

A VARIANT OF THE HYPERGRAPH REMOVAL LEMMA

TERENCE TAO

ABSTRACT. Recent work of Gowers [10] and Nagle, Rödl, Schacht, and Skokan [15], [19], [20] has established a hypergraph removal lemma, which in turn implies some results of Szemerédi [26] and Furstenberg-Katznelson [7] concerning one-dimensional and multi-dimensional arithmetic progressions respectively. In this paper we shall give a self-contained proof of this hypergraph removal lemma. In fact we prove a slight strengthening of the result, which we will use in a subsequent paper [29] to establish (among other things) infinitely many constellations of a prescribed shape in the Gaussian primes.

1. INTRODUCTION

In this paper we prove a slight variant of the hypergraph removal lemma established recently and independently by Gowers [10] and Nagle, Rödl, Schacht and Skokan [15], [19], [20]. To motivate this lemma, let us first recall the more well-known triangle removal lemma from graph theory of Ruzsa and Szemerédi [22]. It will be convenient to work in the setting of tripartite graphs, though we will comment about the generalization to general graphs shortly. We adopt the following $o()$ and $O()$ notation: If $x, y_1, \dots, y_n$ are parameters, we use $o_{x \rightarrow 0; y_1, \dots, y_n}(X)$ to denote any quantity bounded in magnitude by $Xc(x, y_1, \dots, y_n)$, where $c()$ is a function which goes to zero as $x \rightarrow 0$ for each fixed choice of $y_1, \dots, y_n$. Similarly, we use $O_{y_1, \dots, y_n}(X)$ to denote any quantity bounded by $XC(y_1, \dots, y_n)$, for some function $C()$ of $y_1, \dots, y_n$. If A is a finite set, we use $|A|$ to denote the cardinality of A .

Theorem 1.1 (Triangle removal lemma, tripartite graph version). [22] *Let V_1, V_2, V_3 be finite non-empty sets of vertices, and let $G = (V_1, V_2, V_3, E_{12}, E_{23}, E_{31})$ be a tripartite graph on these sets of vertices, thus $E_{ij} \subseteq V_i \times V_j$ for $ij = 12, 23, 31$. Suppose that the number of triangles in this graph does not exceed $\delta|V_1||V_2||V_3|$ for some $0 < \delta < 1$. Then there exists a graph $G' = G'(V_1, V_2, V_3, E'_{12}, E'_{23}, E'_{31})$ which contains no triangles whatsoever, and such that $|E_{ij} \setminus E'_{ij}| = o_{\delta \rightarrow 0}(|V_i \times V_j|)$ for $ij = 12, 23, 31$.*

One can view G' as a “triangle-free approximation” to G . Note that we do not assume that G' is a subgraph of G , but one can easily obtain this conclusion by replacing E'_{ij} with $E'_{ij} \cap E_{ij}$ if desired (i.e. one replaces G' by $G' \cap G$). As we shall see, however, it will be convenient to allow the possibility that G' is not a subgraph of G .

Remark 1.2. The above theorem is phrased for tri-partite graphs, but it quickly implies an analogous version for non-partite graphs $G = (V, E)$, by taking three

copies $V_1 = V_2 = V_3 = V$ of the vertex set V , and constructing the bipartite graph $\tilde{G} = (V_1, V_2, V_3, E_{12}, E_{23}, E_{31})$, where E_{ij} consists of those pairs (x, y) which are the endpoints of an edge in E . We omit the details.

It was observed in [22] that Theorem 1.1 implies Roth’s famous theorem [21] that subsets of integers of positive density contain infinitely many progressions of length three. In [24] it was also observed that Theorem 1.1 also implies that subsets of $\mathbf{Z}^2$ with positive density contain infinitely many right-angled triangles (a result first obtained in [1]). It was observed earlier (for instance in [16] or [5]) that an extension of the triangle removal lemma to hypergraphs would similarly imply Szemerédi’s famous theorem [26] on progressions of arbitrary length; by modifying the observation in [24], it would also imply a multidimensional extension of that theorem due to Furstenberg and Katznelson [7]. We shall return to this issue in the sequel [29] to this paper, and discuss the above hypergraph removal lemma in detail later in this introduction.

Theorem 1.1 was proven using the *Szemerédi regularity lemma* (see e.g. [27], [14] for a survey of this lemma and its applications), which roughly speaking allows one to approximate an arbitrary large and complex graph to arbitrary accuracy by a much simpler object; see also [32], [23] for further refinements of Theorem 1.1. This proof in fact yields a little bit more information on the triangle-free approximation G' to G , namely that G' can be chosen to be “bounded complexity”. More precisely:

Theorem 1.3 (Strong triangle removal lemma, tripartite graph version). [22] *Let V_1, V_2, V_3 be finite non-empty sets of vertices, and let $G = (V_1, V_2, V_3, E_{12}, E_{23}, E_{31})$ be a tri-partite graph on these sets of vertices. Suppose that G contains at most $\delta|V_1||V_2||V_3|$ triangles. Then there exists a graph $G' = G'(V_1, V_2, V_3, E'_{12}, E'_{23}, E'_{31})$ which contains no triangles whatsoever, and such that $|E_{ij} \setminus E'_{ij}| = o_{\delta \rightarrow 0}(|V_i \times V_j|)$ for $ij = 12, 23, 31$. Furthermore, there exists a quantity $M = O_{\delta}(1)$, and partitions $V_i = V_{i,1} \cup \dots \cup V_{i,M}$ for each $i = 1, 2, 3$ into sets $V_{i,a}$ (some of which may be empty) such that for each $ij = 12, 23, 31$, E'_{ij} is the union of sets of the form $V_{i,a} \times V_{j,b}$.*

Note that the graph G' constructed in Theorem 1.3 will typically not be a subgraph of G . One could make the sets $V_{i,1}, \dots, V_{i,M}$ to be the same size (with at most one exception for each i) without much difficulty but we will not endeavour to do so here. There is also a version of this lemma for non-tripartite graphs which is well known (and essentially equivalent to the tripartite version) but we will not reproduce it here.

It turns out that Theorem 1.1 and Theorem 1.3 can be rephrased in a more “probabilistic” manner. One reason for doing this is because in our arguments we will need two basic concepts from probability theory, which are *conditional expectation* and *complexity* respectively. It seems that with the aid of these concepts, the proofs become somewhat cleaner to give¹. To explain these concepts we need some notation. For reasons which will become clearer later, we shall use a rather general notation which incorporates the above Theorems as a special case.

¹For a more traditional combinatorial approach to these problems, see [17].

Definition 1.4 (Hypergraphs). If J is a finite set and $d \geq 0$, we define $\binom{J}{d} := \{e \subseteq J : |e| = d\}$ to be the set of all subsets of J of cardinality d . A d -uniform hypergraph on J is then defined to be any subset $H_d \subseteq \binom{J}{d}$ of $\binom{J}{d}$. For instance, an undirected graph $G = (V, E)$ without loops can be viewed as a 2-uniform hypergraph on V .

Example 1.5. If $J := \{1, 2, 3\}$, then the triangle $H_2 := \binom{J}{2} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$ is a 2-uniform hypergraph on J .

Definition 1.6 (Hypergraph systems). A *hypergraph system* is a quadruplet $V = (J, (V_j)_{j \in J}, d, H_d)$, where J is a finite set, $(V_j)_{j \in J}$ is a collection of finite non-empty sets indexed by J , $d \geq 1$ is positive integer, and $H_d \subseteq \binom{J}{d}$ is a d -uniform hypergraph. For any $e \subseteq J$, we set $V_e := \prod_{j \in e} V_j$, and let $\pi_e : V_J \rightarrow V_e$ be the canonical projection map.

Remark 1.7. Very roughly speaking, a hypergraph system corresponds to the notion of a *measure-preserving system*² in ergodic theory, though with the notable difference that no analogue of the shift operator exists in a hypergraph system. Indeed the V_j are simply finite sets, and need not have any additive structure whatsoever.

Definition 1.8 (Conditional expectation). Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system. If $f : V_J \rightarrow \mathbf{R}$ is a function, we define the expectation $\mathbf{E}(f) = \mathbf{E}(f(x)|x \in V_J)$ by the formula

$$\mathbf{E}(f) = \mathbf{E}(f(x)|x \in V_J) := \frac{1}{|V_J|} \sum_{x \in V_J} f(x).$$

Similarly, if $\mathcal{B}$ is a σ -algebra³ on V_J , i.e. a collection of sets in V_J which contains $\emptyset$ and V_J , and is closed under unions, intersections, and complementation, we define the *conditional expectation* $\mathbf{E}(f|\mathcal{B}) : V_J \rightarrow \mathbf{R}$ by the formula

$$\mathbf{E}(f|\mathcal{B})(x) := \frac{1}{|\mathcal{B}(x)|} \sum_{y \in \mathcal{B}(x)} f(y),$$

where $\mathcal{B}(x)$ is the smallest element of $\mathcal{B}$ which contains x . For each $e \subseteq J$, let $\mathcal{A}_e$ be the σ -algebra on V_J defined by $\mathcal{A}_e := \{\pi_e^{-1}(E) : E \subseteq V_e\}$. In other words, $\mathcal{A}_e$ consists of those subsets of V_J , membership of which is determined solely by the co-ordinates of V_J indexed by e .

²A measure preserving system is a probability space $(X, \mathcal{B}, \mu)$ together with a shift $T : X \rightarrow X$ that preserves the measure μ . The ergodic approach to Szemerédi's theorem, as introduced by Furstenberg[6], recasts the problem of finding arithmetic progressions as that of understanding averages such as $\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^n A \cap \dots \cap T^{(k-1)n} A)$. This can in turn be viewed as the problem of understanding shift operators such as $(T, T^2, \dots, T^{k-1})$ on a product space $X \times \dots \times X$. This has some intriguing parallels with the combinatorial approach, in which the problem of obtaining arithmetic progressions in a set V is reduced to that of analyzing Cayley-type graphs or hypergraphs, which can be viewed as subsets of $V \times \dots \times V$. We do not know of any formal connection between these two approaches, nevertheless there do appear to be some interesting similarities.

³Of course, since V_J is finite, we do not need to distinguish finite unions and countable unions, and could simply call $\mathcal{B}$ an “algebra”, or even a “partition”; the latter notation is in fact used in most treatments of the regularity lemma. However we prefer the notation of σ -algebra as being highly suggestive, evoking ideas and insights from probability theory, measure theory, and information theory.

One can interpret the usage of these averages as imposing the uniform probability distribution on each V_e , which basically amounts to introducing a set $(x_j)_{j \in J}$ of independent random variables, with each x_j ranging uniformly in V_j .

If $\mathcal{B}_1$ and $\mathcal{B}_2$ are two σ -algebras on V_J , we use $\mathcal{B}_1 \vee \mathcal{B}_2$ to denote the smallest σ -algebra that contains both $\mathcal{B}_1$ and $\mathcal{B}_2$; this corresponds to the familiar concept of the *common refinement* of two partitions. We can more generally define $\bigvee_{i \in I} \mathcal{B}_i$ for any collection $(\mathcal{B}_i)_{i \in I}$ of σ -algebras.

Example 1.9. For any finite non-empty sets V_1, V_2, V_3 , the quadruplet $V = (J, (V_j)_{j \in J}, 2, H_2)$ is a hypergraph system, where $J := \{1, 2, 3\}$ and $H_2 := \binom{J}{2}$ are as in Example 1.5. The σ -algebra $\mathcal{A}_{\{1,2\}}$ is the algebra of all subsets of $V_1 \times V_2 \times V_3$ which do not depend on the third variable, and thus take the form $E \times V_3$ for some $E \subseteq V_1 \times V_2$. Similarly for $\mathcal{A}_{\{2,3\}}$ and $\mathcal{A}_{\{3,1\}}$.

Definition 1.10 (Complexity). Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system. If $\mathcal{B}$ is a σ -algebra in V_J , we define the *complexity* $\text{complex}(\mathcal{B})$ of $\mathcal{B}$ to be the least number of sets in V_J needed to generate $\mathcal{B}$ as a σ -algebra; this can be viewed as a simplified version of the Shannon entropy $\mathbf{H}(\mathcal{B})$, which we will not use here. We observe the obvious inequalities

$$\text{complex}(\mathcal{B}_1 \vee \mathcal{B}_2) \leq \text{complex}(\mathcal{B}_1) + \text{complex}(\mathcal{B}_2) \text{ for arbitrary } \mathcal{B}_1, \mathcal{B}_2 \quad (1)$$

and

$$|\mathcal{B}| \leq 2^{2^{\text{complex}(\mathcal{B})}}. \quad (2)$$

Remark 1.11. If one views $\mathcal{B}$ as a partition, the complexity is essentially the logarithm of the number of cells in the partition. From an information-theoretic perspective, the complexity measures how many bits of information are needed to know which atom of $\mathcal{B}$ a given point in V_J lies in.

If E is a subset of V_J , we let $1_E : V_J \rightarrow \mathbf{R}$ be the indicator function, thus $1_E(x) := 1$ when $x \in E$ and $1_E(x) := 0$ otherwise. In particular, $\mathbf{E}(1_E) = |E|/|V_J|$ can be viewed as the “density” or “probability” of E in V_J .

With all this notation, Theorem 1.3 becomes

Theorem 1.12 (Strong triangle removal lemma, σ -algebra version). *Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system with $J = \{1, 2, 3\}$, $d = 2$, and $H_d = \binom{J}{2} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$. For each $e \in H_d$, let E_e be a set in $\mathcal{A}_e$ such that*

$$\mathbf{E}\left(\prod_{e \in H_d} 1_{E_e}\right) \leq \delta$$

for some $0 < \delta < 1$. Then there exist sets $E'_e \in \mathcal{A}_e$ for $e \in H_d$ such that

$$\bigcap_{e \in H_d} E'_e = \emptyset$$

and

$$\mathbf{E}(1_{E_e \setminus E'_e}) = o_{\delta \rightarrow 0}(1) \text{ for all } e \in H_d.$$

Furthermore, for each $i \in J$ there exists sub-algebras $\mathcal{B}_i \subseteq \mathcal{A}_{\{i\}}$ such that

$$\text{complex}(\mathcal{B}_i) = O_{\delta}(1) \text{ for } i \in J$$

and

$$E'_e \in \bigvee_{i \in e} \mathcal{B}_i \text{ for } e \in H_d.$$

It is easy to see that Theorem 1.3 and Theorem 1.12 are equivalent. The notation here may appear quite cumbersome, but the advantages of these notations will hopefully become more apparent when we prove a generalization of this result shortly.

The case of $d = 2$, and J and H_d arbitrary, was treated in [3]. It was then conjectured in that paper that a result of the above type should also hold for higher d . The generalization of Theorem 1.1 to the higher d case was accomplished only recently and independently by Gowers [11] and Nagle, Rödl, Schacht, Skokan [15], [19], [20], using the language of hypergraphs. It turns out that Theorem 1.3 or Theorem 1.12 can similarly be generalized, and with the notation already developed, the extension is very easy to state:

Theorem 1.13 (Hypergraph removal lemma). [11], [15], [19], [20] *Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system. For each $e \in H$, let E_e be a set in $\mathcal{A}_e$ such that*

$$\mathbf{E}\left(\prod_{e \in H_d} 1_{E_e}\right) \leq \delta \quad (3)$$

for some $0 < \delta < 1$. Then for each $e \in H_d$ there exists a set $E'_e \in \mathcal{A}_e$ such that

$$\bigcap_{e \in H_d} E'_e = \emptyset \quad (4)$$

and

$$\mathbf{E}(1_{E_e \setminus E'_e}) = o_{\delta \rightarrow 0; J}(1) \text{ for all } e \in H_d. \quad (5)$$

Furthermore, there exist sub-algebras $\mathcal{B}_{e'} \subseteq \mathcal{A}_{e'}$ whenever $e' \subset J$ and $|e'| < d$ obeying the complexity estimate

$$\text{complex}(\mathcal{B}_{e'}) = O_{J, \delta}(1) \text{ whenever } e' \subseteq J \text{ and } |e'| < d \quad (6)$$

(so in particular $|\mathcal{B}_{e'}| = O_{J, \delta}(1)$, thanks to (2)) and

$$E'_e \in \bigvee_{e' \subsetneq e} \mathcal{B}_{e'} \text{ for all } e \in H_d. \quad (7)$$

Clearly Theorem 1.12 is a special case of Theorem 1.13. We have attributed this theorem to Gowers [11] and Nagle-Rödl-Schacht-Skokan [15], [19], [20] because it follows from their methods, although a theorem of this type is not stated explicitly in those papers. One can formulate variants of this removal lemma in the case when H_d is not d -uniform but we will not do so here. A related result has recently been obtained in [17], using techniques similar in spirit to those here (though with substantially different notation).

The main purpose of this paper is to explicitly prove Theorem 1.13 in a completely self-contained manner. In a subsequent paper [29], we will then transfer this theorem (as in [12]) to obtain a relative version of Theorem 1.13, restricted to a suitably

pseudorandom subset of $\prod_j V_j$. This will then be used (again following [12]) to deduce the existence of infinitely many constellations of a prescribed shape in the Gaussian primes and similar sets.

As a corollary of Theorem 1.13, we obtain the hypergraph removal lemma in a formulation closer to that of Gowers or Nagle-Rödl-Schacht-Skokan:

Corollary 1.14 (Hypergraph removal lemma, partite hypergraph version). [11], [15], [19], [20] *Let $(V_j)_{j \in J}$ be a collection of finite non-empty sets. Let $0 \leq d \leq |J|$, and let $H_d \subseteq \binom{J}{d}$ be a d -uniform hypergraph on J . For each $e \in H_d$, let E_e be a subset of $\prod_{j \in e} V_j$. Suppose that*

$$|\{(x_j)_{j \in J} \in \prod_{j \in J} V_j : (x_j)_{j \in e} \in E_e \text{ for all } e \in H_d\}| \leq \delta \prod_{j \in J} |V_j|$$

for some $0 < \delta \leq 1$; in other words, the J -partite hypergraph $G = ((V_j)_{j \in J}, (E_e)_{e \in H_d})$ contains at most $\delta \prod_{j \in J} |V_j|$ copies of H_d . Then for each $e \in H_d$ there exists $E'_e \subset \prod_{j \in e} V_j$ such that

$$\{(x_j)_{j \in J} \in \prod_{j \in J} V_j : (x_j)_{j \in e} \in E'_e \text{ for all } e \in H_d\} = \emptyset$$

(i.e. the J -partite hypergraph $G' = G'((V_j)_{j \in J}, (E'_e)_{e \in H_d})$ contains no copies of H_d whatsoever), and such that $|E_e \setminus E'_e| = o_{\delta \rightarrow 0; |J|}(\prod_{j \in e} |V_j|)$ for all $e \in H_d$.

The deduction of Corollary 1.14 from Theorem 1.13 is analogous to the deduction of Theorem 1.1 from Theorem 1.12 and is omitted. It seems quite likely that we can obtain similar analogues for non-partite hypergraphs, just as was the case with the non-partite version of Theorem 1.1; see [11], [15], [19], [20] for some examples of this, though for applications to Szemerédi-type theorems it is the partite version which is of importance. It should be unsurprising that Theorem 1.1 is then the special case of Corollary 1.14 applied to the (hyper)graph in Example 1.5. The case $|J| = 4$ and $H_3 = \binom{J}{3}$ was treated in [5]. Just as Theorem 1.1 implies Roth's theorem, Corollary 1.14 implies Szemerédi's theorem [26] on arithmetic progressions, as well as the multidimensional generalization of that theorem due to Furstenberg and Katznelson [7]; see [25], [5], [11], [20] for further discussion⁴. Thus this paper provides a moderately short and self-contained proof of these theorems, although we emphasize that this goal was already achieved in the prior work of [11], [15], [19], [20].

The remainder of this paper is devoted to proving Theorem 1.13. As one might expect from the previous proofs of these types of results, our proof shall proceed by proving a “hypergraph regularity lemma” and a “hypergraph counting lemma”. The arguments are broadly along similar lines to those of Gowers or Nagle, Rödl, Schacht, and Skokan, although it seems that using the notation of σ -algebras and probability theory allows for slightly cleaner arguments.

⁴It was also recently observed that this hypergraph removal result also implies another theorem of Furstenberg and Katznelson [8] on affine subspaces of dense subsets of high-dimensional finite field vector spaces; see [18].

The author thanks Fan Chung Graham, Vojtěch Rödl, Mathias Schacht, and Jozsef Solymosi for helpful comments and references. He is particularly indebted to Mathias Schacht for supplying the recent preprint [17], and to the anonymous referees for a careful reading of the paper and many cogent suggestions and corrections. The author is supported by a grant from the Packard foundation.

2. PSEUDORANDOMNESS AND THE REGULARITY LEMMA

Henceforth the hypergraph system $V = (J, (V_j)_{j \in J}, d, H_d)$ will be fixed. In this section we shall state and prove a σ -algebra version of the hypergraph regularity lemma (Lemma 2.9). This lemma establishes a dichotomy between pseudorandomness (or ε -regularity, or small discrepancy) on one hand, and bounded complexity⁵ on the other; the regularity lemma then asserts, very roughly speaking, that any given set or σ -algebra (or family of σ -algebras) can be split into a component with bounded complexity, and a component which is pseudorandom (has small discrepancy).

In order to state the regularity lemma we need to formalize the notion of pseudorandomness (or more precisely, of discrepancy). We shall also need a notion of the *energy* of a σ -algebra in order to keep track of the inductions that go into the proof of the regularity lemma, and also in the final statement of our regularity lemma.

We shall not state the final regularity lemma we need (Lemma 2.9) immediately. To begin with, we set out our notation for discrepancy and energy. Initially we shall be focusing primarily on a single edge $e \subseteq J$, as opposed to an entire hypergraph H_d , though this hypergraph shall emerge later in this section.

Definition 2.1 (*e-discrepancy*). For any $e \subseteq J$, we define the *skeleton* ∂e of e to be the set $\{f \subsetneq e : |f| = |e| - 1\}$. If $e \subseteq J$, $E_e \subseteq V_J$, and $\mathcal{B}$ is a σ -algebra on V_J , we define the *e-discrepancy* $\Delta_e(E_e|\mathcal{B})$ of the set E_e with respect to the σ -algebra $\mathcal{B}$ to be the quantity⁶

$$\Delta_e(E_e|\mathcal{B}) := \sup_{E_f \in \mathcal{A}_f \forall f \in \partial e} |\mathbf{E}((1_{E_e} - \mathbf{E}(1_{E_e}|\mathcal{B})) \prod_{f \in \partial e} 1_{E_f})| \quad (8)$$

where the supremum is over all collections of sets $(E_f)_{f \in \partial e}$, where each E_f lies in the σ -algebra $\mathcal{A}_f$. Note that since V_J is finite, so is $\Delta_e(E_e|\mathcal{B})$.

Roughly speaking, the *e-discrepancy* $\Delta_e(E_e|\mathcal{B})$ measures the amount of “structure” in E_e which is not already captured by the σ -algebra $\mathcal{B}$. By “structure”, we mean sets which can be easily described by sets from the lower order σ -algebras $\mathcal{A}_f$, as opposed to a generic set in $\mathcal{A}_e$ which in general is likely to have no good decomposition (or approximate decomposition) into sets from the $\mathcal{A}_f$. Thus if $\Delta_e(E_e|\mathcal{B})$ is small, we expect E_e to behave randomly (i.e. in an unstructured way) on most

⁵This is very similar to the dichotomy between weak mixing and compactness in ergodic theory, which is of great utility in proving statements such as Szemerédi’s theorem; it seems of interest to explore these connections further.

⁶This quantity is related to the Gowers uniformity norms used for instance in [10], [11], [12], but we will not explicitly introduce those norms here. This quantity is also related to the notion of a pseudorandom hypergraph, studied for instance in [13].

atoms of $\mathcal{B}$. The $\Delta_e(E_e|\mathcal{B})$ generalize the concept of ε -regularity, as the following example shows:

Example 2.2. Let $G = (V_1, V_2, E_{12})$ be a bipartite graph between two finite non-empty sets V_1, V_2 ; we can thus view E_{12} as a set in $\mathcal{A}_{\{1,2\}}$, where V is the hypergraph system $V = (J, (V_j)_{j \in J}, d, H_d)$ with $J = \{1, 2\}$, $d = 2$, and $H_d = \binom{J}{d} = \{\{1, 2\}\}$. Suppose that E_{12} has density $\mathbf{E}(1_{E_{12}}) = \sigma$ (i.e. $\sigma = |E_{12}|/|V_1||V_2|$), and that

$$\Delta_{\{1,2\}}(E_{12}|\mathcal{A}_\emptyset) \leq \varepsilon$$

for some $\varepsilon > 0$. Then by definition we have

$$|\mathbf{E}((1_{E_{12}} - \sigma)1_{E_1}1_{E_2})| \leq \varepsilon \text{ whenever } E_1 \in \mathcal{A}_{\{1\}}, E_2 \in \mathcal{A}_{\{2\}}.$$

In the original setting of the bipartite graph G , this is equivalent to asserting that

$$||E_{12} \cap (E_1 \times E_2)| - \sigma|E_1||E_2|| \leq \varepsilon|V_1||V_2|$$

for all $E_1 \subseteq V_1$ and $E_2 \subseteq V_2$. The reader may recognize this as a pseudorandomness condition or ε -regularity condition on the graph G . If we replace $\mathcal{A}_\emptyset$ by a finer σ -algebra such as $\mathcal{B}_1 \vee \mathcal{B}_2$ for some $\mathcal{B}_1 \subseteq \mathcal{A}_{\{1\}}$ and $\mathcal{B}_2 \subseteq \mathcal{A}_{\{2\}}$, where the complexity of $\mathcal{B}_1$ and $\mathcal{B}_2$ is small compared to $1/\varepsilon$, then a condition such as $\Delta_{\{1,2\}}(E_{12}|\mathcal{B}_1 \vee \mathcal{B}_2) \leq \varepsilon$ states, roughly speaking, that the graph G is ε -regular on “most” of the atoms $A_1 \times A_2$ in the partition associated to $\mathcal{B}_1 \vee \mathcal{B}_2$.

If $\mathcal{B}$ is a σ -algebra on V_J and E is a set in V_J (not necessarily in $\mathcal{B}$), we define the *E-energy* of $\mathcal{B}$ to be the quantity

$$\mathcal{E}_E(\mathcal{B}) := \mathbf{E}(|\mathbf{E}(1_E|\mathcal{B})|^2).$$

Clearly, the *E-energy* $\mathcal{E}_E(\mathcal{B})$ ranges between 0 and 1; intuitively, $\mathcal{E}_E(\mathcal{B})$ is a measure of how much information about E is captured by $\mathcal{B}$, and is thus in many ways complementary to the *e-discrepancy* $\Delta_e(E|\mathcal{B})$. From Pythagoras’ theorem we can verify the identity

$$\mathcal{E}_E(\mathcal{B}') = \mathcal{E}_E(\mathcal{B}) + \mathbf{E}(|\mathbf{E}(1_E|\mathcal{B}') - \mathbf{E}(1_E|\mathcal{B})|^2) \text{ whenever } \mathcal{B} \subseteq \mathcal{B}', \quad (9)$$

thus finer σ -algebras have larger *E-energy*.

Remark 2.3. In the setting of Example 2.2 with $\mathcal{B} = \mathcal{B}_1 \vee \mathcal{B}_2$ for some $\mathcal{B}_1 \subseteq \mathcal{A}_{\{1\}}$ and $\mathcal{B}_2 \subseteq \mathcal{A}_{\{2\}}$, the energy is a familiar quantity in the theory of the regularity lemma, and is usually referred to as the *index* of the partition; see [27].

Let us informally say that a set $E_e \in \mathcal{A}_e$ is *e-pseudorandom with respect to $\mathcal{B}$* if the *e-discrepancy* $\Delta_e(E_e|\mathcal{B})$ is small. A fundamental fact (which was already exploited in [26], [27]) is that if E is *not e-pseudorandom* with respect to $\mathcal{B}$, then we can find a refinement of $\mathcal{B}$ with higher energy and not much larger complexity:

Lemma 2.4 (Large discrepancy implies energy increment). *Let $e \subseteq J$, let $E_e \in \mathcal{A}_e$ be a set, and for each $f \in \partial e$ let $\mathcal{B}_f \subseteq \mathcal{A}_f$ be a σ -algebra such that*

$$\Delta_e(E_e|\bigvee_{f \in \partial e} \mathcal{B}_f) \geq \varepsilon$$

for some $\varepsilon > 0$. Then there exists a σ -algebra $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ for all $f \in \partial e$ such that

$$\text{complex}(\mathcal{B}'_f) \leq \text{complex}(\mathcal{B}_f) + 1 \quad (10)$$

and

$$\mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}'_f) \geq \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}_f) + \varepsilon^2. \quad (11)$$

Proof By (8) (and the finiteness of V_J) we can find sets $E_f \in \mathcal{A}_f$ for all $f \in \partial e$ such that

$$|\mathbf{E} \left((1_{E_e} - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}_f)) \prod_{f \in \partial e} 1_{E_f} \right) | \geq \varepsilon.$$

For each $f \in \partial e$, let $\mathcal{B}'_f$ be the σ -algebra

$$\mathcal{B}'_f := \mathcal{B}_f \vee \mathcal{B}(E_f)$$

then we have $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$, and obtain (10) from (1). Since $\prod_{f \in \partial e} 1_{E_f}$ is measurable with respect to $\bigvee_{f \in \partial e} \mathcal{B}'_f$, and $1_{E_e} - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f)$ has zero conditional expectation with respect to $\bigvee_{f \in \partial e} \mathcal{B}'_f$ we see that

$$\mathbf{E} \left((1_{E_e} - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f)) \prod_{f \in \partial e} 1_{E_f} \right) = 0$$

and hence

$$|\mathbf{E} \left((\mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}_f)) \prod_{f \in \partial e} 1_{E_f} \right) | \geq \varepsilon.$$

By the boundedness of $\prod_{f \in \partial e} 1_{E_f}$ and the Cauchy-Schwarz inequality we conclude

$$\mathbf{E} \left(|\mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}_f)|^2 \right) \geq \varepsilon^2,$$

and (11) then follows from (9). $\blacksquare$

By iterating Lemma 2.4, one expects to be able to show that any given set $E_e \in \mathcal{A}_e$ must be e -pseudorandom with respect to a σ -algebra $\mathcal{B}$ of bounded complexity, since otherwise we could create a tower of σ -algebras whose energy increments indefinitely. Such statements can be viewed as σ -algebra analogues of the Szemerédi regularity lemma. There are several such lemmas available; the final lemma which we need is a bit lengthy to state, so we begin by stating some simpler regularity lemmas which we will then iterate to obtain the stronger lemmas which we need. We first obtain a preliminary iteration of Lemma 2.4, in which the single set $E_e \in \mathcal{A}_e$ is replaced by an ensemble of sets, or more precisely an ensemble $(\mathcal{B}_e)_{e \in H}$ of σ -algebras with bounded complexity.

If H_d is a d -uniform hypergraph, we define ∂H_d to be the $(d-1)$ -uniform hypergraph $\partial H_d := \bigcup_{e \in H_d} \partial e$.

Lemma 2.5 (Dichotomy between randomness and structure). *Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system. For each $e \in H_d$, let $\mathcal{B}_e \subseteq \mathcal{A}_e$ be a σ -algebra with the complexity bounds*

$$\text{complex}(\mathcal{B}_e) \leq m \text{ for all } e \in H_d$$

for some $m > 0$, and for each $f \in \partial H_d$, let $\mathcal{B}_f \subseteq \mathcal{A}_f$ be a σ -algebra with the complexity bounds

$$\text{complex}(\mathcal{B}_f) \leq M \text{ for all } f \in \partial H_d$$

for some $M > 0$. Let $\varepsilon, \delta > 0$. Then one of the following statements must hold.

- (Randomness) *There exists σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ for all $f \in \partial H_d$ such that*

$$\mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}'_f) < \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}_f) + \varepsilon^2 \text{ for all } e \in H_d \text{ and } E_e \in \mathcal{B}_e \quad (12)$$

and

$$\Delta_e(E_e | \bigvee_{f \in \partial e} \mathcal{B}'_f) \leq \delta \text{ for all } e \in H_d \text{ and } E_e \in \mathcal{B}_e. \quad (13)$$

- (Structure) *There exist σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ for all $f \in \partial H_d$ such that*

$$\mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}'_f) \geq \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}_f) + \varepsilon^2 \text{ for some } e \in H_d \text{ and } E_e \in \mathcal{B}_e \quad (14)$$

and

$$\text{complex}(\mathcal{B}'_f) \leq M + O_{|J|, m, \varepsilon, \delta}(1) \text{ for all } f \in \partial H_d. \quad (15)$$

Proof We run the following algorithm:

- Step 0. Initialize $\mathcal{B}'_f := \mathcal{B}_f$ for all $f \in \partial H_d$. Note that (12) and (15) currently hold.
- Step 1. If (13) holds, then we halt the algorithm (we are in the “randomness” half of the dichotomy). Otherwise, there exists an $e \in H$ and $E_e \in \mathcal{B}_e$ such that

$$\Delta_e(E_e | \bigvee_{f \in \partial e} \mathcal{B}'_f) > \delta.$$

We can then invoke Lemma 2.4 to locate refinements $\mathcal{B}'_f \subseteq \mathcal{B}''_f \subseteq \mathcal{A}_f$ for all $f \in \partial H_d$ (note that $\mathcal{B}''_f$ will just equal $\mathcal{B}'_f$ if $f \not\subset e$) such that

$$\text{complex}(\mathcal{B}''_f) \leq \text{complex}(\mathcal{B}'_f) + 1 \text{ for all } f \in \partial H_d$$

and

$$\mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}''_f) \geq \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}'_f) + \delta^2.$$

- Step 2. We replace $\mathcal{B}'_f$ with $\mathcal{B}''_f$ for all $f \in \partial H_d$. If (12) fails (i.e. (14) holds), then we halt the algorithm (we are in the “structure” half of the dichotomy). Otherwise, we return to Step 1.

Observe that every time we return from Step 2 to Step 1, the quantity

$$\sum_{e \in H_d} \sum_{E_e \in \mathcal{B}_e} \mathcal{E}_{E_e} \left(\bigvee_{f \in \partial e} \mathcal{B}'_f \right)$$

increases by at least δ^2 . On the other hand, if this quantity ever increases by more than $|H_d|2^{2^m}\varepsilon^2 = O_{|J|,m,\varepsilon}(1)$, then by (2) and the pigeonhole principle (12) will necessarily fail. Since we only return to Step 1 when (12) holds, we see that the algorithm can only iterate at most $O_{|J|,m,\varepsilon,\delta}(1)$ times. Thus when we terminate we must have (15). The claim then follows. $\blacksquare$

We now iterate Lemma 2.5 to obtain the following preliminary regularity lemma. Define a *growth function* to be an increasing function $F : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ such that $F(x) \geq 1 + x$ for all x .

Lemma 2.6 (Preliminary regularity lemma). *Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system. For each $e \in H_d$ let $\mathcal{B}_e \subseteq \mathcal{A}_e$ be a σ -algebra, and suppose that we have the bound*

$$\text{complex}(\mathcal{B}_e) \leq m \text{ for all } e \in H_d$$

for some $m > 0$. Let $\varepsilon > 0$, and let F be a growth function (possibly depending on ε). Then there exists $M > 0$, and for each $f \in \partial H_d$ there exists a pair of σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ such that we have the estimates

$$F(m) \leq M \leq O_{|J|,\varepsilon,m,F}(1) \tag{16}$$

$$\text{complex}(\mathcal{B}_f) \leq M \text{ for all } f \in \partial H_d \tag{17}$$

$$\mathcal{E}_{E_e} \left(\bigvee_{f \in \partial e} \mathcal{B}'_f \right) - \mathcal{E}_{E_e} \left(\bigvee_{f \in \partial e} \mathcal{B}_f \right) \leq \varepsilon^2 \text{ for all } e \in H_d, E_e \in \mathcal{B}_e \tag{18}$$

$$\Delta_e(E_e | \bigvee_{f \in \partial e} \mathcal{B}'_f) \leq \frac{1}{F(M)} \text{ for all } e \in H_d, E_e \in \mathcal{B}_e \tag{19}$$

Remark 2.7. Lemma 2.6 provides a coarse low-order approximation $(\mathcal{B}_f)_{f \in \partial H_d}$ and a fine low-order approximation $(\mathcal{B}'_f)_{f \in \partial H_d}$ to the high-order σ -algebras $(\mathcal{B}_e)_{e \in H_d}$. The coarse approximation has bounded complexity, the fine approximation is close to the coarse approximation in an L^2 sense, and the high order σ -algebras are pseudorandom with respect to the fine approximation. The key point here is that the discrepancy control on the fine approximation given by (19) is superior to the complexity control on the coarse approximation given by (17) by an *arbitrary* growth function F . If one were to try to use a single approximation instead of a pair of coarse and fine approximations, it appears impossible to obtain such a crucial gain.

Proof We perform the following iteration.

- Step 0. Initialize $\mathcal{B}_f = \{\emptyset, V_J\}$ to be the trivial σ -algebra for all $f \in \partial H_d$, thus $\mathcal{B}_f$ has complexity 0 initially.
- Step 1. Set $M := \max(F(m), \sup_{f \in \partial H_d} \text{complex}(\mathcal{B}'_f))$, and $\delta := 1/F(M)$. We apply Lemma 2.5, and end up in either the randomness or structure half of the dichotomy. In either case we generate σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ for each $f \in \partial H_d$.

- Step 2. If we are in the randomness half of the dichotomy, we terminate the algorithm. Otherwise, if we are in the structure half of the dichotomy, we replace $\mathcal{B}_f$ with $\mathcal{B}'_f$ for each $f \in \partial H_d$, and return to Step 1.

Observe that every time we return from Step 2 to Step 1, the quantity

$$\sum_{e \in H_d} \sum_{E_e \in \mathcal{B}_e} \mathcal{E}_{E_e} \left(\bigvee_{f \in \partial e} \mathcal{B}_f \right)$$

increases by at least ε^2 . On the other hand, this quantity is non-negative and does not exceed $|H_d|2^{2^m} = O_{|J|,m}(1)$, thanks to (2). Thus this algorithm terminates after $O_{|J|,m,\varepsilon}(1)$ steps. By (15), we see that at each of these steps, the quantity M increases to be at most $M + O_{J,m,\varepsilon,F(M)}(1)$, while initially M is equal to $F(m)$. Thus at the end of the algorithm we have (16) as desired. The remaining claims (17), (18), (19) follow from construction (and (12), (13)). $\blacksquare$

Remark 2.8. Lemma 2.6 already implies the Szemerédi regularity lemma in its usual form (and with the usual tower-exponential bounds); see [28] for further discussion. The above lemma is also similar in spirit to the modern regularity lemmas that appear for instance in [17] (except for an issue of obtaining regularity at all orders less than d , which we shall address in Lemma 2.9 below). In such lemmas, the objective is not to obtain a partition for which the original graph or hypergraph is regular, but instead to obtain a partition for which a *modified* graph or hypergraph is *very* regular, where the modification consists of adding or subtracting a small number of edges. The analogue of such a modification in our context is the decomposition

$$1_{E_e} = F_{\text{regular}} + F_{\text{small}}$$

where

$$F_{\text{regular}} := \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}_f) + (1_{E_e} - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f))$$

and

$$F_{\text{small}} := \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}_f).$$

The function F_{small} is small thanks to (18) and (9). Now consider F_{regular} . On a typical atom of $\bigvee_{f \in \partial e} \mathcal{B}_f$, the first term is constant, and the second term is going to be very pseudorandom (have small correlation with sets of the form $\bigcap_{f \in \partial e} E_f$ for $E_f \in \mathcal{A}_f$) thanks to (19) and (8).

Lemma 2.6 regularizes the σ -algebras $\mathcal{B}_e$ on the d -uniform hypergraph H_d in terms of σ -algebras $\mathcal{B}_f, \mathcal{B}'_f$ on the $(d-1)$ -uniform hypergraph ∂H_d . However it does not regularize the σ -algebras on ∂H_d . This can be accomplished by one final iteration, which gives our final regularity lemma (which is essentially the same lemma⁷ as that in [11], [19], or [17]).

⁷In contrast, the earlier regularity lemmas of Chung [2] and Frankl-Rodl [4] are closer to Lemma 2.6, with ∂H_d generalized to $\partial^l H_d$ for any fixed l . The case $l = d-1$ in particular is essentially a routine generalization of the ordinary regularity lemma and appears to have been folklore for quite some time.

Lemma 2.9 (Full regularity lemma). *Let $V = (J, (V_j)_{j \in J}, d, H_d)$ be a hypergraph system, and define the j -uniform hypergraphs H_j for all $0 \leq j < d$ recursively backwards from $j = d$ by the formula $H_j := \partial H_{j+1}$. (In particular, if H_d is non-empty then we have $H_0 = \{\emptyset\}$.) For all $e \in H_d$ let $\mathcal{B}_e \subseteq \mathcal{A}_e$ be a σ -algebra, and suppose that we have the bound*

$$\text{complex}(\mathcal{B}_e) \leq M_d \text{ for all } e \in H_d$$

for some $M_d > 0$. Let F be a growth function. Then there exists numbers

$$M_d \leq F(M_d) \leq M_{d-1} \leq F(M_{d-1}) \leq \dots \leq M_0 \leq F(M_0) \leq O_{|J|, M_d, F}(1) \quad (20)$$

and for each $0 \leq j < d$ and $f \in H_j$ there exist σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$, such that we have the estimates

$$\text{complex}(\mathcal{B}_f) \leq M_j \text{ for all } 0 \leq j < d, f \in H_j \quad (21)$$

$$\mathcal{E}_{E_e} \left(\bigvee_{f \in \partial e} \mathcal{B}'_f \right) - \mathcal{E}_{E_e} \left(\bigvee_{f \in \partial e} \mathcal{B}_f \right) \leq \frac{1}{F(M_j)^2} \text{ for all } 1 \leq j \leq d, e \in H_j, E_e \in \mathcal{B}_e \quad (22)$$

$$\Delta_e(E_e | \bigvee_{f \in \partial e} \mathcal{B}'_f) \leq \frac{1}{F(M_0)} \text{ for all } 1 \leq j \leq d, e \in H_j, E_e \in \mathcal{B}_e. \quad (23)$$

Remark 2.10. At every order $0 \leq j \leq d$, Lemma 2.9 gives coarse and fine approximations $(\mathcal{B}_f)_{f \in H_{j-1}}$, $(\mathcal{B}'_f)_{f \in H_{j-1}}$ at the $(j-1)$ -uniform level to the σ -algebras $(\mathcal{B}'_e)_{e \in H_j}$ at the j -uniform level. As one goes down in order, the σ -algebras rapidly become more complex⁸ (though lower order, of course). However, the bounds in (22) and (23) will keep pace with this growth in complexity (see [17] for some related discussion concerning the desirability of having the constants grow along such a hierarchy). Indeed the bound (23) is extremely strong, as $F(M_0)$ dominates all the other quantities which appear in the above lemma; it is effectively as if the fine approximation was perfectly accurate (so that 1_{E_e} is approximable by $\mathbf{E}(1_{E_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f)$ with only negligible error). The main remaining difficulty when using this lemma is to exploit the estimate (22) measuring the gap between the coarse and fine approximations; one has to take some care here because the error bound $1/F(M_j)^2$ here safely exceeds the complexity⁹ of the higher-order objects $(\mathcal{B}_e)_{e \in H_j}$, but not that of the lower-order objects $(\mathcal{B}_e)_{e \in H_{j-1}}$.

Proof We induct on d (keeping J fixed); the implicit constants in (20) will change when one does this, but the induction will only run for at most $|J|$ steps and so this will not cause a difficulty. When $d = 0$ the claim is trivial (and the claim (21) has an enormous amount of room available!) so assume that $d \geq 1$ and the claim has already been proven for all smaller d . We will need a growth function F^{fast} to be chosen later; as the name suggests, this function will grow substantially faster than F , in particular we assume $F^{\text{fast}}(n) \geq F(n)$ for all n . Applying Lemma 2.6 with m equal to M_d , with ε equal to $1/F(M_d)$, and the growth function F^{fast} , we

⁸At the zeroth order $j = 0$, all σ -algebras have complexity zero, but this is a degenerate exception to the above general rule.

⁹We will only need to bound the complexity of the coarse algebras $\mathcal{B}_e$. Some (very weak) bounds on the complexity of the fine algebras $\mathcal{B}'_e$ are available but they seem to be useless for applications and so we have not stated them explicitly here.

can create σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ for all $f \in H_{d-1}$ and a quantity M_{d-1} such that

$$F(M_d) \leq F^{\text{fast}}(M_d) \leq M_{d-1} \leq O_{|J|, \varepsilon, M_d, F^{\text{fast}}}(1) = O_{|J|, M_d, F, F^{\text{fast}}}(1) \quad (24)$$

$$\text{complex}(\mathcal{B}_f) \leq M_{d-1} \text{ for all } f \in H_{d-1}$$

$$\begin{aligned} \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}'_e) - \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}_f) &\leq \frac{1}{F(M_d)^2} \text{ for all } e \in H_d, E_e \in \mathcal{B}_e \\ \Delta_e(E_e | \bigvee_{f \in \partial e} \mathcal{B}'_f) &\leq \frac{1}{F^{\text{fast}}(M_{d-1})} \text{ for all } e \in H_d, E_e \in \mathcal{B}_e. \end{aligned} \quad (25)$$

Now we apply the induction hypothesis with d replaced by $d-1$, and H_d replaced by H_{d-1} . This generates numbers

$$M_{d-1} \leq F(M_{d-1}) \leq \dots \leq M_0 \leq F(M_0) \leq O_{|J|, M_{d-1}, F}(1) \quad (26)$$

and for each $0 \leq j < d-1$ and $f \in H_j$ there exist σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$, such that we have the estimates

$$\begin{aligned} \text{complex}(\mathcal{B}_f) &\leq M_j \text{ for all } 0 \leq j < d-1, f \in H_j \\ \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}'_e) - \mathcal{E}_{E_e}(\bigvee_{f \in \partial e} \mathcal{B}_f) &\leq \frac{1}{F(M_j)^2} \text{ for all } 1 \leq j \leq d-1, e \in H_j, E_e \in \mathcal{B}_e \\ \Delta_e(E_e | \bigvee_{f \in \partial e} \mathcal{B}'_f) &\leq \frac{1}{F(M_0)} \text{ for all } 1 \leq j \leq d-1, e \in H_j, E_e \in \mathcal{B}_e. \end{aligned}$$

Comparing this with the conclusion of Lemma 2.9, we see that we can obtain all the claims we need except for (23) when $j = d$, as well as the final bound in (20). To obtain (23), we see from (25) that it would suffice to ensure that

$$F^{\text{fast}}(M_{d-1}) \geq F(M_0).$$

But since $F(M_0) = O_{|J|, M_{d-1}, F}(1)$, this can be achieved simply by choosing the growth function F^{fast} to be sufficiently large and rapidly increasing depending on F and $|J|$. By (26), (24), we then have

$$F(M_0) = O_{|J|, M_{d-1}, F}(1) = O_{|J|, M_d, F, F^{\text{fast}}}(1) = O_{|J|, M_d, F}(1)$$

and the claim (20) follows. $\blacksquare$

Remark 2.11. The dependence of constants here is quite terrible. Typically F will be an exponential function. In the graph case $d = 2$ one can take M_0 to be a tower of exponentials, whose height is bounded by some polynomial of $F(M_2)$; a modification of the arguments in [9] shows that this tower bound is essentially best possible. However, for $d = 3$, both M_0 and M_1 will be an *iterated* tower of exponentials of iterated height equal to a polynomial in $F(M_3)$, basically because of the need for F^{fast} to exceed the bounds one obtains from the $d = 2$ case. The situation of course gets even worse for larger values of d , though for any fixed d the bounds are still primitive recursive. As stated earlier, the complexity bounds for the fine approximations $\mathcal{B}'_f$ will be even worse than this, perhaps by yet another layer of iteration. Nevertheless, this regularity lemma is still sufficient for applications in which one is willing to have qualitative control only on the error terms (e.g. $o(1)$ type bounds) rather than quantitative control. (As we shall see in [29], obtaining

infinitely many constellations in the Gaussian primes will be one such application.) In view of recent results on effective bounds on Szemerédi-type theorems (see e.g. [10], [23]) it seems quite possible that these very rapid bounds, while perhaps necessary in order to have a regularity lemma, are not needed for the hypergraph removal lemma.

3. STATEMENT OF COUNTING LEMMA

As is customary in these arguments, the regularity lemma must be complemented with a counting lemma in order for it to be applicable to proving results such as Theorem 1.13. In the σ -algebra language, the setup is as follows. Suppose we start with σ -algebras $(\mathcal{B}_e)_{e \in H_d}$ as in the hypotheses of Lemma 2.9. Then, among other things, this lemma yields further σ -algebras $(\mathcal{B}_e)_{e \in H_j}$ for $0 \leq j < d$, each of which has some complexity bound. Combining all of these σ -algebras together, one obtains a somewhat large (but still bounded complexity) σ -algebra $\bigvee_{e \in H} \mathcal{B}_e$, where $H := \bigcup_{0 \leq j \leq d} H_j$. In particular, if E_e are sets in $\mathcal{B}_e$ for all $e \in H_d$, then $\bigcap_{e \in H_d} E_e$ is the union of atoms in $\bigvee_{e \in H} \mathcal{B}_e$. Here, of course, an atom of a σ -algebra $\mathcal{B}$ is a non-empty set in $\mathcal{B}$ of minimal size; since the ambient space V_J is finite, every point is contained in exactly one atom of $\mathcal{B}$.

Roughly speaking, the counting lemma we give below (Lemma 3.4) gives a formula for computing the probability of atoms in $\bigvee_{e \in H} \mathcal{B}_e$, or at least those atoms which are “good”. It can be informally described as follows. For each $e \in H$, let A_e be an atom of $\mathcal{B}_e$, thus $\bigcap_{e \in H} A_e$ will be an atom of $\bigvee_{e \in H} \mathcal{B}$ (if it is non-empty). The counting lemma then says that under most circumstances we have the approximate formula¹⁰

$$\mathbf{E}\left(\prod_{e \in H} 1_{A_e}\right) \approx \prod_{e \in H} \mathbf{E}(1_{A_e} \mid \bigcap_{f \in \partial e} A_f) \quad (27)$$

where we use $\mathbf{E}(f|A)$ to denote the conditional expectation

$$\mathbf{E}(f|A) := \frac{1}{|A|} \sum_{x \in A} f(x).$$

This can be viewed as an assertion that higher order atoms A_e are approximately independent of each other, conditioning on lower order atoms A_f , although a precise formulation of this heuristic is somewhat difficult to quantify. In particular, if we remove those “bad” atoms $\bigcap_{e \in H} A_e$ for which $\mathbf{E}(1_{A_e} \mid \bigcap_{f \in \partial e} A_f)$ is small for at least one $e \in H$, then all the remaining non-empty atoms will have fairly large size. Thus if the set $\bigcap_{e \in H} E_e$ has very small size, then after removing all the bad atoms we expect this set to in fact be empty. This is the strategy behind proving Theorem 1.13.

We now formalize the above discussion. We begin by describing the good atoms. Informally speaking, the good atoms are going to be those which are fairly large (at

¹⁰The reader may wish to interpret $\mathbf{E}(1_A)$ as being the “probability” of the “event” A , thus for instance $\mathbf{E}(\prod_{e \in H} 1_{A_e})$ is the probability of the joint event $\bigcap_{e \in H} A_e$. Similarly, many of the arguments in the sequel also have a strongly probabilistic flavour.

all orders) and also fairly regular (at all orders). This is consistent with previous experience with counting lemmas (say in the graph case), in which one must first throw away all cells of the partition which are too small (or have too few edges), as well as all pairs of cells for which the graph is irregular, before one can obtain a useful estimate for (say) the number of triangles in a graph.

Definition 3.1 (Good atoms). Let the notation, assumptions, and conclusions be as in Lemma 2.9, and let $H := \bigcup_{0 \leq j \leq d} H_j$. Let $\bigcap_{e \in H} A_e$ be a (possibly empty) atom of $\bigvee_{e \in H} \mathcal{B}_e$, where for each $e \in H$, A_e is an atom of $\mathcal{B}_e$. We say that this atom is *good* if for all $0 \leq j \leq d$ and $e \in H_j$ we have the largeness estimates

$$\mathbf{E}(1_{A_e} \prod_{f \in \partial e} 1_{A_f}) \geq \frac{1}{\log F(M_j)} \mathbf{E}(\prod_{f \in \partial e} 1_{A_f}) \quad (28)$$

as well as the regularity estimates

$$\mathbf{E} \left(\left| \mathbf{E}(1_{A_e} | \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{A_e} | \bigvee_{f \in \partial e} \mathcal{B}_f) \right|^2 \prod_{f \subsetneq e} 1_{A_f} \right) \leq \frac{1}{F(M_j)} \mathbf{E}(\prod_{f \subsetneq e} 1_{A_f}). \quad (29)$$

Remark 3.2. While the definition of a good atom allows for $\bigcap_{e \in H} A_e$ to be empty, the counting lemma we prove below will show that in fact good atoms are always non-empty (assuming F is sufficiently rapid). The reader should not take the logarithmic factor in (28) too seriously; the point is that $\log F(M_j)$ is smaller than any power of $F(M_j)$ but still much larger than any given function of M_j .

One can easily verify that most atoms are good in the following sense. For any $0 \leq j \leq d$, $e \in H_j$, and any atom A_e of $\mathcal{B}_e$, let B_{e,A_e} be the union of all the sets $\bigcap_{f \subsetneq e} A_f$ for which (28) or (29) fails. We remark for future reference that the set B_{e,A_e} lies in $\bigvee_{f \subsetneq e} \mathcal{B}_f$. Note also that if the atom $\bigcap_{e \in H} A_e$ is not good, then there exists $e \in H$ such that $\bigcap_{e' \in H} A_{e'} \subseteq A_e \cap B_{e,A_e}$.

Lemma 3.3 (Most atoms are good). *Let the notation, assumptions, and conclusions be as in Lemma 2.9 and Definition 3.1. For any $0 \leq j \leq d$, $e \in H_j$, and any atom A_e of $\mathcal{B}_e$, we have $\mathbf{E}(1_{A_e} 1_{B_{e,A_e}}) = O(1/\log F(M_j))$.*

Proof Consider the contribution to $\mathbf{E}(1_{A_e} 1_{B_{e,A_e}})$ from the case where (28) fails. This contribution is bounded by¹¹

$$\sum_{(A_f)_{f \in \partial e} \text{ atoms in } (\mathcal{B}_f)_{\partial e}: (28) \text{ fails}} \mathbf{E}(1_{A_e} \prod_{f \in \partial e} 1_{A_f})$$

which by failure of (28) is bounded by

$$\leq \sum_{(A_f)_{f \in \partial e} \text{ atoms in } (\mathcal{B}_f)_{\partial e}} \frac{1}{\log F(M_j)} \mathbf{E}(\prod_{f \in \partial e} 1_{A_f}) = \frac{1}{\log F(M_j)}.$$

Next, consider the contribution to $\mathbf{E}(1_{A_e} 1_{B_{e,A_e}})$ arising from the case when (29) fails. The total contribution of this case is

$$\sum_{(A_f)_{f \subsetneq e}: (29) \text{ fails}} \mathbf{E}(\prod_{f \subsetneq e} 1_{A_f})$$

¹¹Note that (28) depends only on those A_f for which $f \in \partial e$, as opposed to the larger class of events A_f for which $f \subsetneq e$.

which by failure of (29) is at most

$$F(M_j) \sum_{(A_f)_{f \subseteq e}} \mathbf{E} \left(\left| \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}_f) \right|^2 \prod_{f \subseteq e} 1_{A_f} \right)$$

which in turn is at most

$$F(M_j) \mathbf{E} \left(\left| \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}_f) \right|^2 \right).$$

But by (9), (22) we have

$$\mathbf{E} \left(\left| \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}_f) \right|^2 \right) \leq \frac{1}{F(M_j)^2}.$$

Combining all of these estimates, the claim follows. $\blacksquare$

We can now state the counting lemma; closely related results appear in the work of Gowers [10], Nagle, Rödl, and Schacht [15], and Rödl and Schacht [17].

Lemma 3.4 (Counting lemma). *Let the notation, assumptions, and conclusions be as in Lemma 2.9 and Definition 3.1, and let $H := \bigcup_{0 \leq j \leq d} H_j$. Let $\bigcap_{e \in H} A_e$ be a good atom of $\bigvee_{e \in H} \mathcal{B}_e$. Then, if the growth function F is sufficiently rapid depending on $|J|$, we have that $\bigcap_{e \in H} A_e$ is non-empty, and more precisely*

$$\mathbf{E} \left(\prod_{e \in H} 1_{A_e} \right) = (1 + o_{M_d \rightarrow \infty; |J|}(1)) \prod_{e \in H} \mathbf{E}(1_{A_e} \mid \bigcap_{f \in \partial e} A_f) + O_{|J|, M_0} \left(\frac{1}{F(M_0)} \right)$$

(compare with (27)).

This lemma is a little lengthy (though straightforward) to prove, and we defer it to the next section. Let us assume it for now, and conclude the proof of Theorem 1.13.

Proof [of Theorem 1.13 assuming Lemma 3.4] Let $V = (J, (V_j)_{j \in J}, d, H_d), (E_e)_{e \in H_d}$, δ be as in Theorem 1.13. We define H_j recursively for $0 \leq j < d$ by setting $H_j := \partial H_{j+1}$, and then set $H := \bigcup_{0 \leq j \leq d} H_j$. For any $e \in H_d$ we set $\mathcal{B}_e := \mathcal{B}(E_e)$, thus each $\mathcal{B}_e$ has complexity at most 1. Let $M_d \geq 1$ be a quantity to be chosen later, and let F be a growth function depending on $|J|$ (but not on δ) to be chosen later. We apply the regularity lemma, Lemma 2.9, to obtain quantities (20) and σ -algebras $\mathcal{B}_f \subseteq \mathcal{B}'_f \subseteq \mathcal{A}_f$ for all $f \in H$.

Suppose that $\bigcap_{e \in H} A_e$ is a (possibly empty) atom of $\bigvee_{e \in H} \mathcal{B}_e$ such that $A_e = E_e$ for $e \in H_d$. If this atom is good, then by the counting Lemma (Lemma 3.4) and Definition 3.1 we have

$$\mathbf{E}(1_{\bigcap_{e \in H} A_e}) = (1 + o_{M_d \rightarrow \infty; |J|}(1)) \prod_{0 \leq j \leq d} \prod_{e \in H_j} \frac{1}{F(M_j)^{1/10}} + O_{|J|, M_0} \left(\frac{1}{F(M_0)} \right),$$

if F is sufficiently rapid depending on $|J|$. Using (20), we thus see that (if M_d is sufficiently large depending on J)

$$\mathbf{E}(1_{\bigcap_{e \in H} A_e}) \geq c(|J|, M_d, F)$$

for some $c(|J|, M_d, F) > 0$. On the other hand, $\bigcap_{e \in H} A_e$ is contained in $\bigcap_{e \in H_d} E_e$, which has density at most δ by the hypothesis (3). Thus if δ is sufficiently small depending on $|J|, M_d, F$, we see that no atom $\bigcap_{e \in H} A_e$ with $A_e = E_e$ for $e \in H_d$ can possibly be good.

Now let B_{e, A_e} be as in Lemma 3.3. Let us define

$$E'_e := V_J \setminus (B_{e, E_e} \cup \bigcup_{f \subsetneq e} A_f \cap B_{f, A_f})$$

for all $e \in H_d$, where for brevity we adopt the convention that A_f is always understood to range over the atoms of $\mathcal{B}_f$. Then we observe that $E'_e \in \bigvee_{f \subsetneq e} \mathcal{B}_f$. The claims (6), (7) then follow from (21). Also, from Lemma 3.3, (21) we see that for any $e \in H_d$,

$$\begin{aligned} \mathbf{E}(1_{E_e \setminus E'_e}) &\leq \mathbf{E}(1_{E_e} 1_{B_{e, E_e}}) + \sum_{f \subsetneq e} \sum_{A_f} \mathbf{E}(1_{A_f} 1_{B_{f, A_f}}) \\ &\leq O(F(M_d)^{-1/10}) + \sum_{0 \leq j < d} \sum_{f \in H_j} \sum_{A_f} O(1/\log F(M_j)) \\ &\leq O(F(M_d)^{-1/10}) + \sum_{0 \leq j < d} \sum_{f \in H_j} O_{M_j}(1/\log F(M_j)) \\ &\leq \sup_{0 \leq j \leq d} O_{M_j, |J|}(1/\log F(M_j)). \end{aligned}$$

If one chooses F sufficiently rapidly growing (depending only on $|J|$), we conclude from (20) that we have

$$\mathbf{E}(1_{E_e \setminus E'_e}) = o_{M_d \rightarrow 0; |J|}(1).$$

By choosing M_d sufficiently large depending on $|J|$, and then letting δ be sufficiently small depending on M_d and $|J|$, we conclude (5).

The final thing to verify is (4). To see this, first observe that this set lies in $\bigvee_{f \in H \setminus H_d} \mathcal{B}_f$ and thus is the union of atoms of the form $\bigcap_{f \in H \setminus H_d} A_f$. Suppose for contradiction that $\bigcap_{e \in H_d} E'_e$ contains a non-empty atom of the form $\bigcap_{f \in H \setminus H_d} A_f$. Set $A_e := E_e$ for $e \in H_d$. By the preceding discussion we know that $\bigcap_{e \in H} A_e$ cannot be good, thus there exists an $f' \in H$ such that $\bigcap_{g \subsetneq f'} A_g$ lies in $B_{f', A_{f'}}$. From construction of H , there exists $e \in H_d$ which contains f' . But then by definition of E'_e , $\bigcap_{f \in H \setminus H_d} A_f$ cannot lie in E'_e , contradiction. Thus $\bigcap_{e \in H_d} E'_e$ is empty, which is (4), and Theorem 1.13 follows. $\blacksquare$

It remains to prove the counting lemma. This will be accomplished in the next section.

4. PROOF OF COUNTING LEMMA

We now prove Lemma 3.4. Fix a good collection $(A_e)_{e \in H}$ of atoms. We introduce the numbers $p_e \in \mathbf{R}$, the functions $b_e, c_e : V_J \rightarrow \mathbf{R}$, and the sets $A_{<e} \subseteq V_J$ for all $e \in H$ by the formulae

$$\begin{aligned} p_e &:= \mathbf{E}(1_{A_e} \mid \bigcap_{f \in \partial e} A_f) \\ b_e &:= \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}'_f) - \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}_f) \\ c_e &:= 1_{A_e} - \mathbf{E}(1_{A_e} \mid \bigvee_{f \in \partial e} \mathcal{B}'_f) \\ A_{<e} &:= \bigcap_{f \subsetneq e} A_f. \end{aligned}$$

Note that we have not yet shown that $\bigcap_{f \in \partial e} A_f$ is non-empty; for now, let us just assign an arbitrary value to p_e (e.g. $p_e = 1$) when $\bigcap_{f \in \partial e} A_f$ is empty. We thus have the decomposition

$$1_{A_e} = p_e + b_e + c_e \quad (30)$$

on the set $\bigcap_{f \in \partial e} A_f$. One should think of the constant p_e as the main term, and the other two terms as error terms. The c_e error term will be very easy to handle, whereas the b_e error term will cause somewhat more difficulty. Since $(A_e)_{e \in H}$ is good, we have the estimates

$$p_e \geq 1/\log F(M_j) \text{ for all } 0 \leq j \leq d \text{ and } e \in H_j \quad (31)$$

and

$$\mathbf{E}(|b_e|^2 1_{A_{<e}}) \leq F(M_j)^{-1} \mathbf{E}(1_{A_{<e}}) \text{ for all } 0 \leq j \leq d \text{ and } e \in H_j. \quad (32)$$

From (23) and (8), we also have

$$|\mathbf{E}(c_e \prod_{f \in \partial e} 1_{E_f})| \leq \frac{1}{F(M_0)} \text{ whenever } E_f \in \mathcal{A}_f \text{ for } f \in \partial e. \quad (33)$$

Our objective is to use the above estimates (30), (31), (32), (33) to conclude that

$$\mathbf{E}(\prod_{e \in H} 1_{A_e}) = (1 + o_{M_d \rightarrow \infty; |J|}(1)) \prod_{e \in H} p_e + O_{|J|, M_0}(\frac{1}{F(M_0)}). \quad (34)$$

This will be achieved by several applications of the Cauchy-Schwarz and triangle inequalities. However, there is a certain amount of notational burden in order to keep track of the expressions in the successive applications of these inequalities. It will be convenient to return to the original sets $(V_j)_{j \in J}$. We can identify $A_e \in \mathcal{B}_e$ as a subset $\overline{A_e}$ of $V_e = \prod_{j \in e} V_j$, and similarly we can view the $\mathcal{A}_e$ -measurable b_e

and c_e as functions $\overline{b}_e$ and $\overline{c}_e$ on V_e . One can then write (34) in the form

$$\begin{aligned} & \frac{1}{\prod_{j \in J} |V_j|} \sum_{(v_j)_{j \in J} \in \prod_{j \in J} V_j} \prod_{e \in H} 1_{\overline{A}_e}((v_j)_{j \in e}) \\ &= (1 + o_{M_d \rightarrow \infty; |J|}(1)) \prod_{e \in H} p_e + O_{|J|, M_0} \left(\frac{1}{F(M_0)} \right). \end{aligned} \quad (35)$$

For inductive purposes we will need to generalize¹² this formula.

Definition 4.1 (Hypergraph bundle). A *hypergraph bundle* over H is a hypergraph $G \subseteq 2^K$ on a finite set K , together with a map $\pi : K \rightarrow J$ (which we call the *projection map* of the bundle), which is a hypergraph homomorphism (i.e. for each edge $g \in G$, the function π is injective on g and $\pi(g) \in H$). For any $g \subseteq K$, we write V_g for the product set $V_g := \prod_{k \in g} V_{\pi(k)}$. We say that the bundle is *closed under set inclusion* if whenever $g \in G$ and $g' \subset g$, we have $g' \in G$.

Remark 4.2. From a probabilistic viewpoint, the probability space V_J corresponds to sampling one vertex independently from each of the vertex classes V_j of V_J , whereas the more general spaces V_g correspond to the possibility of sampling more than one vertex independently from each of the vertex classes.

The generalization of the formula (35) is then

Lemma 4.3 (Generalized counting lemma). *Let $G \subseteq 2^K$ be a hypergraph bundle over H which is closed under set inclusion, with projection map $\pi : K \rightarrow J$. Let $d' := \sup_{g \in G} |g|$ be the order of G . Then, if F is sufficiently rapidly growing depending on d' , $|J|$ and $|K|$, we have*

$$\begin{aligned} & \frac{1}{|V_K|} \sum_{(v_k)_{k \in K} \in V_K} \prod_{g \in G} 1_{\overline{A}_{\pi(g)}}((v_k)_{k \in g}) \\ &= (1 + o_{M_d \rightarrow \infty; d', |J|, |K|}(1)) \prod_{g \in G} p_{\pi(g)} + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right). \end{aligned} \quad (36)$$

Observe that (35) is the special case of this lemma with $G = H$ (and $K = J$, and π being the identity map); note from construction of H that H is automatically closed under set inclusion.

Proof We shall use a double induction. Firstly, we shall induct on the order d' of the bundle G . When $d' = 0$ the claim is vacuously true (the left-hand side and the main term of the right-hand side is equal to 1), so we may assume $d' \geq 1$ and the claim has already been proven for $d' - 1$ and for all choices of hypergraph bundle $G \subseteq 2^K$ which are closed under set inclusion.

¹²The basic problem is that we need the Cauchy-Schwarz inequality to eliminate each of the $\overline{b}_e$ factors in turn (using (32)), but each time we apply this inequality we essentially double the number of free variables that one has to sum or average over. In particular, one ends up sampling more than one point from each vertex class V_j , which forces us to leave the probabilistic framework that has been so convenient for us in preceding sections and return to a combinatorial framework. One could stay in the probabilistic framework using the machinery of tensor products (and conditional tensor products) of probability spaces, but this would introduce even more excessive notation into an already notation-heavy argument and would probably not be helpful to the reader.

Next, we fix K and induct on the quantity $r := |\{g \in G : |g| = d'\}|$, which is a positive integer between 1 and $2^{|K|}$. We thus assume that the claim has already been proven for all smaller values of r (note that for $r = 0$ this follows from the previous induction hypothesis). The constants may change as we progress in this induction, but since the number of steps in the induction cannot exceed $2^{|K|}$, this will not be a concern.

Let $g_0 \in G$ be such that $|g_0| = d'$. We use (30) to split

$$\prod_{g \in G} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) = \left[\prod_{g \in G \setminus \{g_0\}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] (p_{\pi(g_0)} + \overline{b_{\pi(g_0)}}((v_k)_{k \in g_0}) + \overline{c_{\pi(g_0)}}((v_k)_{k \in g_0}))$$

and consider the contribution of the three terms separately.

We first consider the contribution of the $p_{\pi(g)}$ term, which is the main term. Applying the second induction hypothesis to $G \setminus \{g_0\}$ we see from (36) that

$$\begin{aligned} & \frac{1}{|V_K|} \sum_{(v_k)_{k \in K} \in V_K} \prod_{g \in G \setminus \{g_0\}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \\ &= (1 + o_{M_d \rightarrow \infty; d', |J|, |K|}(1)) \prod_{g \in G \setminus \{g_0\}} p_{\pi(g)} + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right). \end{aligned}$$

Multiplying this by the quantity $p_{\pi(g_0)}$, which is between 0 and 1, we see that the contribution of this term to (36) is

$$(1 + o_{M_d \rightarrow \infty; d', |J|, |K|}(1)) \prod_{g \in G} p_{\pi(g)} + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right). \quad (37)$$

Next we consider the $\overline{c_{\pi(g_0)}}$ term. We split $V_K = V_{g_0} \times V_{K \setminus g_0}$. Let us temporarily freeze the values of v_k for $k \in K \setminus g_0$, and consider the expression

$$\frac{1}{|V_{g_0}|} \sum_{(v_k)_{k \in g_0} \in V_{g_0}} \left[\prod_{g \in G \setminus \{g_0\}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] \overline{c_{\pi(g_0)}}((v_k)_{k \in g_0}).$$

Observe that for each $g \in G \setminus \{g_0\}$, we have $g \neq g_0$ and $|g| \leq d' = |g_0|$. Thus $g \cap g_0$ is a proper subset of g_0 , and thus there exists an element of ∂g_0 which contains $g \cap g_0$. Thus one can rewrite the product $\prod_{g \in G \setminus \{g_0\}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g})$ in the form

$$\prod_{f \in \partial g_0} 1_{E_f}((v_k)_{k \in \pi(f)})$$

for some sets $E_f \subseteq V_f$ whose exact form is not important here (we allow the E_f to depend on the frozen v_k). Applying (33), we conclude that

$$\left| \frac{1}{|V_{g_0}|} \sum_{(v_k)_{k \in g_0} \in V_{g_0}} \left[\prod_{g \in G \setminus \{g_0\}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] \overline{c_{\pi(g_0)}}((v_k)_{k \in g_0}) \right| \leq 1/F(M_0).$$

Averaging this over all choices of the frozen variables $k \in K \setminus g_0$, we conclude that the contribution of this term to (36) is at most

$$1/F(M_0). \quad (38)$$

Finally we consider the contribution of the $\overline{b_{\pi(g_0)}}$ term, which is the most difficult from a notational viewpoint to handle, mainly because of the need to invoke the Cauchy-Schwarz inequality. We expand this contribution as

$$\frac{1}{|V_K|} \sum_{(v_k)_{k \in K} \in V_K} \left[\prod_{g \in G \setminus \{g_0\}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] \overline{b_{\pi(g_0)}}((v_k)_{k \in g_0}).$$

We take absolute values and discard¹³ the bounded factors $1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g})$ with $|g| = d'$, to estimate this expression by

$$O \left(\frac{1}{|V_K|} \sum_{(v_k)_{k \in K} \in V_K} \left[\prod_{g \in G_{\subsetneq g_0} \cup G'} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] |\overline{b_{\pi(g_0)}}((v_k)_{k \in g_0})| \right)$$

where $G_{\subsetneq g_0} := \{g : g \subsetneq g_0\}$ and $G' := \{g \in G \setminus G_{\subsetneq g_0} : |g| \leq d' - 1\}$. We factorize this as

$$O \left(\frac{1}{|V_{g_0}|} \sum_{(v_k)_{k \in g_0} \in V_{g_0}} \left[\prod_{g \in G_{\subsetneq g_0}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] |\overline{b_{\pi(g_0)}}((v_k)_{k \in g_0})| \right. \\ \left. \left[\frac{1}{|V_{K \setminus g_0}|} \sum_{(v_k)_{k \in K \setminus g_0} \in V_{K \setminus g_0}} \prod_{g \in G'} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] \right). \quad (39)$$

On the other hand, from (32) we have

$$\frac{1}{|V_{g_0}|} \sum_{(v_k)_{k \in g_0} \in V_{g_0}} \left[\prod_{g \in G_{\subsetneq g_0}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k \in g}) \right] |\overline{b_{\pi(g_0)}}((v_k)_{k \in g_0})|^2 \leq \frac{1}{F(M_{d'})} \mathbf{E}(1_{\overline{A_{\pi(g_0)}}}),$$

and hence by Cauchy-Schwarz we can estimate (39) by

¹³This discarding step is important as it lowers the total order of the expression being computed, which compensates for a certain doubling of the hypergraph bundle which shall occur shortly when we apply Cauchy-Schwarz. We can get away with this step because the smallness of $b_{\pi(g_0)}$, as given by (32), safely dominates any loss we absorb by discarding these high-order factors.

$$O\left(F(M_{d'})^{-1/2}\mathbf{E}(1_{\overline{A_{<\pi(g_0)}}})^{1/2}\left(\frac{1}{|V_{g_0}|}\sum_{(v_k)_{k\in g_0}\in V_{g_0}}\left[\prod_{g\in G_{\subseteq g_0}}1_{\overline{A_{\pi(g)}}}((v_k)_{k\in g})\right]\right.\right. \\ \left.\left.\left[\frac{1}{|V_{K\setminus g_0}|}\sum_{(v_k)_{k\in K\setminus g_0}\in V_{K\setminus g_0}}\prod_{g\in G'}1_{\overline{A_{\pi(g)}}}((v_k)_{k\in g})\right]^2\right)^{1/2}\right) \quad (40)$$

From the first induction hypothesis we have

$$\mathbf{E}(1_{\overline{A_{<\pi(g_0)}}}) = (1 + o_{M_d\rightarrow\infty;d',|J|}(1)) \prod_{g\in G_{\subseteq g_0}} p_{\pi(g)} + O_{d',|J|,M_0}\left(\frac{1}{F(M_0)}\right)$$

and thus

$$\mathbf{E}(1_{\overline{A_{<\pi(g_0)}}}) = O_{M_d,d',|J|}\left(\prod_{g\in G_{\subseteq g_0}} p_{\pi(g)}\right) + O_{d',|J|,M_0}\left(\frac{1}{F(M_0)}\right). \quad (41)$$

Now we estimate the expression in parentheses in (40). As we shall see, this expression can be rewritten in a form which can be handled by the induction hypothesis, but with the hypergraph bundle G replaced by a hypergraph of approximately twice the size (roughly speaking, we throw away all edges of top order d' , and double all the remaining edges that are not contained in $G_{\subseteq g_0}$). It is this doubling which forces us to work with a generalized counting lemma¹⁴ rather than the original counting lemma.

Let $\tilde{K} = K \oplus_{g_0} K$ be the set $K \times \{0, 1\}$, with the elements $(k, 0)$ and $(k, 1)$ identified for all $k \in g_0$. There is an obvious projection $\phi : \tilde{K} \mapsto K$, and hence a map $\pi \circ \phi : \tilde{K} \rightarrow H$. On $\tilde{K}$ we also place a hypergraph bundle $\tilde{G}$, defined as the set $\{g \times \{i\} : g \in G_{\subseteq g_0} \cup G', i \in 1, 2\}$; note that $g \times \{0\}$ and $g \times \{1\}$ will be identified when $g \in G_{\subseteq g_0}$. From the definitions we observe that

$$\frac{1}{|V_{g_0}|} \sum_{(v_k)_{k\in g_0}\in V_{g_0}} \left[\prod_{g\in G_{\subseteq g_0}} 1_{\overline{A_{\pi(g)}}}((v_k)_{k\in g}) \right] \left[\frac{1}{|V_{K\setminus g_0}|} \sum_{(v_k)_{k\in K\setminus g_0}\in V_{K\setminus g_0}} \prod_{g\in G'} 1_{\overline{A_{\pi(g)}}}((v_k)_{k\in g}) \right]^2 \\ = \frac{1}{|V_{\tilde{K}}|} \sum_{(v_{\tilde{k}})_{\tilde{k}\in \tilde{K}}\in V_{\tilde{K}}} \prod_{\tilde{g}\in \tilde{G}} 1_{\overline{A_{\pi\circ\phi(\tilde{g})}}}((v_{\tilde{k}})_{\tilde{k}\in \tilde{g}}).$$

Applying the first induction hypothesis, we can write this expression as

$$(1 + o_{M_d\rightarrow\infty;d',|J|,|K|}(1)) \prod_{\tilde{g}\in \tilde{G}} p_{\pi\circ\phi(\tilde{g})} + O_{d',|J|,|K|,M_0}\left(\frac{1}{F(M_0)}\right). \quad (42)$$

¹⁴There is a possible alternate approach which avoids the Cauchy-Schwarz inequality, and hence the need to work with hypergraph bundles. One can attempt to use the lower-order induction hypothesis to show some uniform distribution properties concerning the intersections of the lower-order atoms with each other, in order that the contribution of the b_{g_0} error be shown to be negligible. A model example of such a statement, in the graph setting, would be the assertion that in an ε -regular graph H , the number of copies of a fixed small graph G in H , with one edge specified to be (x, y) , is usually close to a fixed quantity independent of x and y , except for a small number of exceptional pairs (x, y) . We will not pursue such an alternate approach here.

By the definition of $\tilde{G}$, we can write

$$\prod_{\tilde{g} \in \tilde{G}} p_{\pi \circ \phi(\tilde{g})} = \prod_{g \in G \subsetneq g_0} p_{\pi(g)} \times \left[\prod_{g \in G'} p_{\pi(g)} \right]^2$$

and thus by (31) and (20) we can rewrite (42) as

$$O_{M_d, d', |J|, |K|} \left(\prod_{g \in G \subsetneq g_0} p_{\pi(g)} \left[\prod_{g \in G'} p_{\pi(g)} \right]^2 \right) + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right).$$

Inserting this and (41) back into (40), we can estimate (40) by

$$O_{M_d, d', |J|, |K|} \left(F(M_{d'})^{-1/2} \prod_{g \in G \subsetneq g_0} p_{\pi(g)} \prod_{g \in G'} p_{\pi(g)} \right) + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right).$$

Re-inserting those elements g of G for which $|g| = d'$ using (31), we can estimate this by

$$O_{M_d, d', |J|, |K|} (F(M_{d'})^{-1/4} \prod_{g \in G} p_{\pi(g)}) + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right)$$

(for instance). By choosing F sufficiently rapid depending on d' , $|J|$, $|K|$, we can write this as

$$O_{M_d \rightarrow \infty; d', |J|, |K|} \left(\prod_{g \in G} p_{\pi(g)} \right) + O_{d', |J|, |K|, M_0} \left(\frac{1}{F(M_0)} \right).$$

Combining this with the bounds (37), (38) we obtain (36), which closes the induction. This completes the proof of Lemma 4.3, and hence Lemma 3.4. $\blacksquare$

REFERENCES

- [1] M. Ajtai, E. Szemerédi, *Sets of lattice points that form no squares*, Studia Scientiarum Mathematicarum Hungarica. **9** (1974), 9–11.
- [2] F. Chung, *Regularity lemmas for hypergraphs and quasi-randomness*, Random Struct. Alg. **2** (1991), 241–252.
- [3] P. Erdős, P. Frankl, V. Rödl, *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin. **2** (1986), no. 2, 113–121.
- [4] P. Frankl, V. Rödl, *The uniformity lemma for hypergraphs*, Graphs Combinat. **8**(4) (1992), 309–312.
- [5] P. Frankl, V. Rödl, *Extremal problems on set systems*, Random Struct. Algorithms **20** (2002), no. 2, 131–164.
- [6] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
- [7] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291.
- [8] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Analyse Math. **45** (1985), 117–168.
- [9] T. Gowers, *Lower bounds of tower type for Szemerédi’s uniformity lemma*, Geom. Func. Anal. **7** (1997), 322–337.
- [10] T. Gowers, *A new proof of Szemerédi’s theorem*, GAFA **11** (2001), 465–588.
- [11] T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, preprint.
- [12] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, preprint.

- [13] Y. Kohayakawa, V. Rödl, J. Skokan, *Hypergraphs, quasi-randomness, and conditions for regularity*, J. Combin. Theory Ser. A **97** (2002), no. 2, 307–352.
- [14] J. Komlós, M. Simonovits, *Szemerédi’s regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), 295–352, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.
- [15] B. Nagle, V. Rödl, M. Schacht, *The counting lemma for regular k -uniform hypergraphs*, to appear, Random Structures and Algorithms.
- [16] V. Rödl, *Some developments in Ramsey theory*, Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990), 1455–1466, Math. Soc. Japan, Tokyo, 1991.
- [17] V. Rödl, M. Schacht, *Regular partitions of hypergraphs*, preprint.
- [18] V. Rödl, M. Schacht, E. Tengan, N. Tokushige, *Density theorems and extremal hypergraph problems*, preprint.
- [19] V. Rödl, J. Skokan, *Regularity lemma for k -uniform hypergraphs*, to appear, Random Structures and Algorithms.
- [20] V. Rödl, J. Skokan, *Applications of the regularity lemma for uniform hypergraphs*, preprint.
- [21] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245–252.
- [22] I. Ruzsa, E. Szemerédi, *Triple systems with no six points carrying three triangles*, Colloq. Math. Soc. J. Bolyai **18** (1978), 939–945.
- [23] I.D. Shkredov, *On a problem of Gowers*, preprint.
- [24] J. Solymosi, *Note on a generalization of Roth’s theorem*, Discrete and computational geometry, 825–827, Algorithms Combin. **25**, Springer Verlag, 2003.
- [25] J. Solymosi, *A note on a question of Erdős and Graham*, Combinatorics, Probability and Computing **13** (2004), 263–267.
- [26] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.
- [27] E. Szemerédi, *Regular partitions of graphs*, in “Problèmes Combinatoires et Théorie des Graphes, Proc. Colloque Inter. CNRS,” (Bermond, Fournier, Las Vergnas, Sotteau, eds.), CNRS Paris, 1978, 399–401.
- [28] T. Tao, *Szemerédi’s regularity lemma revisited*, preprint.
- [29] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, preprint.
- [30] B.L. Van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.
- [31] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. **39** (1959), 358–360.
- [32] V. Vu, *On a question of Gowers*, Ann. of Combinatorics, **6** (2002), 229–233.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: tao@math.ucla.edu