, \varepsilon}(d \log^2(2/\alpha))} |B|.$$

Moreover, by **(P4)**, we have  $B'' \subseteq B'_\rho$  and  $B' \subseteq B_\rho$ .

- By **(P6)**, we have  $|\mu_{B'}(A') - \alpha_*| \leq \gamma \alpha_*$  and  $|\mu_{B''}(A') - \alpha_*| \leq \gamma \alpha_*$ .
- $T \subseteq B'$  and  $S \subseteq B''$  and they satisfy

$$\mathbb{E}_{\mathbf{s} \in S, \mathbf{t} \in T} A'(\mathbf{s} + \mathbf{t}) \geq \left(1 + \frac{\delta'}{2}\right) \alpha_*, \quad (34)$$

with  $\delta' = \Omega_{m, \varepsilon}(1)$  and

$$\mu_{B''}(S), \mu_{B'}(T) \geq 2^{-O_{m, \varepsilon}(\log^2(2/\alpha))}.$$

b) *Stage 2: Density increment on a Bohr set.*: We can rewrite (34) as

$$\begin{aligned} \mathbb{E}_{\mathbf{b}'' \in B'', \mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') S(\mathbf{b}'') T(\mathbf{b}') \geq \\ \left(1 + \frac{\delta'}{2}\right) \alpha_* \cdot \mu_{B''}(S) \cdot \mu_{B'}(T) \geq 2^{-O_{m, \varepsilon}(\log^2(2/\alpha))}. \end{aligned} \quad (35)$$

We wish to apply a change of variables and then apply Hölder's inequality to eliminate the set  $S$ . Since the averages are over Bohr sets rather than the entire group, we will incur some small errors as a result of the change of variables. By Lemma III.3 and our choice of  $\rho$  in (27), if  $\gamma = \Omega_{\varepsilon, m}(1)$  sufficiently small, for every  $z \in B''_\rho \subseteq B'_\rho$ ,

$$\begin{aligned} & |\mathbb{E}_{\mathbf{b}'' \in B'' - z} \mathbb{E}_{\mathbf{b}' \in B' + z} A'(\mathbf{b}'' + \mathbf{b}') S(\mathbf{b}'') T(\mathbf{b}') \\ & - \mathbb{E}_{\mathbf{b}'' \in B'', \mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') S(\mathbf{b}'') T(\mathbf{b}')| \\ & \leq 400\rho d \leq \gamma \alpha_* \cdot \mu_{B''}(S) \cdot \mu_{B'}(T). \end{aligned}$$

By averaging over  $z \in B''_\rho$  and combining it with (35), if  $\gamma \leq \delta'/8$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B'', \mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') S(\mathbf{b}'' - \mathbf{z}) T(\mathbf{z} + \mathbf{b}') \geq \\ \left(1 + \frac{\delta'}{4}\right) \alpha_* \cdot \mu_{B''}(S) \cdot \mu_{B'}(T). \end{aligned} \quad (36)$$

Next, we will apply Hölder's inequality to remove  $S$ . For any  $p \geq 2$  and  $q \in (1, 2]$  with  $\frac{1}{p} + \frac{1}{q} = 1$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B'', \mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') S(\mathbf{b}'' - \mathbf{z}) T(\mathbf{z} + \mathbf{b}') \leq \\ \left( \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B''} (\mathbb{E}_{\mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') T(\mathbf{z} + \mathbf{b}'))^p \right)^{1/p} \\ \cdot \left( \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B''} S(\mathbf{b}'' - \mathbf{z}) \right)^{1/q}. \end{aligned}$$

By Lemma III.3, we have

$$|\mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B''} S(\mathbf{b}'' - \mathbf{z}) - \mu_{B''}(S)| \leq 200d\rho \leq \gamma \mu_{B''}(S),$$

which with (36) gives us

$$\begin{aligned} & \left( \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B''} (\mathbb{E}_{\mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') T(\mathbf{z} + \mathbf{b}'))^p \right)^{1/p} \\ & \cdot (1 + \gamma)^{1/q} \geq \\ & \left(1 + \frac{\delta'}{4}\right) \alpha_* \cdot \mu_{B'}(T) \cdot \mu_{B''}(S)^{1 - \frac{1}{q}}. \end{aligned}$$

By (28), we have  $\alpha_* > (1 - \gamma)\mu_{B'}(A')$ . If  $\gamma \leq \delta'/100$ , we have

$$\begin{aligned} \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B''} (\mathbb{E}_{\mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') T(\mathbf{z} + \mathbf{b}'))^p \geq \\ \left(1 + \frac{\delta'}{8}\right)^p \mu_{B'}(A')^p \mu_{B'}(T)^p \cdot \mu_{B''}(S). \end{aligned}$$

Taking  $p = O(\log(2\mu_{B''}(S)^{-1})/\delta') = O(\log^2(2/\alpha))$  to be sufficiently large, we obtain that

$$\begin{aligned} \mathbb{E}_{\mathbf{z} \in B''_\rho, \mathbf{b}'' \in B''} (\mathbb{E}_{\mathbf{b}' \in B'} A'(\mathbf{b}'' + \mathbf{b}') T(\mathbf{z} + \mathbf{b}'))^p \geq \\ \left(1 + \frac{\delta'}{16}\right)^p \mu_{B'}(A')^p \mu_{B'}(T)^p. \end{aligned} \quad (37)$$

Now, we are in the position to apply Theorem III.7 with  $\mathcal{X}_1 = B''_\rho$ ,  $\mathcal{X}_2 = B''$ , and  $\mathcal{Y} = B'$ . Therefore, there exists sets  $A_1 \subseteq B''_\rho$  and  $A_2 \subseteq B''$ , and  $\delta'' = \frac{\delta'}{32}$ , such that

$$\begin{aligned} \mu_{B''_\rho}(A_1) \mu_{B''}(A_2) \geq \mu_{B'}(A')^p \mu_{B'}(T)^p = \\ 2^{-O_{\varepsilon, m}(p \log(2/\alpha) + p \log^2(2/\alpha))} = 2^{-O_{\varepsilon, m}(\log^4(2/\alpha))}. \end{aligned} \quad (38)$$

and

$$\begin{aligned} \Pr_{(\mathbf{x}_1, \mathbf{x}_2) \in A_1 \times A_2} [\mathbb{E}_{\mathbf{z} \in B'} A'(\mathbf{x}_1 + \mathbf{z}) T(\mathbf{x}_2 + \mathbf{z}) > \\ (1 + \delta'') \mu_{B'}(A') \mu_{B'}(T)] \geq \\ 1 - \frac{\delta''}{100}. \end{aligned} \quad (39)$$

Note that  $(x_1, x_2) \in A_1 \times A_2$  implies  $x_2 \in B'' \subseteq B'_\rho$  where the latter inclusion is by **(P4)**. Therefore, by Lemma III.3, assuming  $\gamma \leq \delta''/10$ , for every  $(x_1, x_2) \in A_1 \times A_2$ , we have

$$\begin{aligned} |\mathbb{E}_{\mathbf{z} \in B'} A'(x_1 + \mathbf{z}) T(x_2 + \mathbf{z}) - \mathbb{E}_{\mathbf{z} \in B'} A'(x_1 - x_2 + \mathbf{z}) T(\mathbf{z})| \leq \\ 200\rho d \leq \frac{\delta''}{10} \mu_{B'}(A') \mu_{B'}(T). \end{aligned} \quad (40)$$

Let  $\Gamma \subseteq B''$  be the set of all  $b \in B''$  such that

$$\mathbb{E}_{\mathbf{z} \in B'} A'(b + \mathbf{z}) T(\mathbf{z}) \geq (1 + 0.9\delta'') \mu_{B'}(A') \mu_{B'}(T).$$

By (39) and (40), we have

$$\mathbb{E}_{(\mathbf{x}_1, \mathbf{x}_2) \in A_1 \times A_2} [\mathbf{x}_1 - \mathbf{x}_2 \in \Gamma] \geq 1 - \frac{\delta''}{100}.$$

By Lemma III.1, we may assume that  $B''' := B''_\rho$  is regular. We invoke Theorem III.6 with  $2\varepsilon = 0.9\delta''$ ,  $\tau = 2^{-O_{\varepsilon, m}(\log^2(2/\alpha))}$ , and  $A', T \subseteq B', A_1 \subseteq B'''$  and  $A_2 \subseteq B''$ . Note that  $|\Gamma| \leq |B''|$ , and we have

- $\mu_{B'}(A') \geq (1 - \gamma)\alpha \geq \tau$ ;
- $\mu_{B'}(T) \geq 2^{-O_{\varepsilon, m}(\log^2(2/\alpha))} \geq \tau$ ;
- $\alpha_1 := \mu_{B'''}(A_1) \geq 2^{-O_{\varepsilon, m}(\log^4(2/\alpha))}$  and  $\alpha_2 := \mu_{B''}(A_2) \geq 2^{-O_{\varepsilon, m}(\log^4(2/\alpha))}$  by (38).

We obtain a regular Bohr set  $B^*$  and  $x \in G$  such that

$$\begin{aligned} \frac{|(A' + x) \cap B^*|}{|B^*|} \geq \left(1 + \frac{\delta''}{10}\right) \frac{|A'|}{|B'|} \geq \\ \left(1 + \frac{\delta''}{10}\right) (1 - \gamma) \frac{|A|}{|B|} \geq \left(1 + \frac{\delta''}{20}\right) \frac{|A|}{|B|}. \end{aligned}$$

The last inequality is by assuming  $\gamma \leq \delta''/40$ . Moreover,

$$\begin{aligned}\text{rk}(B^*) &\leq \text{rk}(B') + O_\varepsilon(\log^2(2\tau^{-1}) \log(2\alpha_1^{-1}) \log(2\alpha_2^{-1})) \\ &\leq \text{rk}(B') + O_\varepsilon(\log^{12}(2\alpha^{-1}))\end{aligned}$$

and

$$|B^*| \geq 2^{-O_{m,\varepsilon}(r)} |B'''|,$$

where

$$\begin{aligned}C &= \log(2/\tau) + \log(\text{rk}(B')) + \log \log(2/\alpha_1) + \log \log(2/\alpha_2) \\ &= O(\log(2/\tau)) = O(\log^2(2/\alpha))\end{aligned}$$

and

$$\begin{aligned}r &\leq C \cdot (\text{rk}(B') + \log^2(2\tau^{-1}) \log(2\alpha_1^{-1}) \log(2\alpha_2^{-1})) \\ &\leq C \cdot (\text{rk}(B') + \log^{12}(2\alpha^{-1})) \\ &\leq O(\log^2(2\alpha^{-1})d + \log^{14}(2\alpha^{-1})).\end{aligned}$$

Combining  $|B^*| \geq 2^{-O_{m,\varepsilon}(r)} |B'''|$  with

$$\begin{aligned}|B'''| &= |B''_\rho| \geq \left(\frac{\rho}{4}\right)^d |B''| \geq \\ &\exp(-O_{m,\varepsilon}(d \log^2(2\alpha^{-1}))) |B|,\end{aligned}$$

we have

$$|B^*| \geq \exp(-O_{m,\varepsilon}(d \log^2(2\alpha^{-1}) + \log^{14}(2\alpha^{-1}))) |B|.$$

By taking  $\gamma = \Omega_{m,\varepsilon}(1)$  to be a small enough constant, we can satisfy all the requirements on  $\gamma$ . Taking  $\delta = \min\left(\frac{\delta''}{20}, \frac{\gamma}{10m}\right)$  finishes the proof of Theorem III.5.

### C. Proof of Theorem I.5 (i)

Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  and  $\alpha = \mu(A)$ . Suppose  $d' = O_\mathcal{L}(\log^{12}(2/\alpha))$  and  $r = O_\mathcal{L}(\log^{15}(2/\alpha))$  are sufficiently large. Let  $\delta = \delta(\varepsilon, \mathcal{L}) > 0$  given by Theorem III.5 with  $\varepsilon = 1/2$ . We wish to obtain an algebraically spread set  $A_*$  to apply Theorem III.5. If  $A$  is  $(\delta, d', r)$ -algebraically spread then  $A_* = A$ , otherwise, from Definition 8 (where  $B = G$ ), we find a regular Bohr set  $B^{[1]} \subseteq G$  of rank  $d'$  and density at least  $2^{-r}$  and an  $x_1 \in G$  such that  $A_1 := (A + x_1) \cap B^{[1]}$  satisfies  $\mu_{B^{[1]}}(A_1) \geq (1 + \delta)\alpha$ . We can repeat this density increment process at most  $O_\delta(\log(2/\alpha))$  times and obtain a regular Bohr set  $B^*$  and  $x_* \in G$  such that  $A_* := (A + x_*) \cap B^*$  is  $(\delta, d', r)$ -algebraically spread in  $B^*$ . At this point, we have

$$\text{rk}(B^*) \leq O_\mathcal{L}(d' \log(2/\alpha)) = O_\mathcal{L}(\log^{13}(2/\alpha))$$

and for

$$\mu(B^*) \geq \exp(-O_\mathcal{L}(\log^{16}(2/\alpha))).$$

Note  $\alpha_* := \mu_{B^*}(A_*) \geq \alpha$ . Since  $r = O_\delta(\text{rk}(B^*) \log^2(2/\alpha) + \log^{14}(2/\alpha)) = O_\mathcal{L}(\log^{15}(2/\alpha))$  is sufficiently large, by Theorem III.5, there are Bohr sets  $B^{(1)}, \dots, B^{(k)}$  such that

- For every  $i \in [k]$ , we have

$$\mu(B^{(i)}) \geq 2^{-O_\mathcal{L}(\log^{15}(2/\alpha))} \mu(B^*) \geq 2^{-O_\mathcal{L}(\log^{16}(2/\alpha))}. \quad (41)$$

- We have

$$\Pr_{\mathbf{x}_1 \in B^{(1)}, \dots, \mathbf{x}_k \in B^{(k)}} [\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_k) \in A_*^m] \geq \alpha^m/2. \quad (42)$$

Note that  $x \mapsto \lambda_{ij}x$  and  $x \mapsto \eta_{ij}x$  are group automorphisms. Hence, given any two distinct linear forms  $L, L' \in \mathcal{L}$ , we have

$$\Pr_{\mathbf{x}_1 \in B^{(1)}, \dots, \mathbf{x}_k \in B^{(k)}} [L(\mathbf{x}_1, \dots, \mathbf{x}_k) = L'(\mathbf{x}_1, \dots, \mathbf{x}_k)] \leq \frac{1}{\min_i |B^{(i)}|}. \quad (43)$$

This implies that

$$\Pr_{\mathbf{x}_1 \in B^{(1)}, \dots, \mathbf{x}_k \in B^{(k)}} [\mathcal{L}(\mathbf{x}_1, \dots, \mathbf{x}_k) \text{ is degenerate}] \leq \frac{m^2}{\min_i |B^{(i)}|}. \quad (44)$$

Since we have assumed that  $A$  does not contain non-degenerate instances of  $\mathcal{L}$ , by combining (41), (42), and (44), we get

$$\frac{\alpha^m}{2} \leq \frac{m^2}{\min_i |B^{(i)}|} \leq m^2 \cdot |G|^{-1} \cdot \exp(O_\mathcal{L}(\log^{16}(2/\alpha))). \quad (45)$$

If  $\alpha \geq \exp(-O_\mathcal{L}(\log^{1/16}(|G|)))$ , we can ensure the right hand side in (45) is at most, say,  $|G|^{-1/2}$ , which violates (45) if  $|G| = \Omega_\mathcal{L}(1)$  is large enough. This is a contradiction.  $\square$

### D. Proof of Theorem I.5 (ii)

Recall that we want to leverage the assumption that the underlying graph of  $\mathcal{L}$  is 2-degenerate to conclude the stronger bound of

$$|A| \leq |G| \cdot 2^{-\Omega_\mathcal{L}(\log^{1/9} |G|)}.$$

We will describe how the proof of Theorem III.5 can be modified to yield a stronger density increment. We change the definition of  $\rho$  in (27) to the following, shaving off a factor of  $\log(2/\alpha)$ .

$$K := \prod_{(i,j) \in E} |\eta_{ij}| |\lambda_{ij}| \quad \text{and} \quad \rho := \frac{\gamma \exp(-\gamma^{-1} \log(2/\alpha))}{10^3 d K^{k+1}}. \quad (46)$$

Let  $H$  be the underlying graph of  $\mathcal{L}$ . Since  $H$  is 2-degenerate, there exists an ordering of the vertices such that every vertex has at most 2 preceding neighbours. We then orient each edge of  $H$  from the smaller to the larger vertex in this ordering. Note that the indegree of every vertex is at most 2. Therefore, if one applies Lemma II.8 with this graph  $H$ , one would obtain

$$\|G_{ij}\|_{U(2,p)} \geq (1 + \delta')\alpha_{ij},$$

for  $p = O_{m,\varepsilon}(\log \frac{2}{\alpha})$  and  $\delta' = \varepsilon^{2-4m-10}$ . Recalling the definitions of  $B'$  and  $B''$ , we can express this as,

$$\mathbb{E}_{\mathbf{x} \in B'', \mathbf{y} \in B''} \mathbb{E}_{\mathbf{z} \in B'} (A'(\mathbf{x} + \mathbf{z}) A'(\mathbf{y} + \mathbf{z}))^p \geq (1 + \delta')^{2p} \alpha_{ij}^{2p}. \quad (47)$$

Note that this allows us to bypass obtaining the large rectangle  $S \times T$  and directly jump to (37) with  $T = A'$ , where we also use  $\alpha_{ij} \geq (1 - O(\gamma))\mu_{B'}(A')$ . Furthermore, we are in a better situation in terms of the parameters: Firstly,  $p = O_{m,\varepsilon}(\log(2/\alpha))$  instead of  $p = O_{m,\varepsilon}(\log^2(2/\alpha))$ . Secondly, in (47), we have  $\mu_{B'}(A')^{2p} = 2^{-O(\log^2(2/\alpha))}$  instead of  $\mu_{B'}(A')^p \mu_{B'}(T)^p = 2^{-O_{m,\varepsilon}(p\log(2/\alpha) + p\log^2(2/\alpha))} = 2^{-O(\log^4(2/\alpha))}$  in (37). Note also that the reason we could choose a larger  $\rho$  is that in (40), we now only need  $200\rho d$  to be sufficiently small compared to  $\frac{\delta''}{10}\mu_{B'}(A')\mu_{B'}(A') = \Omega(\alpha^2)$ . Therefore, when we apply Theorem III.7, we obtain  $A_1, A_2 \subseteq B''$  with

$$\mu_{B''}(A_1), \mu_{B''}(A_2) \geq 2^{-O_{m,\varepsilon}(\log^2(2/\alpha))}.$$

Now, we invoke Theorem III.6 with  $U_1 = U_2 := A' \subseteq B'$  and  $A_1, A_2 \subseteq B''$ , and obtain  $B'''$  and  $x_0 \in G$  with  $(A + x_0) \cap B''' \geq (1 + \delta)\alpha$ , such that

$$\begin{aligned} \text{rk}(B''') &\leq \text{rk}(B') + O_{m,\varepsilon}(\log^2(2\alpha^{-1})\log(2\alpha_1^{-1})\log(2\alpha_2^{-1})) \\ &\leq \text{rk}(B') + O_{m,\varepsilon}(\log^6(2\alpha^{-1})) \end{aligned}$$

and

$$|B'''| \geq 2^{-O_{m,\varepsilon}(r)}|B''|,$$

where

$$\begin{aligned} C = \\ \log(2/\alpha) + \log(\text{rk}(B')) + \log\log(2/\alpha_1) + \log\log(2/\alpha_2) = \\ O(\log(2/\alpha)) \end{aligned}$$

and

$$\begin{aligned} r &\leq C \cdot (\text{rk}(B') + \log^2(2\alpha^{-1})\log(2\alpha_1^{-1})\log(2\alpha_2^{-1})) \\ &\leq C \cdot (\text{rk}(B') + \log^6(2\alpha^{-1})) \\ &\leq O(\log(2\alpha^{-1})d + \log^7(2\alpha^{-1})). \end{aligned}$$

By an iterative density increment argument, similar to Theorem III.5, we obtain a Bohr set  $B^*$  and  $x_* \in G$  such that  $A_* := (A + x_*) \cap B^*$  is  $(\delta', d', r)$ -algebraically spread in  $B^*$  where  $d' = O_m(\log^6(2/\alpha))$  and  $r = O_m(\log^8(2/\alpha))$ ,  $\delta' = \delta(\varepsilon, m)$ , and  $\alpha_* := \mu_{B^*}(A_*) \geq \alpha$ . We have

$$\text{rk}(B^*) \leq O_{m,\varepsilon}(\log^7(2/\alpha))$$

and

$$\mu(B^*) \geq \exp(-O_{m,\varepsilon}(\log^9(2/\alpha))).$$

A similar argument as in the proof of Theorem I.5 (i) concludes the proof.  $\square$

#### IV. CONCLUDING REMARKS: RESILIENCE TO SPREADNESS

In this section, we discuss negative results and argue that, essentially, the approach based on the graph and algebraic spreadness does not apply to most systems of linear forms. For simplicity, we focus on the case of the group  $\mathbb{F}^n$  where  $\mathbb{F} = \mathbb{F}_q$  is a finite field, and  $q$  is a fixed prime. The following notion of pseudo-randomness, based on graph spreadness, is implicit in the proof of Theorem I.5.

**Definition 9** (Spreadness [12], [19]). Let  $r \geq 1$  and  $\delta > 0$ . A set  $A \subseteq \mathbb{F}^n$  of density  $\alpha$  is called  $(\delta, r)$ -combinatorially spread if, for all subsets  $S$  and  $T$  of density at least  $2^{-r}$ , we have

$$\mathbb{E}_{\mathbf{x} \sim S}[\mathbb{E}_{\mathbf{y} \sim T}[A(\mathbf{x} + \mathbf{y})]] \leq (1 + \delta)\alpha.$$

In the literature, two main notions are commonly used to characterize the complexity of systems of linear forms. The first notion is called the *Cauchy-Schwarz complexity* (*CS-complexity* for short), defined by Green and Tao [33] in their study of solutions of linear equations in the primes. The second notion, defined by Gower and Wolf [34], [35], is called the *true complexity*. In both definitions, the complexity of a system is a positive integer. The following inclusions follow from the definitions of these complexity measures.

$$\begin{aligned} \{\mathcal{L} : \mathcal{L} \text{ is binary}\} &\subseteq \{\mathcal{L} : \mathcal{L} \text{ has CS-complexity 1}\} \subseteq \\ &\subseteq \{\mathcal{L} : \mathcal{L} \text{ has true complexity 1}\}. \end{aligned}$$

Our sparse graph counting lemma shows that for binary systems of linear forms, combinatorial spreadness implies  $t_{\mathcal{L}}(A) \approx \alpha^m$ . On the other hand, we will show in Theorem IV.3 that if the true complexity of a linear system  $\mathcal{L}$  is larger than 1, then there are sets  $A$  with density  $\alpha$  that are combinatorially spread with extremely good parameters, and yet  $t_{\mathcal{L}}(A) > (1 + \Omega(1))\alpha^m$ .

*a) True complexity and CS-complexity.*: Gowers, in his proof of Szemerédi's theorem, introduced a hierarchy of increasingly stronger notions of pseudo-randomness based on the so-called Gowers uniformity norms  $\|\cdot\|_{U^k}$  for  $k \geq 1$ . He used iterated applications of the classical Cauchy-Schwarz inequality to prove

$$|\mathbb{E}_{\mathbf{x}, \mathbf{y}}[A(\mathbf{x})A(\mathbf{x} + \mathbf{y}) \cdots A(\mathbf{x} + (k-1)\mathbf{y})] - \alpha^k| \leq k\|A - \alpha\|_{U^{k-1}},$$

showing that the density of  $k$ -progressions in  $A$  is controlled by the pseudo-randomness condition  $\|A - \alpha\|_{U^{k-1}} = o(1)$ .

Green and Tao [33] determined the most general class of systems of linear forms that can be handled by Gowers' iterated Cauchy-Schwarz argument.

**Definition 10** (The Cauchy-Schwarz Complexity). Let  $\mathcal{L} = \{L_1, \dots, L_m\}$  be a system of linear forms. The *Cauchy-Schwarz complexity* of  $\mathcal{L}$  is the minimal  $k$  such that the following holds. For every  $1 \leq i \leq m$ , we can partition  $\{L_j\}_{j \in [m] \setminus \{i\}}$  into  $k+1$  subsets, such that  $L_i$  does not belong to the linear span of any of these subsets.

In particular, Green and Tao [33] showed that if the Cauchy-Schwarz complexity of  $\mathcal{L}$  is  $k$ , then

$$|t_{\mathcal{L}}(A) - \alpha^m| \leq m\|A - \alpha\|_{U^{k+1}}. \quad (48)$$

**Example IV.1.** The system of linear forms  $(x, x + y, \dots, x + (k-1)y)$ , which represents  $k$ -term arithmetic progressions, has CS-complexity  $k-2$ . It is also straightforward to verify that binary systems of linear forms have CS-complexity 1.

Later, in a series of articles [34], [35], Gowers and Wolf initiated a systematic study of classifying the systems of linear forms that are controlled by the  $k$ -th Gowers uniformity norm. They defined the *true complexity* of a system of linear forms  $L = (L_1, \dots, L_m)$  as the smallest  $k$  such that the pseudo-randomness condition of  $\|A - \alpha\|_{U^{k+1}} = o(1)$  implies  $|t_L(A) - \alpha^m| = o(1)$ .

By (48), the true complexity is at most the CS-complexity. Gowers and Wolf [34] fully characterized the true complexity of systems of linear forms, provided that the field size is not too small. Let  $L = \lambda_1 x_1 + \dots + \lambda_d x_d$  be a linear form in  $d$  variables. We identify  $L \equiv (\lambda_1, \dots, \lambda_d) \in \mathbb{F}^d$ , and define the  $k$ -th tensor power of  $L$  by

$$L^k = \left( \prod_{j=1}^k \lambda_{i_j} : i_1, \dots, i_k \in [d] \right) \in \mathbb{F}^{d^k}.$$

**Theorem IV.2** ([34]). *Let  $\mathcal{L} = (L_1, \dots, L_m)$  be a system of linear forms of CS-complexity at most  $|\mathbb{F}|$ . The true complexity of  $\mathcal{L}$  is the least integer  $k \geq 1$  such that the forms  $L_1^{k+1}, \dots, L_m^{k+1}$  are linearly independent.*

It is known that the true complexity of the linear system identifying 4-progressions is 2 [34], [35]. Moreover, it is well-known that there is a set  $A$  of density  $\alpha$  such that  $\|A - \alpha\|_{U^2} = o(1)$  (which implies combinatorial spreadness) but the count of 4-progressions in  $A$  is far from what is expected from a random set of density  $\alpha$  [22]. Here, we generalize this result. We essentially show that if the true complexity of a linear system  $\mathcal{L}$  is larger than 1, then there are sets  $A$  with density  $\alpha$  that are combinatorially spread with extremely good parameters, and yet  $t_{\mathcal{L}}(A) \not\approx \alpha^m$ .

**Theorem IV.3.** *Let  $\mathbb{F}$  be a finite field of prime order. Suppose that  $\mathcal{L} = (L_1, \dots, L_m)$  is a system of linear forms over  $d$  variables such that  $L_1^2, \dots, L_m^2$  are linearly dependent. There exists a set  $A \subseteq \mathbb{F}^n$  with density  $\alpha \geq \frac{1}{3}$  that is  $(e^{-\Omega(n)}, \Omega(n))$ -combinatorially spread, yet*

$$t_{\mathcal{L}}(A) > (1 + \Omega(1))\alpha^m.$$

As the proof makes clear, the constant  $\frac{1}{3}$  can be replaced by any constant strictly smaller than 1. We omit the proof and refer the reader to the full version of this paper [36].

#### ACKNOWLEDGMENT

We wish to thank Zander Kelley for bringing [19, Theorem 4.10] to our attention, which proved invaluable in extending Theorem I.5 from the 2-degenerate case to all binary systems. We wish to thank Tsun-Ming Cheung and Shachar Lovett for their valuable feedback on an earlier draft of this paper. K.H. would like to thank Anup Rao for valuable conversations. We wish to thank Dion Gijswijt for bringing [30] to our attention. We also thank Seth Pettie for pointing out a mistake in a previous version of the paper.

#### REFERENCES

- [1] A. Thomason, “Pseudorandom graphs,” in *Random graphs '85 (Poznań, 1985)*, ser. North-Holland Math. Stud. North-Holland, Amsterdam, 1987, vol. 144, pp. 307–331.
- [2] F. R. K. Chung, R. L. Graham, and R. M. Wilson, “Quasi-random graphs,” *Combinatorica*, vol. 9, pp. 345–362, 1989.
- [3] L. Lovász, *Large networks and graph limits*, ser. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2012, vol. 60.
- [4] Y. Kohayakawa, “Szemerédi’s regularity lemma for sparse graphs,” in *Foundations of computational mathematics (Rio de Janeiro, 1997)*. Springer, Berlin, 1997, pp. 216–230.
- [5] S. Gerke and A. Steger, “The sparse regularity lemma and its applications,” in *Surveys in combinatorics 2005*, ser. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2005, vol. 327, pp. 227–258.
- [6] F. Chung and R. Graham, “Sparse quasi-random graphs,” *Combinatorica*, vol. 22, no. 2, pp. 217–244, 2002.
- [7] Y. Kohayakawa and V. Rödl, “Regular pairs in sparse random graphs i,” *Random Structures & Algorithms*, vol. 22, no. 4, pp. 359–434, 2003.
- [8] Y. Kohayakawa, V. Rödl, and P. Sissokho, “Embedding graphs with bounded degree in sparse pseudorandom graphs,” *Israel Journal of Mathematics*, vol. 139, pp. 93–137, 2004.
- [9] Y. Kohayakawa, V. Rödl, M. Schacht, and J. Skokan, “On the triangle removal lemma for subgraphs of sparse pseudorandom graphs,” in *An Irregular Mind: Szemerédi is 70*. Springer, 2010, pp. 359–404.
- [10] D. Conlon, J. Fox, and Y. Zhao, “Extremal results in sparse pseudorandom graphs,” *Advances in Mathematics*, vol. 256, pp. 206–290, 2014.
- [11] P. Allen, J. Böttcher, J. Skokan, and M. Stein, “Regularity inheritance in pseudorandom graphs,” *Random Structures & Algorithms*, vol. 56, no. 2, pp. 306–338, 2020.
- [12] Z. Kelley, S. Lovett, and R. Meka, “Explicit separations between randomized and deterministic number-on-forehead communication,” in *Proceedings of the 56st Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, 2024.
- [13] K. F. Roth, “On certain sets of integers,” *J. London Math. Soc.*, vol. 28, no. 104–109, p. 3, 1953.
- [14] D. R. Heath-Brown, “Integer sets containing no arithmetic progressions,” *J. London Math. Soc. (2)*, vol. 35, no. 3, pp. 385–394, 1987.
- [15] E. Szemerédi, “Integer sets containing no arithmetic progressions,” *Acta Math. Hungar.*, vol. 56, no. 1–2, pp. 155–158, 1990.
- [16] J. Bourgain, “On triples in arithmetic progression,” *Geom. Funct. Anal.*, vol. 9, no. 5, pp. 968–984, 1999.
- [17] T. Sanders, “On Roth’s theorem on progressions,” *Ann. of Math. (2)*, vol. 174, no. 1, pp. 619–636, 2011.
- [18] T. F. Bloom and O. Sisask, “Breaking the logarithmic barrier in Roth’s theorem on arithmetic progressions,” 2021, <https://arxiv.org/abs/2007.03528>.
- [19] Z. Kelley and R. Meka, “Strong bounds for 3-progressions,” in *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2023, pp. 933–973.
- [20] T. F. Bloom and O. Sisask, “An improvement to the Kelley-Meka bounds on three-term arithmetic progressions,” 2023, <https://arxiv.org/abs/2309.02353>.
- [21] F. A. Behrend, “On sets of integers which contain no three terms in arithmetical progression,” *Proc. Nat. Acad. Sci. U.S.A.*, vol. 32, pp. 331–332, 1946.
- [22] W. T. Gowers, “A new proof of Szemerédi’s theorem,” *Geom. Funct. Anal.*, vol. 11, no. 3, pp. 465–588, 2001.
- [23] E. Szemerédi, “On sets of integers containing no  $k$  elements in arithmetic progression,” *Acta Arith.*, vol. 27, no. 199–245, p. 2, 1975.
- [24] J. Leng, A. Sah, and M. Sawhney, “Improved bounds for Szemerédi’s theorem,” 2024, <https://arxiv.org/abs/2402.17995>.
- [25] I. Z. Ruzsa and E. Szemerédi, “Triple systems with no six points carrying three triangles,” in *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, ser. Colloq. Math. Soc. János Bolyai. North-Holland, 1978, vol. 18, pp. 939–945.
- [26] P. Erdős, P. Frankl, and V. Rödl, “The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent,” *Graphs Combin.*, vol. 2, no. 2, pp. 113–121, 1986.
- [27] T. Sanders, “On the Bogolyubov–Ruzsa lemma,” *Analysis & PDE*, vol. 5, no. 3, pp. 627 – 655, 2012.

- [28] E. Croot, V. F. Lev, and P. P. Pach, “Progression-free sets in  $\mathbb{Z}_4^n$  are exponentially small,” *Ann. of Math. (2)*, vol. 185, no. 1, pp. 331–337, 2017.
- [29] J. S. Ellenberg and D. Gijswijt, “On large subsets of  $\mathbb{F}_q^n$  with no three-term arithmetic progression,” *Ann. of Math. (2)*, vol. 185, no. 1, pp. 339–343, 2017.
- [30] D. Gijswijt, “Excluding affine configurations over a finite field,” *Discrete Anal.*, pp. Paper No. 21, 24, 2023.
- [31] T. Tao and V. H. Vu, *Additive combinatorics*, paperback ed., ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010, vol. 105.
- [32] T. F. Bloom and O. Sisask, “Logarithmic bounds for Roth’s theorem via almost-periodicity,” *Discrete Anal.*, pp. Paper No. 4, 20, 2019.
- [33] B. Green and T. Tao, “Linear equations in primes,” *Ann. of Math. (2)*, pp. 1753–1850, 2010.
- [34] W. T. Gowers and J. Wolf, “Linear forms and higher-degree uniformity for functions on  $\mathbb{F}_p^n$ ,” *Geom. Funct. Anal.*, vol. 21, no. 1, pp. 36–69, 2011.
- [35] ———, “The true complexity of a system of linear equations,” *Proc. Lond. Math. Soc. (3)*, vol. 100, no. 1, pp. 155–176, 2010.
- [36] Y. Filmus, H. Hatami, K. Hosseini, and E. Kelman, “Sparse graph counting and kelley-meka bounds for binary systems,” 2024. [Online]. Available: <https://arxiv.org/abs/2311.12248>