

A RELATIVE SZEMERÉDI THEOREM

DAVID CONLON, JACOB FOX, AND YUFEI ZHAO

ABSTRACT. The celebrated Green-Tao theorem states that there are arbitrarily long arithmetic progressions in the primes. One of the main ingredients in their proof is a relative Szemerédi theorem which says that any subset of a pseudorandom set of integers of positive relative density contains long arithmetic progressions.

In this paper, we give a simple proof of a strengthening of the relative Szemerédi theorem, showing that a much weaker pseudorandomness condition is sufficient. Our strengthened version can be applied to give the first relative Szemerédi theorem for k -term arithmetic progressions in pseudorandom subsets of \mathbb{Z}_N of density N^{-c_k} .

The key component in our proof is an extension of the regularity method to sparse pseudorandom hypergraphs, which we believe to be interesting in its own right. From this we derive a relative extension of the hypergraph removal lemma. This is a strengthening of an earlier theorem used by Tao in his proof that the Gaussian primes contain arbitrarily shaped constellations and, by standard arguments, allows us to deduce the relative Szemerédi theorem.

1. INTRODUCTION

The Green-Tao theorem [22] states that the primes contain arbitrarily long arithmetic progressions. This result, along with their subsequent work [23] on determining the asymptotics for the number of prime k -tuples in arithmetic progression, constitutes one of the great breakthroughs in 21st century mathematics.

The proof of the Green-Tao theorem has two key steps. The first step, which Green and Tao refer to as the “main new ingredient” of their proof, is to establish a relative Szemerédi theorem. Szemerédi’s theorem [43] states that any dense subset of the integers contains arbitrarily long arithmetic progressions. More formally, we have the following theorem, which easily implies an equivalent statement in the set $[N] := \{1, 2, \dots, N\}$.

Theorem 1.1 (Szemerédi’s theorem). *For every natural number $k \geq 3$ and every $\delta > 0$, as long as N is sufficiently large, any subset of \mathbb{Z}_N of density at least δ contains an arithmetic progression of length k .*

A relative Szemerédi theorem is a similar statement where the ground set is no longer the set \mathbb{Z}_N but rather a sparse pseudorandom subset of \mathbb{Z}_N .

The second step in their proof is to show that the primes are a dense subset of a pseudorandom set of “almost primes”, sufficiently pseudorandom that the relative Szemerédi theorem holds. Then, since the primes are a dense subset of this pseudorandom set, an application of the relative Szemerédi theorem implies that the primes contain arbitrarily long arithmetic progressions. This part of the proof makes use of some ideas from the work of Goldston and Yıldırım [16] (and was subsequently simplified in [44]).

In the work of Green and Tao, the pseudorandomness conditions on the ground set are known as the linear forms condition and the correlation condition. Roughly speaking, both of these conditions say that, in terms of the number of solutions to certain linear systems of equations, the set behaves

The first author was supported by a Royal Society University Research Fellowship, the second author was supported by a Simons Fellowship, NSF grant DMS-1069197, by an Alfred P. Sloan Fellowship, and by an MIT NEC Corporation Fund Award, and the third author was supported by a Microsoft Research PhD Fellowship.

like a random set of the same density. A natural question is whether these pseudorandomness conditions can be weakened. We address this question by giving a simple proof for a strengthening of the relative Szemerédi theorem, showing that a weak linear forms condition is sufficient for the theorem to hold.

This improvement has two aspects. We remove the correlation condition entirely but we also reduce the set of linear forms for which the correct count is needed. In particular, we remove those corresponding to the dual function condition, a pointwise boundedness condition stated explicitly by Tao [45] in his work on constellations in the Gaussian primes but also used implicitly in [22].

To state the main theorem, we will assume the definition of the k -linear forms condition. The formal definition, which may be found in Section 2 below, is stated for measures rather than sets but we will ignore this relatively minor distinction here, reserving a more complete discussion of our terminology for there.

Theorem 1.2 (Relative Szemerédi theorem). *For every natural number $k \geq 3$ and every $\delta > 0$, if $S \subset \mathbb{Z}_N$ satisfies the k -linear forms condition and N is sufficiently large, then any subset of S of relative density at least δ contains an arithmetic progression of length k .*

One of the immediate advantages of this theorem is that it simplifies the proof of the Green-Tao theorem. In addition to giving a simple proof of the relative Szemerédi theorem, it removes the need for the number-theoretic estimates involved in establishing the correlation condition for the almost primes. A further advantage is that, by removing the correlation condition, the relative Szemerédi theorem now applies to pseudorandom subsets of \mathbb{Z}_N of density N^{-c_k} . With the correlation condition, one could only hope for such a theorem down to densities of the form $N^{-o(1)}$.

While the relative Szemerédi theorem is the main result of this paper, the main advance is an approach to regularity in sparse pseudorandom hypergraphs. This allows us to prove analogues of several well-known combinatorial theorems relative to sparse pseudorandom hypergraphs. In particular, we prove a sparse analogue of the hypergraph removal lemma. It is from this that we derive our relative Szemerédi theorem. As always, applying the regularity method has two steps, a regularity lemma and a counting lemma. We provide novel approaches to both.

A counting lemma for subgraphs of sparse pseudorandom graphs was already proved by the authors in [7]. In this paper, we simplify and streamline the approach taken there in order to prove a counting lemma for subgraphs of sparse pseudorandom hypergraphs. This result is the key technical step in our proof and, perhaps, the main contribution of this paper. Apart from the obvious difficulties in passing from graphs to hypergraphs, the crucial difference between this paper and [7] is in the type of pseudorandomness considered. For graphs, we have a long-established notion of pseudorandomness known as jumbledness. The greater part of [7] is then concerned with optimizing the jumbledness condition which is necessary for counting a particular graph H . For hypergraphs, we use an analogue of the linear forms condition first considered by Tao [45]. It says that our hypergraph is pseudorandom enough for counting H within subgraphs if it contains asymptotically the correct count for the 2-blow-up of H and all its subgraphs.

We also use an alternative approach to regularity in sparse hypergraphs. While it would be natural to use a sparse hypergraph regularity lemma (and, following our approach in [7], this was how we initially proceeded), it suffices to use a weak sparse hypergraph regularity lemma which is an extension of the weak regularity lemma of Frieze and Kannan [14]. This is also closely related to the transference theorem used by Green and Tao (see, for example, [19] or [33, 49], where it is also referred to as the dense model theorem).

With both a regularity lemma and a counting lemma in place, it is then a straightforward matter to prove a relative extension of the famous hypergraph removal lemma [18, 32, 34, 35, 46]. Such a theorem was first derived by Tao [45] in his work on constellations in the Gaussian primes but, like the Green-Tao relative Szemerédi theorem, needs both a correlation condition and a dual

function condition.¹ Our approach removes these conditions. The final step in the proof of the relative Szemerédi theorem is then a standard reduction used to derive Szemerédi's theorem from the hypergraph removal lemma. The details of this reduction already appear in [45] but we include them here for completeness. In fact, the paper is self-contained apart from assuming the hypergraph removal lemma.

In Section 2, we state our results including the relative Szemerédi theorem and the removal, regularity, and counting lemmas. In Section 3, we deduce the relative multidimensional Szemerédi theorem from our relative hypergraph removal lemma. In Section 4, we prove the removal lemma from the regularity and counting lemmas. We will prove our weak sparse hypergraph regularity lemma in Section 5 and the associated counting lemma in Section 6. We conclude, in Section 7, with some remarks.

2. DEFINITIONS AND RESULTS

Notation. *Dependence on N .* We consider functions $\nu = \nu^{(N)}$, where N (usually suppressed) is assumed to be some large integer. We write $o(1)$ for a quantity that tends to zero as $N \rightarrow \infty$.

Expectation. We write $\mathbb{E}[f(x_1, x_2, \dots) | x_1 \in A_1, x_2 \in A_2, \dots]$ for the expectation of $f(x_1, x_2, \dots)$ when each x_i is chosen uniformly and independently at random from A_i .

2.1. A relative Szemerédi theorem. Here is an equivalent weighted version of Szemerédi's theorem as formulated, for example, in [22, Prop. 2.3].

Theorem 2.1 (Szemerédi's theorem, weighted version). *For every $k \geq 3$ and $\delta > 0$, there exists $c > 0$ such that for N sufficiently large and any nonnegative function $f: \mathbb{Z}_N \rightarrow [0, 1]$ satisfying $\mathbb{E}[f] \geq \delta$,*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d) | x, d \in \mathbb{Z}_N] \geq c. \quad (1)$$

A relative Szemerédi theorem would instead ask for the nonnegative function f to be bounded above by a measure ν instead of the constant function f . For us, a measure will be any nonnegative function on \mathbb{Z}_N . We do not explicitly assume the additional condition that

$$\mathbb{E}[\nu(x) | x \in \mathbb{Z}_N] = 1 + o(1),$$

but this property follows from the linear forms condition that we will now assume. Such measures are more general than subsets, as any subset $S \subseteq \mathbb{Z}_N$ (e.g., in Theorem 1.2) can be thought of as a measure on \mathbb{Z}_N taking value $N/|S|$ on S and 0 elsewhere. The dense case, as in Theorem 2.1, corresponds to taking $\nu = 1$. Our notion of pseudorandomness for measures ν on \mathbb{Z}_N is now as follows.

Definition 2.2 (Linear forms condition). A nonnegative function $\nu = \nu^{(N)}: \mathbb{Z}_N \rightarrow \mathbb{R}_{\geq 0}$ is said to obey the *k -linear forms condition* if one has

$$\mathbb{E}\left[\prod_{j=1}^k \prod_{\omega \in \{0,1\}^{[k] \setminus \{j\}}} \nu\left(\sum_{i=1}^k (i-j)x_i^{(\omega_i)}\right)^{n_{j,\omega}} \middle| x_1^{(0)}, x_1^{(1)}, \dots, x_k^{(0)}, x_k^{(1)} \in \mathbb{Z}_N\right] = 1 + o(1) \quad (2)$$

for any choice of exponents $n_{j,\omega} \in \{0, 1\}$.

Example 2.3. For $k = 3$, condition (2) says that

$$\begin{aligned} &\mathbb{E}[\nu(y+2z)\nu(y'+2z)\nu(y+2z')\nu(y'+2z')\nu(-x+z)\nu(-x'+z)\nu(-x+z')\nu(-x'+z') \\ &\quad \nu(-2x-y)\nu(-2x'-y)\nu(-2x-y')\nu(-2x'-y') | x, x', y, y', z, z' \in \mathbb{Z}_N] = 1 + o(1) \end{aligned}$$

and similar conditions hold if one or more of the twelve ν factors in the expectation are erased.

¹The problem of relative hypergraph removal was also recently considered by Towsner [48].

Our linear forms condition is much weaker than that used in Green and Tao [22]. In particular, Green and Tao need to assume that pointwise estimates such as

$$\mathbb{E}[\nu(a+x)\nu(a+y)\nu(a+x+y)|x, y \in \mathbb{Z}_N] = 1 + o(1)$$

hold uniformly over all $a \in \mathbb{Z}_N$. Such linear forms do not arise in our proof. Moreover, to prove their relative Szemerédi theorem, Green and Tao need to assume a further pseudorandomness condition, which they call the correlation condition. This condition also doesn't arise in our proofs. Indeed, we prove that a relative Szemerédi theorem holds given only the linear forms condition defined above.

Theorem 2.4 (Relative Szemerédi theorem). *For every $k \geq 3$ and $\delta > 0$, there exists $c > 0$ such that if $\nu: \mathbb{Z}_N \rightarrow \mathbb{R}_{\geq 0}$ satisfies the k -linear forms condition, N is sufficiently large, and $f: \mathbb{Z}_N \rightarrow \mathbb{R}_{\geq 0}$ satisfies $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}[f] \geq \delta$, then*

$$\mathbb{E}[f(x)f(x+d)f(x+2d)\cdots f(x+(k-1)d)|x, d \in \mathbb{Z}_N] \geq c. \quad (3)$$

We note that both here and in Theorem 1.2, the phrase “ N is sufficiently large” indicates not only a dependency on δ and k as in the usual version of Szemerédi's theorem but also a dependency on the $o(1)$ term in the linear forms condition. We will make a similar assumption in many of the theorems stated below.

We prove Theorem 2.4 using a new relative hypergraph removal lemma.² In the next subsection we set up the notation for hypergraphs and state the corresponding pseudorandomness hypothesis.

2.2. Hypergraphs. We borrow most of our notation and definitions from Tao [45, 46].

Definition 2.5 (Hypergraphs). Let J be a finite set and $r > 0$. Define $\binom{J}{r} = \{e \subseteq J : |e| = r\}$ to be the set of all r -element subsets of J . An r -uniform hypergraph on J is defined to be any subset $H \subseteq \binom{J}{r}$.

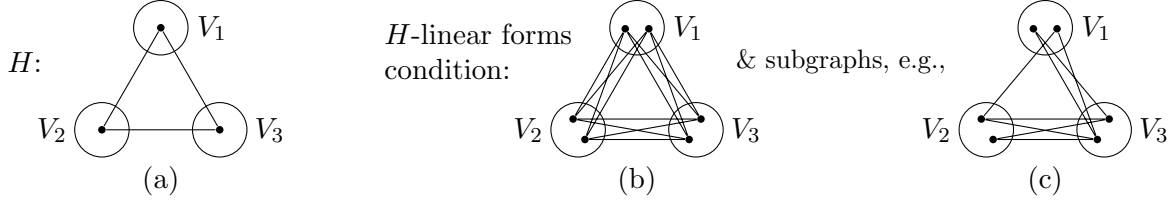
Definition 2.6 (Hypergraph system). A *hypergraph system* is a quadruple $V = (J, (V_j)_{j \in J}, r, H)$, where J is a finite set, $(V_j)_{j \in J}$ is a collection of finite non-empty sets indexed by J , $r \geq 1$ is a positive integer, and $H \subseteq \binom{J}{r}$ is an r -uniform hypergraph. For any $e \subseteq J$, we set $V_e := \prod_{j \in e} V_j$. For any $x = (x_j)_{j \in J} \in V_J$ and any subset $J' \subseteq J$, we write $x_{J'} = (x_j)_{j \in J'} \in V_{J'}$ to mean the natural projection of x onto the coordinates J' . Finally, for any $e \subseteq J$, we write ∂e for the set $\{f \subsetneq e : |f| = |e| - 1\}$, the skeleton of e .

Definition 2.7 (Weighted hypergraphs). Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system. A *weighted hypergraph* on V is a collection $g = (g_e)_{e \in H}$ of functions $g_e: V_e \rightarrow \mathbb{R}_{\geq 0}$ indexed by H . We write 0 and 1 to denote the constant-valued weighted hypergraphs of uniform weight 0 and 1, respectively. Given two weighted hypergraphs g and ν on the same hypergraph system, we write $g \leq \nu$ to mean that $g_e \leq \nu_e$ for all e , which in turn means that $g_e(x_e) \leq \nu_e(x_e)$ for all $x_e \in V_e$.

The weighted hypergraph ν plays an analogous role to the ν in Theorem 2.4, with $\nu = 1$ again corresponding to the dense case. We have an analogous linear forms condition for ν as a weighted hypergraph. We use the following indexing notation. For a finite set e and $\omega \in \{0, 1\}^e$, we write $x_e^{(\omega)}$ to mean the tuple $(x_j^{(\omega_j)})_{j \in e}$. We also write $x_e^{(0)} := (x_j^{(0)})_{j \in e}$ and similarly with $x_e^{(1)}$.

Definition 2.8 (Linear forms condition). A weighted hypergraph $\nu = \nu^{(N)}$ on the hypergraph system $V = V^{(N)} = (J, (V_j^{(N)})_{j \in J}, r, H)$ is said to obey the H -linear forms condition (or simply

²Green and Tao [22] prove a transference result that allows them to apply the dense version of Szemerédi's theorem as a black box. This allows them to show that the optimal c in (3) can be taken to be the same as the optimal c in (1). Since our proof goes through the hypergraph removal lemma, we do not obtain the same c . However, it seems plausible that one could obtain our result with the same c by modifying our argument to an arithmetic setting.

FIGURE 1. Linear forms conditions for $H = K_3$. See Example 2.9.

the *linear forms condition* if there is no confusion) if one has

$$\mathbb{E} \left[\prod_{e \in H} \prod_{\omega \in \{0,1\}^e} \nu_e(x_e^{(\omega)})^{n_{e,\omega}} \middle| x_J^{(0)}, x_J^{(1)} \in V_J \right] = 1 + o(1) \quad (4)$$

for any choice of exponents $n_{e,\omega}$.

Example 2.9. Let H be the set of all pairs in $J = \{1, 2, 3\}$. The linear forms condition says that

$$\mathbb{E} \left[\prod_{ij=12,13,23} \nu_{ij}(x_i, x_j) \nu_{ij}(x'_i, x_j) \nu_{ij}(x_i, x'_j) \nu_{ij}(x'_i, x'_j) \middle| x_1, x'_1 \in V_1, x_2, x'_2 \in V_2, x_3, x'_3 \in V_3 \right] = 1 + o(1).$$

and similarly if one or more of the twelve ν factors are deleted. This expression represents the weighted homomorphism density of $K_{2,2,2}$ in the weighted tripartite graph given by ν , as illustrated in Figure 1(b) (the vertices of $K_{2,2,2}$ must map into the corresponding parts). Deleting some ν factors corresponds to considering various subgraphs of $K_{2,2,2}$, e.g., Figure 1(c).

In general, the H -linear forms condition says that ν has roughly the expected density for the 2-blow-up³ of H as well as any subgraph of the 2-blow-up. Our linear forms condition for hypergraphs coincides with the one used by Tao [45, Def. 2.8], although in [45] one assumes additional pseudorandomness hypotheses on ν known as the dual function condition and the correlation condition.

2.3. Hypergraph removal lemma. The hypergraph removal lemma was first proved by Gowers [18] and by Nagle, Rödl, Schacht, and Skokan [32, 34, 35]. It states that for every r -uniform hypergraph H on h vertices, every r -uniform hypergraph on n vertices with $o(n^h)$ copies of H can be made H -free by removing $o(n^r)$ edges. As first explicitly stated and proved by Tao [46], the proof of the hypergraph removal lemma further gives that the edges can be removed in a low complexity way (this idea will soon be made formal). We will use a slightly stronger version, where edges are given weights in the interval $[0, 1]$. This readily follows from the usual version by a simple rounding argument, as done in [45, Thm. 3.7]. We state this result as Theorem 2.11 below.

Definition 2.10. For any set e of size r and any $E_e \subseteq V_e = \prod_{j \in e} V_j$, we define the *complexity* of E_e to be the minimum integer T such that there is a partition of E_e into T sets $E_{e,1}, \dots, E_{e,T}$, so that each $E_{e,i}$ is the set of r -cliques of some $(r-1)$ -uniform hypergraph, meaning that there exists some $B_{f,i} \subseteq V_f$ for each $f \in \partial e$ so that $1_{E_{e,i}}(x_e) = \prod_{f \in \partial e} 1_{B_{f,i}}(x_f)$ for all $x_e \in V_e$.

³By the 2-blow-up of H we mean the hypergraph consisting of vertices $j^{(0)}, j^{(1)}$ for each $j \in J$, and edges $e^{(\omega)} := \{j^{(\omega_j)} : j \in e\}$ for any $e \in H$ and $\omega \in \{0, 1\}^e$. We actually do not need the full strength of this assumption. It suffices to assume that ν has roughly the expected density for any subgraph of a *weak* 2-blow-up of H , where by a weak 2-blow-up we mean the following. Fix some edge $e_1 \in H$ (we will need to assume the condition for all e_1). The weak 2-blow-up of H with respect to e_1 is the subgraph of the usual 2-blow-up consisting of all edges $e^{(\omega)}$ where $\omega_i = \omega_j$ for any $i, j \in e \setminus e_1$. This weaker version of the H -linear forms condition is all we shall use for the proof, although everything to follow will be stated as in Definition 2.8 for clarity.

Theorem 2.11 (Weighted hypergraph removal lemma). *For every $\epsilon > 0$ and finite set J , there exists $\delta > 0$ and $T > 0$ such that the following holds. Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system. Let g be a weighted hypergraph on V satisfying*

$$\mathbb{E} \left[\prod_{e \in H} g_e(x_e) \mid x \in V_J \right] \leq \delta.$$

Then for each $e \in H$ there exists a set $E'_e \subseteq V_e$ for which $V_e \setminus E'_e$ has complexity at most T and such that

$$\prod_{e \in H} 1_{E'_e}(x_e) = 0 \text{ for all } x \in V_J$$

and for all $e \in H$ one has

$$\mathbb{E}[g_e(x_e) 1_{V_e \setminus E'_e}(x_e) \mid x_e \in V_e] \leq \epsilon.$$

We prove a relativized extension of the hypergraph removal lemma. A relative hypergraph removal lemma was already established by Tao in [45], where he assumed the majorizing measure satisfies three conditions: the linear forms condition, the correlation condition, and the dual function condition. We again show that a linear forms condition is sufficient.

Theorem 2.12 (Relative hypergraph removal lemma). *For every $\epsilon > 0$ and finite set J , there exists $\delta > 0$ and $T > 0$ such that the following holds. Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system. Let ν and g be weighted hypergraphs on V . Suppose $0 \leq g \leq \nu$, ν satisfies the H -linear forms condition, and N is sufficiently large. If*

$$\mathbb{E} \left[\prod_{e \in H} g_e(x_e) \mid x \in V_J \right] \leq \delta,$$

then for each $e \in H$ there exists a set $E'_e \subseteq V_e$ for which $V_e \setminus E'_e$ has complexity at most T and such that

$$\prod_{e \in H} 1_{E'_e}(x_e) = 0 \text{ for all } x \in V_J$$

and for all $e \in H$ one has

$$\mathbb{E}[g_e(x_e) 1_{V_e \setminus E'_e}(x_e) \mid x_e \in V_e] \leq \epsilon.$$

In Section 4 we will deduce Theorem 2.12 from Theorem 2.11 by applying the weak regularity lemma and the counting lemma which are stated in the next two subsections.

2.4. Weak hypergraph regularity. The Frieze-Kannan weak regularity lemma [14] allows one to approximate in cut-norm a matrix (or graph) with entries in the interval $[0, 1]$ by another matrix of low complexity. A major advantage over simply applying Szemerédi's regularity lemma is that the complexity has only an exponential dependence on the approximation parameter, as opposed to the tower-type bound that is incurred by Szemerédi's regularity lemma. Unfortunately, these regularity lemmas are not meaningful for sparse graphs as the error term is too large in this setting. Following sparse extensions of Szemerédi's regularity lemma by Kohayakawa [24] and Rödl, a sparse extension of the weak regularity lemma was proved by Bollobás and Riordan [3] and by Coja-Oghlan, Cooper, and Frieze [6]. In [6], they further generalize this to r -dimensional tensors (or r -uniform hypergraphs), but it only gives an approximation which is close in density on all hypergraphs induced by large vertex subsets. In order to prove a relative hypergraph removal lemma, we will need a stronger approximation, which is close in density on all dense r -uniform hypergraphs formed by the clique set of some $(r - 1)$ -uniform hypergraph. In Section 5, we will prove a more general sparse regularity lemma. For now, we state the result in the form that we need.

The weak regularity lemma approximates a weighted hypergraph g on V by another weighted hypergraph \tilde{g} of bounded complexity which satisfies $0 \leq \tilde{g} \leq 1$. One can think of \tilde{g} as a dense approximation of g . The following definition makes precise in what sense \tilde{g} approximates g .

Definition 2.13 (Discrepancy pair). Let e be a finite set and $g_e, \tilde{g}_e: \prod_{j \in e} V_j \rightarrow \mathbb{R}_{\geq 0}$ be two nonnegative functions. We say that (g_e, \tilde{g}_e) is an ϵ -discrepancy pair if for all subsets $B_f \subseteq V_f$, $f \in \partial e$, one has

$$\left| \mathbb{E} \left[(g_e(x_e) - \tilde{g}_e(x_e)) \prod_{f \in \partial e} 1_{B_f}(x_f) \middle| x_e \in V_e \right] \right| \leq \epsilon. \quad (5)$$

For two weighted hypergraphs g and \tilde{g} on $(J, (V_j)_{j \in J}, r, H)$, we say that (g, \tilde{g}) is an ϵ -discrepancy pair if (g_e, \tilde{g}_e) is an ϵ -discrepancy pair for all $e \in H$.

One needs an additional hypothesis on g in order to prove a weak regularity lemma. The condition roughly says that g contains “no dense spots.”

Definition 2.14 (Upper regular). Let e be a finite set, $g_e: \prod_{j \in e} V_j \rightarrow \mathbb{R}_{\geq 0}$ a nonnegative function, and $\eta > 0$. We say that g_e is *upper η -regular* if for all subsets $B_f \subseteq V_f$, $f \in \partial e$, one has

$$\mathbb{E} \left[(g_e(x_e) - 1) \prod_{f \in \partial e} 1_{B_f}(x_f) \middle| x_e \in V_e \right] \leq \eta. \quad (6)$$

A hypergraph g on $(J, (V_j)_{j \in J}, r, H)$ is *upper η -regular* if g_e is upper η -regular for all $e \in H$.

Note that unlike (5), there is no absolute value on the left-hand side of (6). The upper regularity hypothesis is needed for establishing the sparse regularity lemma. Fortunately, this mild hypothesis is automatically satisfied in our setting. We will say more about this in Section 6.2.

Lemma 2.15. *Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system. Let ν and g be weighted hypergraphs on V . Suppose $0 \leq g \leq \nu$ and ν satisfies the H -linear forms condition. Then g is upper $o(1)$ -regular.*

Define the *complexity* of a function $g: V_e \rightarrow [0, 1]$ to be the minimum T such that there is a partition of V_e into T subgraphs S_1, \dots, S_T , each of which is the set of r -cliques of some $(r-1)$ -uniform hypergraph (see Definition 2.10), and such that g is constant on each S_i . We state the regularity lemma below with a complexity bound on \tilde{g} , although the complexity bound will not actually be needed for our application.

Theorem 2.16 (Sparse weak regularity lemma). *For any $\epsilon > 0$ and function $g: V_1 \times \dots \times V_r \rightarrow \mathbb{R}_{\geq 0}$ which is upper η -regular with $\eta \leq 2^{-40r/\epsilon^2}$, there exists $\tilde{g}: V_1 \times \dots \times V_r \rightarrow [0, 1]$ with complexity at most $2^{20r/\epsilon^2}$ such that (g, \tilde{g}) is an ϵ -discrepancy pair.*

The special case $r = 2$ is the sparse extension of the Frieze-Kannan weak regularity lemma [3, 6].

2.5. Counting lemma. Informally, the counting lemma says that if (g, \tilde{g}) is an ϵ -discrepancy pair, with the additional assumption that $g \leq \nu$ and $\tilde{g} \leq 1$, then the density of H in \tilde{g} is close to the density of H in g . This sparse counting lemma is perhaps the most novel ingredient in this paper.

Theorem 2.17 (Counting lemma). *For every $\gamma > 0$ and finite set J , there exists an $\epsilon > 0$ so that the following holds. Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system and ν, g, \tilde{g} be weighted hypergraphs on V . Suppose that ν satisfies the H -linear forms condition and N is sufficiently large. Suppose also that $0 \leq g \leq \nu$, $0 \leq \tilde{g} \leq 1$, and (g, \tilde{g}) is an ϵ -discrepancy pair. Then*

$$\left| \mathbb{E} \left[\prod_{e \in H} g_e(x_e) \middle| x \in V_J \right] - \mathbb{E} \left[\prod_{e \in H} \tilde{g}_e(x_e) \middle| x \in V_J \right] \right| \leq \gamma. \quad (7)$$

As a corollary, Theorem 2.17 also holds if the hypothesis $0 \leq \tilde{g} \leq 1$ is replaced by $0 \leq \tilde{g} \leq \nu$. Indeed, we can use the weak regularity lemma, Theorem 2.16, to find a common 1-bounded approximation to g and \tilde{g} . The result then follows from Theorem 2.17 and the triangle inequality.

To summarize, to get a counting lemma for a fixed hypergraph H in a subgraph of a pseudorandom host hypergraph, it suffices to know that the host hypergraph has approximately the expected count for a somewhat larger family of hypergraphs (namely, subgraphs of the 2-blow-up of H).

3. THE RELATIVE SZEMERÉDI THEOREM

In this section, we deduce the relative Szemerédi theorem, Theorem 2.4, from the relative hypergraph removal lemma, Theorem 2.12. We use the relative hypergraph removal lemma to prove a relative arithmetic removal lemma, Theorem 3.2. This result then easily implies a relative version of the multidimensional Szemerédi theorem of Furstenberg and Katznelson [15]. This is Theorem 3.1 below. The relative Szemerédi theorem, Theorem 2.4, follows as a special case of Theorem 3.1 by setting $Z = Z' = \mathbb{Z}_N$ and $\phi_j(d) = (j-1)d$. One may easily check that the linear forms condition for the resulting hypergraph is satisfied if $\nu: \mathbb{Z}_N \rightarrow \mathbb{R}_{\geq 0}$ satisfies the k -linear forms condition.

The statement and proof of Theorem 3.1 closely follows the write-up in Tao [45, Thm 2.18], adapted in a straightforward way to our new pseudorandomness conditions as well as to the slightly more general setting of functions instead of subsets. Earlier versions of this type of argument for deducing Szemerédi-type results (in the dense setting) from graph and hypergraph removal lemmas were given by Ruzsa and Szemerédi [36], Frankl and Rödl [13], and Solymosi [41, 42].

Theorem 3.1 (Relative multidimensional Szemerédi theorem). *For a finite set J and $\delta > 0$, there exists $c > 0$ so that the following holds. Let Z, Z' be two finite additive groups and let $(\phi_j)_{j \in J}$ be a finite collection of group homomorphisms $\phi_j: Z \rightarrow Z'$ from Z to Z' . Assume that the elements $\{\phi_i(d) - \phi_j(d) : i, j \in J, d \in Z\}$ generate Z' as an abelian group. Let $\nu: Z' \rightarrow \mathbb{R}_{\geq 0}$ be a nonnegative function with the property that in the hypergraph system $V = (J, (V_j)_{j \in J}, r, H)$, with $V_j := Z$, $r := |J| - 1$, and $H := \binom{J}{r}$, the weighted hypergraph $(\nu_e)_{e \in H}$ defined by*

$$\nu_{J \setminus \{j\}}((x_i)_{i \in J \setminus \{j\}}) := \nu\left(\sum_{i \in J \setminus \{j\}} (\phi_i(x_i) - \phi_j(x_i))\right)$$

satisfies the H -linear forms condition. Assume that N is sufficiently large. Then, for any $f: Z' \rightarrow \mathbb{R}_{\geq 0}$ satisfying $0 \leq f(x) \leq \nu(x)$ for all $x \in Z'$ and $\mathbb{E}[f] \geq \delta$,

$$\mathbb{E}\left[\prod_{j \in J} f(a + \phi_j(d)) \mid a \in Z'; d \in Z\right] \geq c. \quad (8)$$

As in [45, Remark 2.19], we note that the hypothesis that $\{\phi_i(d) - \phi_j(d) : i, j \in J, d \in Z\}$ generate Z' can be dropped by foliating Z' into cosets. However, this results in a change to the linear forms hypothesis on ν , namely, that it must be assumed on every coset.

We shall prove Theorem 3.1 by proving a somewhat more general removal-type result for arithmetic patterns.

Theorem 3.2 (Relative arithmetic removal lemma). *For every finite set J and $\epsilon > 0$, there exists $c > 0$ so that the following holds. Let $Z, Z', (\phi_j)_{j \in J}, \nu$ be the same as in Theorem 3.1. For any collection of functions $\{f_j: Z' \rightarrow \mathbb{R}_{\geq 0}\}_{j \in J}$ satisfying $0 \leq f_j(x) \leq \nu(x)$ for all $x \in Z'$ and $j \in J$, and such that*

$$\mathbb{E}\left[\prod_{j \in J} f_j(a + \phi_j(d)) \mid a \in Z', d \in Z\right] \leq c \quad (9)$$

one can find $A_j \subseteq Z'$ for each $j \in J$ so that

$$\prod_{j \in J} 1_{A_j}(a + \phi_j(d)) = 0 \quad \text{for all } a \in Z', d \in Z \quad (10)$$

and

$$\mathbb{E}[f_j(x) 1_{Z' \setminus A_j}(x) \mid x \in Z'] \leq \epsilon \quad \text{for all } j \in J. \quad (11)$$

Theorem 3.1 follows from Theorem 3.2 by setting $f_j = f$ for all $j \in J$ and $\epsilon < \delta/(r+1)$. Indeed, if the conclusion (8) fails, then Theorem 3.2 implies that there exists $A_j \subseteq Z'$ for each $j \in J$ satisfying (10) and (11). The A_j 's cannot have a common intersection, or else (10) fails for $d = 0$. It follows

that $\{Z' \setminus A_j : j \in J\}$ covers Z' , and hence (11) implies that $\mathbb{E}[f] \leq \sum_j \mathbb{E}[f_j 1_{Z' \setminus A_j}] \leq (r+1)\epsilon < \delta$, which contradicts the hypothesis $\mathbb{E}[f] \geq \delta$.

Proof of Theorem 3.2. Let $V = (J, (V_j), r, H)$ be as in the statement of Theorem 3.1. Write $e_j := J \setminus \{j\} \in H$. Define the weighted hypergraph g on V by setting

$$g_{e_j}(x_{e_j}) := f_j(\psi_j(x_{e_j})) \quad \text{for all } j \in J$$

where $\psi_j : V_{e_j} \rightarrow Z'$ is defined by

$$\psi_j(x_{e_j}) = \sum_{i \in e_j} (\phi_i(x_i) - \phi_j(x_i)) = a + \phi_j(d) \quad (12)$$

where

$$a = \sum_{i \in J} \phi_i(x_i) \quad \text{and} \quad d = - \sum_{i \in J} x_i. \quad (13)$$

Then, for all $x \in V$ and a, d defined in (13), we have

$$\prod_{j \in J} g_{e_j}(x_{e_j}) = \prod_{j \in J} f_j(a + \phi_j(d)). \quad (14)$$

The homomorphism $x \mapsto (a, d) : V \rightarrow Z' \times Z$ given by (13) is surjective: the image contains $\{(\phi_i(d) - \phi_j(d), 0) : i, j \in J, d \in Z\}$ and hence all of $Z' \times \{0\}$. Moreover the image also contains $\{(-\phi_i(d), d) : i \in J, d \in Z\}$. Together, these sets generate all of $Z' \times Z$. It follows that (a, d) varies uniformly over $Z' \times Z$ as x varies uniformly over V_J , and so (14) implies that

$$\mathbb{E}\left[\prod_{j \in J} g_{e_j}(x_{e_j}) \mid x \in V_J\right] = \mathbb{E}\left[\prod_{j \in J} f_j(a + \phi_j(d)) \mid a \in Z'; d \in Z\right] \leq c.$$

By the relative hypergraph removal lemma, for c small enough (depending on J and ϵ), we can find a subset $E'_j \subset V_{e_j}$ for each $j \in J$ such that

$$\prod_{j \in J} 1_{E'_j}(x_{e_j}) = 0 \quad \text{for all } x \in V_J \quad (15)$$

and

$$\mathbb{E}[g_{e_j}(x_{e_j}) 1_{V_{e_j} \setminus E'_j}(x_{e_j}) \mid x_{e_j} \in V_{e_j}] \leq \epsilon / (r+1) \quad \text{for all } j \in J.$$

For each $j \in J$, define $A_j \subseteq Z'$ by

$$A_j := \{z' \in Z' : |\psi_j^{-1}(z') \cap E'_j| > \frac{r}{r+1} |\psi_j^{-1}(z')|\}. \quad (16)$$

In other words, A_j contains $z' \in Z'$ if the hypergraph removal lemma removes less than a $1/(r+1)$ fraction of the edges in V_j representing z' via ψ_j .

For any $z' \in Z' \setminus A_j$, on the fiber $\psi^{-1}(z')$ the function g_{e_j} takes the common value $f_j(z')$. Furthermore, by (16), on this fiber, the expectation of $1_{V_{e_j} \setminus E'_j}$ is at least $1/(r+1)$. Hence

$$\mathbb{E}[f_j(x) 1_{Z' \setminus A_j}(x) \mid x \in Z'] \leq (r+1) \mathbb{E}[g_{e_j}(x_{e_j}) 1_{V_{e_j} \setminus E'_j}(x_{e_j}) \mid x_{e_j} \in V_{e_j}] \leq \epsilon.$$

This proves (11). To prove (10), suppose for some $a \in Z', d \in Z$ we have $a + \phi_j(d) \in A_j$ for all $j \in J$. Let $V_J^{a,d} \subset V_J$ consist of all $x \in V_J$ satisfying (13). Then $\psi_j(x_{e_j}) = a + \phi_j(d)$ for all $x \in V_J^{a,d}$ by (12), and in fact $\psi_j^{-1}(a + \phi_j(d))$ is the projection of $V_J^{a,d}$ onto V_{e_j} . By (16), more than an $\frac{r}{r+1}$ fraction of this projection is in E'_j . It follows by the pigeonhole principle (or a union bound on the complement) that there exists some $x \in V_J^{a,d}$ such that $x_{e_j} \in E'_j$ for every $j \in J$. But this contradicts (15). Thus (10) holds. \blacksquare

4. THE RELATIVE HYPERGRAPH REMOVAL LEMMA

Proof of Theorem 2.12. By Lemma 2.15, ν is upper $o(1)$ -regular, so we can apply the weak sparse hypergraph regularity lemma (Theorem 2.16) to find functions $\tilde{g}_e : V_e \rightarrow [0, 1]$ for every $e \in H$ so that (g, \tilde{g}) is an $o(1)$ -discrepancy pair. By the counting lemma (Theorem 2.17), we have

$$\mathbb{E}\left[\prod_{e \in H} \tilde{g}_e(x_e) \middle| x \in V_J\right] = \mathbb{E}\left[\prod_{e \in H} g_e(x_e) \middle| x \in V_J\right] + o(1) \leq \delta + o(1).$$

The dense weighted hypergraph removal lemma (Theorem 2.11) tells us that for each $e \in H$ we can choose $E'_e \subset V_e$ for which $V_e \setminus E'_e$ has complexity $O_\delta(1)$ (i.e., at most some constant depending on δ) and such that

$$\prod_{e \in H} 1_{E'_e}(x_e) = 0 \quad \text{for all } x \in V_J$$

and, as long as δ is small enough and N is large enough, we have

$$\mathbb{E}[\tilde{g}_e(x_e) 1_{V_e \setminus E'_e}(x_e) | x_e \in V_e] \leq \epsilon/2 \quad \text{for all } e \in H. \quad (17)$$

As $V_e \setminus E'_e$ has complexity $O_\delta(1)$, there is a partition of $V_e \setminus E'_e$ into $O_\delta(1)$ hypergraphs F_{ei} each of which is the set of r -cliques of some $(r-1)$ -uniform hypergraph. We have

$$\begin{aligned} |\mathbb{E}[(\tilde{g}_e - g_e)(x_e) 1_{V_e \setminus E'_e}(x_e) | x_e \in V_e]| &= \sum_i |\mathbb{E}[(\tilde{g}_e - g_e)(x_e) 1_{F_{ei}}(x_e) | x_e \in V_e]| \\ &\leq \sum_i o(1) = O_\delta(1) o(1) \leq \epsilon/2 \quad \text{for all } e \in H. \end{aligned} \quad (18)$$

We used that (g_e, \tilde{g}_e) is an $o(1)$ -discrepancy pair on each of the terms of the sum, and the final inequality is true as long as N is large enough. Combining (17) and (18) we obtain

$$\mathbb{E}[g_e(x_e) 1_{V_e \setminus E'_e}(x_e) | x_e \in V_e] \leq \epsilon \quad \text{for all } e \in H.$$

This proves the claim. ■

5. THE WEAK REGULARITY LEMMA

Let X be a finite set and $g : X \rightarrow \mathbb{R}_{\geq 0}$. Let \mathcal{F} be a family of subsets of X which is closed under intersection, $X \in \mathcal{F}$, all subsets of X of size one are in \mathcal{F} , and such that, for every $S \in \mathcal{F}$, there is a partition of X which contains S and consists of members of \mathcal{F} . For $t \geq 2$, the family \mathcal{F} is *t-splittable* if for every $S \in \mathcal{F}$ there is a partition P of X into members of \mathcal{F} such that $S \in P$ and $|P| \leq t$. The *complexity* $p = p(f)$ of a function $f : X \rightarrow \mathbb{R}_{\geq 0}$ is the minimum p for which there is a partition $X = S_1 \cup \dots \cup S_p$ into p subsets each in \mathcal{F} such that f is constant on each S_i . We call (g, \tilde{g}) an ϵ -discrepancy pair if for all $A \in \mathcal{F}$,

$$|\mathbb{E}[(g - \tilde{g}) 1_A]| \leq \epsilon.$$

All expectations are done with the uniform measure on X . For P a partition of X , let g_P be the function on X given by $g_P(x) = \frac{\mathbb{E}[g 1_A]}{\mathbb{E}[1_A]}$ when $x \in A \in P$. That is, $g_P(x)$ is the conditional expectation of $g(x)$ given the partition P and is constant on any part A of the partition.

The function g we call *upper η -regular* if for every $A \in \mathcal{F}$, we have

$$\mathbb{E}[g 1_A] \leq \mathbb{E}[1_A] + \eta.$$

If g is upper η -regular, $A, B \in \mathcal{F}$, and \mathcal{F} is t -splittable, then

$$\mathbb{E}[g 1_{B \setminus A}] \leq \mathbb{E}[1_{B \setminus A}] + (t-1)\eta. \quad (19)$$

Indeed, in this case $B \setminus A$ can be partitioned into $t-1$ sets in \mathcal{F} (we first split with respect to A and then consider the intersections of the parts of the partition with B). Applying the upper η -regularity condition to each of these sets and summing up the inequalities, we arrive at (19).

Following Scott [39], let $\phi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ be the convex function given by

$$\phi(u) = \begin{cases} u^2 & \text{if } u \leq 2, \\ 4u - 4 & \text{otherwise.} \end{cases}$$

For a partition P of X , let $\phi(P) = \mathbb{E}[\phi(g_P)]$, which is the mean ϕ -density of g with respect to the partition P . As ϕ takes only nonnegative values and $\phi(u) \leq 4u$, we have

$$0 \leq \phi(P) \leq 4\mathbb{E}[g_P] = 4\mathbb{E}[g].$$

Also, by the convexity of ϕ , it follows that if P' is a refinement of P , then $\phi(P') \geq \phi(P)$.

Lemma 5.1. *Let X and \mathcal{F} as above be such that \mathcal{F} is t -splittable. Let $0 < \epsilon, \eta < 1$ and $T = t^{20/\epsilon^2}$. For any $g : X \rightarrow \mathbb{R}_{\geq 0}$ which is upper η -regular with $\eta \leq \frac{\epsilon}{8tT}$, there is $\tilde{g} : X \rightarrow [0, 1]$ with complexity at most T such that (g, \tilde{g}) is an ϵ -discrepancy pair.*

Proof. Let $\alpha = \frac{\epsilon^2}{4}$. We first find a partition P of X into members of \mathcal{F} with $|P| \leq t^{5/\alpha} = T$ such that for any refinement P' of P into members of \mathcal{F} with $|P'| \leq t|P|$, we have $\phi(P') - \phi(P) < \alpha$. In order to construct P , we first recursively construct a sequence P_0, P_1, \dots of finer partitions of X into members of \mathcal{F} so that $|P_j| \leq t^j$ and $\phi(P_j) \geq j\alpha$. We begin by considering the trivial partition $P_0 = \{X\}$, which satisfies $\phi(P_0) \geq 0$. At the beginning of step $j+1$, we have a partition P_j of X into members of \mathcal{F} with $|P_j| \leq t^j$ and $\phi(P_j) \geq j\alpha$. If there exists a refinement P_{j+1} of X into members of \mathcal{F} with $|P_{j+1}| \leq t|P_j|$ and $\phi(P_{j+1}) \geq \phi(P_j) + \alpha$, then we continue to step $j+2$. Otherwise, we may pick $P = P_j$ to be the desired partition. Note that this process must stop after at most $5/\alpha$ steps since $5 > 4(1 + \eta) \geq 4\mathbb{E}[g] \geq \phi(P_j) \geq j\alpha$, where the second inequality follows from g being upper η -regular. We therefore arrive at the desired partition P .

Let $P : X = S_1 \cup \dots \cup S_p$. Let $\tilde{g} : X \rightarrow [0, 1]$, where $\tilde{g} = g_P \wedge 1$ is the minimum of g_P and the constant function 1. We will show that (g_P, \tilde{g}) is an $\frac{\epsilon}{4}$ -discrepancy pair and (g_P, g) is a $\frac{3\epsilon}{4}$ -discrepancy pair, which implies by the triangle inequality that (g, \tilde{g}) is an ϵ -discrepancy pair. As \tilde{g} has complexity at most $|P| \leq T$, this will complete the proof.

We first show (g_P, \tilde{g}) is an $\frac{\epsilon}{4}$ -discrepancy pair. Note that $g_P - \tilde{g}$ is nonnegative and constant on each part of P . If $S_i \in P$ and $g_P - \tilde{g} > 0$ on S_i , then also $g_P > 1$ and $\tilde{g} = 1$ on S_i . As g is upper η -regular, we have $\mathbb{E}[g1_{S_i}] \leq \mathbb{E}[1_{S_i}] + \eta$ and hence $\mathbb{E}[(g - \tilde{g})1_{S_i}] \leq \eta$. Therefore, by summing over all parts in the partition P , we see that if $A \in \mathcal{F}$,

$$0 \leq \mathbb{E}[(g_P - \tilde{g})1_A] \leq \mathbb{E}[(g_P - \tilde{g})] \leq \eta|P| \leq \eta T \leq \frac{\epsilon}{4},$$

and (g_P, \tilde{g}) is an $\frac{\epsilon}{4}$ -discrepancy pair.

We next show that (g_P, g) is a $\frac{3\epsilon}{4}$ -discrepancy pair, which completes the proof. Suppose for contradiction that there is $A \in \mathcal{F}$ such that

$$|\mathbb{E}[(g_P - g)1_A]| > \frac{3\epsilon}{4}.$$

Let B be the union of all $S_i \cap A$, where $S_i \in P$, for which both $\mathbb{E}[1_{S_i \cap A}] \geq t\eta$ and $\mathbb{E}[1_{S_i \setminus A}] \geq t\eta$.

We claim that for each $S_i \in P$, we have

$$|\mathbb{E}[(g_P - g)(1_{A \cap S_i} - 1_{B \cap S_i})]| \leq 2t\eta. \quad (20)$$

Indeed, if $B \cap S_i = A \cap S_i$, then the left hand side of (20) is 0. Otherwise, $\mathbb{E}[1_{A \cap S_i}] \leq t\eta$ or $\mathbb{E}[1_{S_i \setminus A}] \leq t\eta$. In the first case, when $\mathbb{E}[1_{A \cap S_i}] \leq t\eta$, we have $1_{B \cap S_i}$ is identically 0, as well as

$$\mathbb{E}[g1_{A \cap S_i}] \leq \mathbb{E}[1_{A \cap S_i}] + \eta \leq (t+1)\eta$$

and

$$\mathbb{E}[g_P 1_{A \cap S_i}] = \frac{\mathbb{E}[g1_{S_i}]}{\mathbb{E}[1_{S_i}]} \mathbb{E}[1_{A \cap S_i}] \leq \frac{(\mathbb{E}[1_{S_i}] + \eta)}{\mathbb{E}[1_{S_i}]} \mathbb{E}[1_{A \cap S_i}] \leq \mathbb{E}[1_{A \cap S_i}] + \eta \leq (t+1)\eta,$$

from which (20) follows. In the second case, when $\mathbb{E}[1_{S_i \setminus A}] \leq t\eta$, we again have $1_{B \cap S_i}$ is identically 0, so that

$$\begin{aligned} \mathbb{E}[(g - g_P)(1_{A \cap S_i} - 1_{B \cap S_i})] &= \mathbb{E}[(g - g_P)1_{A \cap S_i}] = \mathbb{E}[(g - g_P)(1_{S_i} - 1_{S_i \setminus A})] \\ &= \mathbb{E}[(g - g_P)1_{S_i}] - \mathbb{E}[(g - g_P)1_{S_i \setminus A}] = -\mathbb{E}[(g - g_P)1_{S_i \setminus A}], \end{aligned}$$

and similar to the first case, using (19) to estimate $\mathbb{E}[g1_{S_i \setminus A}]$ and $\mathbb{E}[g_P1_{S_i \setminus A}]$, we get (20).

Notice that

$$|\mathbb{E}[(g_P - g)1_A] - \mathbb{E}[(g_P - g)1_B]| = |\mathbb{E}[(g_P - g)(1_A - 1_B)]| \leq |P|2t\eta \leq \frac{\epsilon}{4},$$

where the first inequality follows by using (20) for each part S_i and the triangle inequality. Hence,

$$|\mathbb{E}[(g_P - g)1_B]| \geq |\mathbb{E}[(g_P - g)1_A]| - |\mathbb{E}[(g_P - g)1_A] - \mathbb{E}[(g_P - g)1_B]| > \frac{3\epsilon}{4} - \frac{\epsilon}{4} = \frac{\epsilon}{2}.$$

Let \hat{P} be the refinement of P where S_i is also in \hat{P} if $B \cap S_i = \emptyset$ and otherwise $S_i \cap B$ and $S_i \setminus B$ are parts of \hat{P} , and let P' be a refinement of \hat{P} into at most $t|P|$ members of \mathcal{F} . The refinement P' exists as \mathcal{F} is t -splittable and is closed under intersections, P consists of members of \mathcal{F} , $A \in \mathcal{F}$, and $S_i \cap B = S_i \cap A \in \mathcal{F}$ if $S_i \cap B \in \hat{P}$. As P' is a refinement of \hat{P} which is a refinement of P , we have $\phi(P') \geq \phi(\hat{P}) \geq \phi(P)$. Let $R \in \{S_i, S_i \cap B, S_i \setminus B\}$, where S_i is a part of P that is refined into two parts in \hat{P} , so that $\mathbb{E}[1_R] \geq t\eta$. Letting $u = \frac{\mathbb{E}[g1_R]}{\mathbb{E}[1_R]}$, we see, since g is upper η -regular and using (19), that $u \leq 1 + t\eta(t\eta)^{-1} = 2$ and hence $\phi(u) = u^2$. It follows, by considering the functions pointwise, that $\phi(g_{\hat{P}}) - \phi(g_P) = g_{\hat{P}}^2 - g_P^2$. Hence,

$$\begin{aligned} \phi(P') - \phi(P) &\geq \phi(\hat{P}) - \phi(P) = \mathbb{E}[g_{\hat{P}}^2] - \mathbb{E}[g_P^2] = \mathbb{E}[g_{\hat{P}}^2 - g_P^2] = \mathbb{E}[(g_{\hat{P}} - g_P)^2] \\ &\geq \mathbb{E}[(g_{\hat{P}} - g_P)1_B]^2 = \mathbb{E}[(g - g_P)1_B]^2 > \frac{\epsilon^2}{4} = \alpha. \end{aligned}$$

The third equality above is the Pythagorean identity, which uses that \hat{P} is a refinement of P , and the second inequality is an application of the Cauchy-Schwarz inequality. However, since P' is a refinement of P consisting of members of \mathcal{F} with $|P'| \leq t|P|$, this contradicts $\phi(P') - \phi(P) < \alpha$ from the definition of P and completes the proof. \blacksquare

To establish the weak hypergraph regularity lemma, Theorem 2.16, we use Lemma 5.1 with $X = V_1 \times \cdots \times V_r$ and \mathcal{F} being the family of subsets of X which form the r -cliques of some r -partite $(r-1)$ -uniform hypergraph with parts V_1, \dots, V_r . Noting that \mathcal{F} is 2^r -splittable in this case, we obtain Theorem 2.16.

6. THE COUNTING LEMMA

The three main ingredients in our proof of the counting lemma (Theorem 2.17) are as follows.

- (1) A standard telescoping argument [4] in the dense case, i.e., when $\nu = 1$.
- (2) Repeated applications of the Cauchy-Schwarz inequality. This is a standard technique in this area, e.g., [17, 18, 22, 45].
- (3) *Densification*. This is the main new ingredient in our proof. At each step, we reduce the problem of counting H in a particular weighted hypergraph to that of counting H in a modified weighted hypergraph. For an edge $e \in H$, we replace the triple $(\nu_e, g_e, \tilde{g}_e)$ by a new triple $(1, g'_e, \tilde{g}'_e)$ with $0 \leq g'_e, \tilde{g}'_e \leq 1$ and such that (g'_e, \tilde{g}'_e) is an ϵ' -discrepancy pair for some $\epsilon' = o_{\epsilon \rightarrow 0}(1)$. By repeatedly applying this reduction to all $e \in H$ (we use induction), we reduce the counting lemma to the dense case.

We developed the densification technique in our earlier paper [7], where we proved a sparse counting lemma in graphs. We have significantly simplified a number of technical steps from [7] in order to extend the densification technique to hypergraphs here.

6.1. Telescoping argument. The following argument allows us to prove the counting lemma in the dense case, i.e., when $0 \leq g \leq 1$.

Lemma 6.1 (Telescoping discrepancy argument for dense hypergraphs). *Theorem 2.17 holds if we assume that there is some $e_1 \in H$ so that $\nu_e = 1$ for all $e \in H \setminus \{e_1\}$. In fact, in this case,*

$$\left| \mathbb{E} \left[\prod_{e \in H} g_e(x_e) \middle| x \in V_J \right] - \mathbb{E} \left[\prod_{e \in H} \tilde{g}_e(x_e) \middle| x \in V_J \right] \right| \leq |H| \epsilon. \quad (21)$$

Lemma 6.1 uses only the assumption that (g_e, \tilde{g}_e) is an ϵ -discrepancy pair for every $e \in H$ and nothing about the linear forms condition on ν . Recall that for each fixed $e \in H$, the condition that (g_e, \tilde{g}_e) is an ϵ -discrepancy pair means that for all subsets $B_f \subseteq V_f$, $f \in \partial e$, we have

$$\left| \mathbb{E} \left[(g_e(x_e) - \tilde{g}_e(x_e)) \prod_{f \in \partial e} 1_{B_f}(x_f) \middle| x_e \in V_e \right] \right| \leq \epsilon. \quad (22)$$

This is equivalent to the condition that for all functions $u_f: V_f \rightarrow [0, 1]$, $f \in \partial e$, we have

$$\left| \mathbb{E} \left[(g_e(x_e) - \tilde{g}_e(x_e)) \prod_{f \in \partial e} u_f(x_f) \middle| x_e \in V_e \right] \right| \leq \epsilon. \quad (23)$$

Indeed, the expectation is linear in each u_f and hence the extrema occur when the u_f 's are $\{0, 1\}$ -valued, thereby reducing to (22).

Proof. Let $h = |H|$ and order the edges of $H \setminus \{e_1\}$ arbitrarily as e_2, \dots, e_h . We can write the left-hand side of (21), without the absolute values, as a telescoping sum

$$\sum_{t=1}^h \mathbb{E} \left[\left(\prod_{s=1}^{t-1} \tilde{g}_{e_s}(x_{e_s}) \right) (g_{e_t}(x_{e_t}) - \tilde{g}_{e_t}(x_{e_t})) \left(\prod_{s=t+1}^h g_{e_s}(x_{e_s}) \right) \middle| x \in V_J \right]. \quad (24)$$

For the t -th term in the sum, when we fix the value of $x_{J \setminus e_t} \in V_{J \setminus e_t}$, the expectation has the form

$$\mathbb{E} \left[(g_{e_t}(x_{e_t}) - \tilde{g}_{e_t}(x_{e_t})) \prod_{f \in \partial e_t} u_f(x_f) \middle| x_{e_t} \in V_{e_t} \right] \quad (25)$$

for some functions $u_f: V_f \rightarrow [0, 1]$ (here we used the key fact that $g_{e_s} \leq 1$ for all $s > 1$ and $\tilde{g}_{e_s} \leq 1$ for all s). Since $(g_{e_t}, \tilde{g}_{e_t})$ is an ϵ -discrepancy pair, (23) implies that (25) is bounded in absolute value by ϵ . The same bound holds after we vary $x_{J \setminus e_t} \in V_{J \setminus e_t}$. So every term in (24) is bounded by ϵ in absolute value, and hence (24) is at most $h\epsilon$ in absolute value. \blacksquare

6.2. Strong linear forms. The main result of this subsection tells us that ν can be replaced by the constant function 1 in counting expressions. Though somewhat technical in detail, the main idea of the proof is quite simple and may be summarized as follows: we use the Cauchy-Schwarz inequality to double each vertex j of a certain edge in turn, at each step majorizing those edges which do not contain j . This method is quite standard in the field. In the work of Green and Tao, it is used to prove generalized von Neumann theorems [22, Prop. 5.3], [45, Thm. 3.8], although the statement of our lemma is perhaps more similar to the uniform distribution property [22, Prop. 6.2], [45, Prop. 5.1].

We begin by using a similar method to prove a somewhat easier result. It shows that if ν satisfies the H -linear forms condition then $(\nu, 1)$ is an $o(1)$ -discrepancy pair, which implies Lemma 2.15.

Lemma 6.2. *Let e be a finite set, V_j a finite set for each $j \in e$, and $V_e = \prod_{j \in e} V_j$. For any function $\nu: V_e \rightarrow \mathbb{R}$, and any collection of $B_f \subseteq V_f$ for $f \in \partial e$, we have*

$$\left| \mathbb{E} \left[(\nu_e(x_e) - 1) \prod_{f \in \partial e} 1_{B_f}(x_f) \middle| x_e \in V_e \right] \right| \leq \mathbb{E} \left[\prod_{\omega \in \{0,1\}^e} (\nu_e(x_e^{(\omega)}) - 1) \middle| x_e^{(0)}, x_e^{(1)} \in V_e \right]^{1/2^{|e|}}. \quad (26)$$

Lemma 6.2 follows from a direct application of the Gowers-Cauchy-Schwarz [17] inequality for hypergraphs (see [8]). We include the proof here for completeness.

Proof. For $\emptyset \subseteq d \subseteq e$, let

$$X_d := \prod_{\omega \in \{0,1\}^d} (v_e(x_{e \setminus d}, x_d^{(\omega)}) - 1), \quad Y_d := \prod_{\substack{f \in \partial e \\ f \supseteq d}} \prod_{\omega \in \{0,1\}^d} 1_{B_f}(x_{f \setminus d}, x_d^{(\omega)}),$$

and

$$Q_d := \mathbb{E}[X_d Y_d | x_{e \setminus d} \in V_{e \setminus d}, x_d^{(0)}, x_d^{(1)} \in V_d].$$

Then (26) can be written as $|Q_\emptyset| \leq Q_e^{1/2^{|e|}}$. By induction, it suffices to show that $Q_d^2 \leq Q_{d \cup \{j\}}$ whenever $j \in e \setminus d$. Let $Y_d = Y_d^{\supseteq j} Y_d^{\not\supseteq j}$ where $Y_d^{\supseteq j}$ consists of all the factors in Y_d that contain x_j in the argument, and $Y_d^{\not\supseteq j}$ consists of all other factors. By the Cauchy-Schwarz inequality, we have

$$Q_d^2 = \mathbb{E}[\mathbb{E}[X_d Y_d^{\supseteq j} | x_j \in V_j] Y_d^{\not\supseteq j}]^2 \leq \mathbb{E}[\mathbb{E}[X_d Y_d^{\supseteq j} | x_j \in V_j]^2] \mathbb{E}[(Y_d^{\not\supseteq j})^2] \leq Q_{d \cup \{j\}},$$

since $Q_{d \cup \{j\}} = \mathbb{E}[\mathbb{E}[X_d Y_d^{\supseteq j} | x_j \in V_j]^2]$ and $0 \leq Y_d^{\not\supseteq j} \leq 1$, where the outer expectations are taken over all free variables. This shows that $Q_d^2 \leq Q_{d \cup \{j\}}$. Hence, $|Q_\emptyset| \leq Q_e^{1/2^{|e|}}$, as desired. \blacksquare

The next lemma is very similar, except that now we need to invoke the linear forms condition.

Lemma 6.3 (Strong linear forms). *Let $V = (J, (V_j)_{j \in J}, r, H)$ be a hypergraph system and let ν be a weighted hypergraph on V satisfying the linear forms condition. Let $e_1 \in H$. For each $\iota \in \{0, 1\}$ and $e \in H \setminus \{e_1\}$, let $g_e^{(\iota)} : V_e \rightarrow \mathbb{R}_{\geq 0}$ be a function so that either $g_e^{(\iota)} \leq 1$ or $g_e^{(\iota)} \leq \nu_e$ holds. Then*

$$\mathbb{E}\left[(\nu_{e_1}(x_{e_1}) - 1) \prod_{\iota \in \{0,1\}} \left(\prod_{e \in H \setminus \{e_1\}} g_e^{(\iota)}(x_e^{(\iota)}) \right) \middle| x_J^{(0)}, x_J^{(1)} \in V_J; x_{e_1}^{(0)} = x_{e_1}^{(1)} = x_{e_1}\right] = o(1). \quad (27)$$

In (27) the notation $x_{e_1}^{(0)} = x_{e_1}^{(1)} = x_{e_1}$ means that $x_j^{(0)}, x_j^{(1)}, x_j$ are taken to be the same for all $j \in e_1$. Recall that we write $o(1)$ for a quantity that tends to zero as $N \rightarrow \infty$.

Proof. For each $\iota \in \{0, 1\}$ and $e \in H \setminus \{e_1\}$, let $\bar{g}_e^{(\iota)}$ be either 1 or ν_e so that $g_e^{(\iota)} \leq \bar{g}_e^{(\iota)}$ holds. For $\emptyset \subseteq d \subseteq e_1$, define

$$X_d := \prod_{\omega \in \{0,1\}^d} (\nu_{e_1}(x_{e_1 \setminus d}, x_d^{(\omega)}) - 1),$$

$$Y_d := \prod_{\iota \in \{0,1\}} \prod_{e \in H \setminus \{e_1\}} \prod_{\omega \in \{0,1\}^{e \cap d}} \left\{ \begin{array}{ll} g_e^{(\iota)}(x_{e \setminus e_1}^{(\iota)}, x_d^{(\omega)}, x_{e \cap e_1 \setminus d}) & \text{if } e \supseteq d \\ \bar{g}_e^{(\iota)}(x_{e \setminus e_1}^{(\iota)}, x_{e \cap d}^{(\omega)}, x_{e \cap e_1 \setminus d}) & \text{if } e \not\supseteq d \end{array} \right\},$$

and

$$Q_d := \mathbb{E}[X_d Y_d | x_{(J \setminus e_1) \cup d}^{(0)}, x_{(J \setminus e_1) \cup d}^{(1)} \in V_{(J \setminus e_1) \cup d}, x_{e_1 \setminus d} \in V_{e_1 \setminus d}].$$

We observe that Q_\emptyset is equal to the left-hand side of (27) and

$$Q_{e_1} = \mathbb{E}\left[\prod_{\omega \in \{0,1\}^{e_1}} (\nu_{e_1}(x_{e_1}^{(\omega)}) - 1) \prod_{\iota \in \{0,1\}} \prod_{e \in H \setminus \{e_1\}} \prod_{\omega \in \{0,1\}^{e \cap e_1}} \bar{g}_e^{(\iota)}(x_{e \setminus e_1}^{(\iota)}, x_{e \cap e_1}^{(\omega)}) \middle| x_J^{(0)}, x_J^{(1)} \in V_J \right] = o(1)$$

by the linear forms condition (4).⁴ Indeed, after we expand $\prod_{\omega \in \{0,1\}^{e_1}} (\nu_{e_1}(x_{e_1}^{(\omega)}) - 1)$, every term in Q_{e_1} has the form of (4) (since $\bar{g}_e^{(\iota)}$ is 1 or ν_e). Thus Q_{e_1} is the sum of $2^{|e_1|}$ terms, each of which is $\pm(1 + o(1))$ by the linear forms condition, and they cancel accordingly to $o(1)$.

⁴This is where the weak 2-blow-up of H arises, since the estimate $Q_{e_1} = o(1)$ only relies upon knowing that ν has roughly the expected density for certain subgraphs of the weak 2-blow-up.

We claim that if $j \in e_1 \setminus d$ then

$$|Q_d| \leq (1 + o(1))Q_{d \cup \{j\}}^{1/2}, \quad (28)$$

from which it would follow by induction that

$$|\text{LHS of (27)}| = |Q_\emptyset| \leq (1 + o(1))Q_{e_1}^{1/2^r} = o(1).$$

Now we prove (28). Let $Y_d = Y_d^{\ni j} Y_d^{\not\ni j}$ where $Y_d^{\ni j}$ consists of all the factors in Y_d that contain x_j in the argument, and $Y_d^{\not\ni j}$ consists of all other factors. Using the Cauchy-Schwarz inequality and $Y_d^{\not\ni j} \leq \bar{Y}_d^{\not\ni j}$ one has

$$\begin{aligned} Q_d^2 &= \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j] Y_d^{\not\ni j}]^2 \leq \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j]^2 Y_d^{\not\ni j}] \mathbb{E}[Y_d^{\not\ni j}] \\ &\leq \mathbb{E}[\mathbb{E}[X_d Y_d^{\ni j} | x_j \in V_j]^2 \bar{Y}_d^{\not\ni j}] \mathbb{E}[\bar{Y}_d^{\not\ni j}] = Q_{d \cup \{j\}} \mathbb{E}[\bar{Y}_d^{\not\ni j}] \end{aligned} \quad (29)$$

where the outer expectations are taken over all free variables. The second factor in (29) is $1 + o(1)$ by the linear forms condition (4) as $\bar{Y}_d^{\not\ni j}$ consists only of ν factors. This proves (28). \blacksquare

6.3. Counting lemma proof. As already mentioned, the main idea of the following proof is a process called densification, where we reduce the problem of counting H in a sparse hypergraph to that of counting H in a dense hypergraph by replacing sparse edges with dense edges one at a time. Several steps are needed to densify a given edge e_1 . The first step is to double all vertices outside of e_1 and to majorize g_{e_1} by ν_{e_1} . We then use the strong linear forms condition to remove the edge corresponding to e_1 entirely. This leaves us with the seemingly harder problem of counting the graph H' consisting of two copies of $H \setminus \{e_1\}$ joined along the vertices of e_1 . However, an inductive hypothesis tells us that we can count copies of $H \setminus \{e_1\}$. The core of the proof is in showing that this allows us to replace one of the copies of $H \setminus \{e_1\}$ in H' by a dense edge, thus reducing our problem to that of counting H with one edge replaced by a dense edge.

Proof of Theorem 2.17. We use induction on $|\{e \in H : \nu_e \neq 1\}|$. When $|\{e \in H : \nu_e \neq 1\}| = 0$ or 1 , the result follows from Lemma 6.1. Now take $e_1 \in H$ so that $\nu_{e_1} \neq 1$.

We assume that $|J|$ is a fixed constant. We write $o(1)$ for a quantity that tends to zero as $N \rightarrow \infty$ and $o_{\epsilon \rightarrow 0}(1)$ for a quantity that tends to zero as $N \rightarrow \infty$ and $\epsilon \rightarrow 0$. We need to show that the following quantity is $o_{\epsilon \rightarrow 0}(1)$:

$$\begin{aligned} &\mathbb{E}\left[\prod_{e \in H} g_e(x_e) \middle| x \in V_J\right] - \mathbb{E}\left[\prod_{e \in H} \tilde{g}_e(x_e) \middle| x \in V_J\right] \\ &= \mathbb{E}\left[g_{e_1}(x_{e_1}) \left(\prod_{e \in H \setminus \{e_1\}} g_e(x_e) - \prod_{e \in H \setminus \{e_1\}} \tilde{g}_e(x_e)\right) \middle| x \in V_J\right] + \mathbb{E}\left[(g_{e_1}(x_{e_1}) - \tilde{g}_{e_1}(x_{e_1})) \left(\prod_{e \in H \setminus \{e_1\}} \tilde{g}_e(x_e)\right) \middle| x \in V_J\right]. \end{aligned} \quad (30)$$

The second term in (30) is at most ϵ in absolute value since $(g_{e_1}, \tilde{g}_{e_1})$ is an ϵ -discrepancy pair and $\tilde{g} \leq 1$ (e.g., see proof of Lemma 6.1). It remains to show that the first term in (30) is $o_{\epsilon \rightarrow 0}(1)$.

Define functions $\nu'_{e_1}, g'_{e_1}, \tilde{g}'_{e_1} : V_{e_1} \rightarrow \mathbb{R}_{\geq 0}$ by

$$\nu'_{e_1}(x_{e_1}) := \mathbb{E}\left[\prod_{e \in H \setminus \{e_1\}} \nu_e(x_e) \middle| x_{J \setminus e_1} \in V_{J \setminus e_1}\right], \quad (31)$$

$$g'_{e_1}(x_{e_1}) := \mathbb{E}\left[\prod_{e \in H \setminus \{e_1\}} g_e(x_e) \middle| x_{J \setminus e_1} \in V_{J \setminus e_1}\right], \quad (32)$$

$$\tilde{g}'_{e_1}(x_{e_1}) := \mathbb{E}\left[\prod_{e \in H \setminus \{e_1\}} \tilde{g}_e(x_e) \middle| x_{J \setminus e_1} \in V_{J \setminus e_1}\right]. \quad (33)$$

We have $g'_{e_1} \leq \nu'_{e_1}$ and $\tilde{g}_{e_1} \leq 1$ (pointwise). In the rest of this proof, unless otherwise specified, expectations are for functions on V_{e_1} with arguments varying uniformly over V_{e_1} . The linear forms condition (4) implies that $\mathbb{E}[\nu'_{e_1}] = 1 + o(1)$ and $\mathbb{E}[(\nu'_{e_1})^2] = 1 + o(1)$, so that⁵

$$\mathbb{E}[(\nu'_{e_1} - 1)^2] = o(1). \quad (34)$$

The square of the first term in (30) equals

$$\begin{aligned} \mathbb{E}[g_{e_1}(g'_{e_1} - \tilde{g}'_{e_1})]^2 &\leq \mathbb{E}[g_{e_1}(g'_{e_1} - \tilde{g}'_{e_1})^2] \mathbb{E}[g_{e_1}] \leq \mathbb{E}[\nu_{e_1}(g'_{e_1} - \tilde{g}'_{e_1})^2] \mathbb{E}[\nu_{e_1}] \\ &= (\mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})^2] + o(1))(1 + o(1)). \end{aligned} \quad (35)$$

The first inequality above is due to the Cauchy-Schwarz inequality. In the final step, both factors are estimated using Lemma 6.3 (for the first factor, expand the square $(g'_{e_1} - \tilde{g}'_{e_1})^2$ and apply Lemma 6.3 term by term). Continuing (35) it suffices to show that the following quantity is $o_{\epsilon \rightarrow 0}(1)$:

$$\mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})^2] = \mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})(g'_{e_1} - g'_{e_1} \wedge 1)] + \mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})(g'_{e_1} \wedge 1 - \tilde{g}'_{e_1})] \quad (36)$$

(here $a \wedge b := \min\{a, b\}$). That is, we are capping the weighted hypergraph g'_{e_1} by 1. Since ν'_{e_1} is very close to 1 by (34), this should not result in a large loss. Indeed, since $0 \leq g'_{e_1} \leq \nu'_{e_1}$, we have

$$0 \leq g'_{e_1} - g'_{e_1} \wedge 1 = \max\{g'_{e_1} - 1, 0\} \leq \max\{\nu'_{e_1} - 1, 0\} \leq |\nu'_{e_1} - 1|. \quad (37)$$

Using (37), $g'_{e_1} \leq \nu'_{e_1}$, and $\tilde{g}'_{e_1} \leq 1$, we bound the magnitude of the first term on the right-hand side of (36) by

$$\mathbb{E}[(\nu'_{e_1} + 1)|\nu'_{e_1} - 1|] = \mathbb{E}[(\nu'_{e_1} - 1)|\nu'_{e_1} - 1|] + 2\mathbb{E}[|\nu'_{e_1} - 1|] \leq \mathbb{E}[(\nu'_{e_1} - 1)^2] + 2\mathbb{E}[(\nu'_{e_1} - 1)^2]^{1/2} = o(1)$$

by the triangle inequality, the Cauchy-Schwarz inequality, and (34). To estimate the second term on the right-hand side of (36), we need the following claim.

Claim. $(g'_{e_1} \wedge 1, \tilde{g}'_{e_1})$ is an ϵ' -discrepancy pair with $\epsilon' = o_{\epsilon \rightarrow 0}(1)$.

Proof of Claim. We need to show that, whenever $B_f \subseteq V_f$ for all $f \in \partial e_1$, we have

$$\mathbb{E}\left[(g'_{e_1}(x_{e_1}) \wedge 1 - \tilde{g}'_{e_1}(x_{e_1})) \prod_{f \in \partial e_1} 1_{B_f}(x_f) \middle| x_{e_1} \in V_{e_1}\right] = o_{\epsilon \rightarrow 0}(1). \quad (38)$$

Define $g''_{e_1} : V_{e_1} \rightarrow \mathbb{R}_{\geq 0}$ by $g''_{e_1}(x_{e_1}) = \prod_{f \in \partial e_1} 1_{B_f}(x_f)$. So the left-hand side of (38) is equal to

$$\mathbb{E}[(g'_{e_1} \wedge 1 - g'_{e_1})g''_{e_1}] + \mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})g''_{e_1}]. \quad (39)$$

Using $0 \leq g''_{e_1} \leq 1$, (37), the Cauchy-Schwarz inequality, and (34), we can bound the magnitude of the first term in (39) by

$$\mathbb{E}[|\nu'_{e_1} - 1|] \leq \mathbb{E}[(\nu'_{e_1} - 1)^2]^{1/2} = o(1).$$

The second term on the right-hand side of (39) is equal to

$$\mathbb{E}\left[\left(\prod_{e \in H \setminus \{e_1\}} g_e(x_e) - \prod_{e \in H \setminus \{e_1\}} \tilde{g}_e(x_e)\right) g''_{e_1}(x_{e_1}) \middle| x \in V_J\right].$$

This is $o_{\epsilon \rightarrow 0}(1)$ by the induction hypothesis applied to new weighted hypergraphs where the old $(\nu_{e_1}, g_{e_1}, \tilde{g}_{e_1})$ gets replaced by $(1, g''_{e_1}, g''_{e_1})$, thereby decreasing $|\{e \in H : \nu_e \neq 1\}|$. Note that the linear forms condition continues to hold. Thus (38) holds, so $(g'_{e_1} \wedge 1, \tilde{g}'_{e_1})$ is an ϵ' -discrepancy pair with $\epsilon' = o_{\epsilon \rightarrow 0}(1)$. \square

We expand the second term of (36) as

$$\mathbb{E}[(g'_{e_1} - \tilde{g}'_{e_1})(g'_{e_1} \wedge 1 - \tilde{g}'_{e_1})] = \mathbb{E}[g'_{e_1}(g'_{e_1} \wedge 1)] - \mathbb{E}[g'_{e_1}\tilde{g}'_{e_1}] - \mathbb{E}[\tilde{g}'_{e_1}(g'_{e_1} \wedge 1)] + \mathbb{E}[(\tilde{g}'_{e_1})^2] \quad (40)$$

⁵In fact, the only assumptions on ν needed for the proof of Theorem 2.17 are (34) and the strong linear forms condition, Lemma 6.3, as well as analogous conditions for other choices of $e_1 \in H$ and allowing some subset of the functions ν_e to be replaced by 1.

We claim that each expectation on the right-hand side of (40) is $\mathbb{E}[(\tilde{g}'_{e_1})^2] + o_{\epsilon \rightarrow 0}(1)$. Indeed, by (32) and (33) we have

$$\mathbb{E}[g'_{e_1}(g'_{e_1} \wedge 1)] - \mathbb{E}[(\tilde{g}'_{e_1})^2] = \mathbb{E}\left[\left((g'_{e_1}(x_{e_1}) \wedge 1) \prod_{e \in H \setminus \{e_1\}} g_e(x_e) - \tilde{g}'_{e_1}(x_{e_1}) \prod_{e \in H \setminus \{e_1\}} \tilde{g}_e(x_e)\right) \middle| x \in V_J\right],$$

which is $o_{\epsilon \rightarrow 0}(1)$ by the induction hypothesis applied to new weighted hypergraphs where the old $(\nu_{e_1}, g_{e_1}, \tilde{g}_{e_1})$ is replaced by $(1, g'_{e_1} \wedge 1, \tilde{g}'_{e_1})$. This is allowed as $(g'_{e_1} \wedge 1, \tilde{g}'_{e_1})$ is an ϵ' -discrepancy pair with $\epsilon' = o_{\epsilon \rightarrow 0}(1)$, the new ν still satisfies the linear forms condition, and $|\{e \in H : \nu_e \neq 1\}|$ has decreased. The claims that other terms on the right-hand side of (40) are each $\mathbb{E}[(\tilde{g}'_{e_1})^2] + o_{\epsilon \rightarrow 0}(1)$ are similar (in fact, easier). It follows that (40) is $o_{\epsilon \rightarrow 0}(1)$, so (36) is $o_{\epsilon \rightarrow 0}(1)$ and we are done. ■

7. CONCLUDING REMARKS

Conditions for counting lemmas. In this paper, we determined sufficient conditions for establishing a relative Szemerédi theorem and, more generally, a counting lemma for sparse hypergraphs. We have assumed that the hypergraph we want to count within is a subgraph of a pseudorandom hypergraph. The main question then is to determine a good notion of pseudorandomness which is sufficient to establish a counting lemma.

There is a marked difference between this paper and our previous paper on graphs [7] in terms of the type of pseudorandom condition assumed for the majorizing hypergraph. In this paper, we prove a counting lemma for a given hypergraph H by assuming that the underlying pseudorandom hypergraph contains approximately the correct count for each hypergraph in a certain collection of hypergraphs \mathcal{H} derived from H . That is, for each $H' \in \mathcal{H}$, we assume that our pseudorandom hypergraph contains $(1 + o(1))p^{e(H')}n^{v(H')}$ labeled copies of H' , where p is the edge density of the pseudorandom hypergraph.

The approach used in [7] is equivalent, up to some polynomial loss in ϵ , to assuming that the number of labeled cycles of length 4 in our pseudorandom graph is $(1 + \epsilon)p^4n^4$, where ϵ is now a carefully controlled term and the question of whether H can be embedded in our pseudorandom graph depends on whether ϵ is sufficiently small with respect to H and p . It is possible to adapt the methods of this paper so that the notion of pseudorandomness used for hypergraphs is more closely related to this latter notion. However, for the purposes of applying the results to a relative Szemerédi theorem, the current formulation seemed more appropriate.

Gowers uniformity norms. For a function $f : \mathbb{Z}_N \rightarrow \mathbb{R}$, the Gowers U^r -norm of f is defined to be

$$\|f\|_{U^r} = \mathbb{E}\left[\prod_{\omega \in \{0,1\}^r} f(x_0 + \omega \cdot \mathbf{x}) \middle| x_0, x_1, \dots, x_r \in \mathbb{Z}_N\right]^{1/2^r},$$

where $\mathbf{x} = (x_1, \dots, x_r)$. The following inequality, referred to as a generalized von Neumann theorem, bounds the weighted count of $(r+1)$ -term arithmetic progressions from functions f_0, \dots, f_r in terms of the Gowers uniformity norm:

$$\left|\mathbb{E}\left[f_0(x)f_1(x+d)f_2(x+2d)\cdots f_r(x+rd) \middle| x, d \in \mathbb{Z}_N\right]\right| \leq \|f_j\|_{U^r} \prod_{i \neq j} \|f_i\|_{\infty}.$$

This fundamental fact is an important starting point for Gowers' celebrated proof [17] of Szemerédi's theorem as well as many later developments in additive combinatorics. For a sparse subset $S \subseteq \mathbb{Z}_N$ of density p , this inequality implies the correct count of $(r+1)$ -term arithmetic progressions in S as long as $\|\nu - 1\|_{U^r} = o(p^r)$, where $\nu = p^{-1}1_S$ (a more careful analysis shows that it suffices to assume $\|\nu - 1\|_{U^r} = o(p^{r/2})$).

Gowers⁶ [19] and Green [20] asked if $\|\nu - 1\|_{U^s} = o(1)$ for some large $s = s(r)$ is sufficient for ν to satisfy a relative Szemerédi theorem for $(r + 1)$ -term arithmetic progressions. Note that this is precisely a linear forms condition and we proved in this paper that a different linear forms condition is sufficient. However, we do not even know if such a condition implies the existence of $(r + 1)$ -term arithmetic progressions. In fact, for $r = 2$ and $s = 2$, it follows from the recent work of Bennett [2] on the random arithmetic-progression-free process that this uniformity norm condition is not enough to imply a 3-term arithmetic progression. In a companion note [8], we show that if a measure ν satisfies the stronger condition $\|\nu - 1\|_{U^r} = o(p^r)$, where $p = \|\nu\|_\infty^{-1}$, then the relative Szemerédi theorem holds with respect to ν for $(r + 1)$ -term arithmetic progressions. This strengthens the consequence of the generalized von Neumann theorem discussed above.

Sparse graph limits. The regularity method played a fundamental role in the development of the theory of dense graph limits [4, 28]. However, no satisfactory theory of graph limits is known for graphs with edge density $o(1)$. Bollobás and Riordan [3] asked a number of questions and made explicit conjectures on suitable conditions for sparse graph limits and counting lemmas. Our work gives some natural sufficient conditions for obtaining a counting lemma in a sequence of sparse graphs G_N . The new counting lemma allows us to transfer the results of Lovász and Szegedy [28, 29] on the existence of the limit graphon, as well as the results of Borgs, Chayes, Lovász, Sós, and Vesztegombi [4] on the equivalence of left-convergence (i.e., convergence in homomorphism densities) and convergence in cut distance. The famous quasirandomness results of Chung, Graham, and Wilson [5] also transfer, namely, that an appropriate relationship between edge density and C_4 -density (of homomorphisms) determines the asymptotic F -density for every graph F . We will explain these connections in more detail in an upcoming survey article [9].

Existing applications of the Green-Tao method. Though our discussion has focused on the relative Szemerédi theorem, we have proved a relative version of the stronger multidimensional Szemerédi theorem. Following Tao [45], this may be used to prove that the Gaussian primes contain arbitrarily shaped constellations, though without the need to verify either the correlation condition or the dual function condition. It seems likely that our method could also be useful for simplifying several other papers where the machinery of Green and Tao is used [12, 23, 27, 30, 31, 47]. In some cases it should be possible to use our results verbatim but in others, such as the paper of Tao and Ziegler [47] proving that there are arbitrarily long polynomial progressions in the primes, it will probably require substantial additional work.

Sparse hypergraph regularity. In proving a hypergraph removal lemma for subgraphs of pseudorandom hypergraphs, we have developed a general approach to regularity and counting in sparse pseudorandom hypergraphs which has the potential for much broader application. It is, for example, quite easy to use our results to prove analogues of well-known combinatorial theorems such as Ramsey’s theorem and Turán’s theorem relative to sparse pseudorandom hypergraphs of density N^{-c_H} . We omit the details. In the graph case, a number of further applications were discussed in [7]. We expect that hypergraph versions of many of these applications should be an easy corollary of our results.

Counting in random hypergraphs. There has been much recent work on counting lemmas and relative versions of combinatorial theorems within random graphs and hypergraphs [1, 10, 11, 37, 38]. Surprisingly, there are a number of disparate approaches to these problems, each having its own strengths and weaknesses. We believe that our results can be used to give an alternative framework for one of these approaches, due to Conlon and Gowers [10].⁷ Their proof relies heavily upon an application of the Green-Tao transference theorem, which we believe can be replaced

⁶This question can be found in the penultimate paragraph in §4 of the arXiv version of [19].

⁷This should at least be true for theorems regarding graphs and hypergraphs, though we feel that a similar approach should also be possible for subsets of the integers.

with an application of the sparse Frieze-Kannan regularity lemma and our densification technique. However, the key technical step in [10], which in our language is to verify that the strong linear forms condition, Lemma 6.3, holds when ν is a random measure, would remain unchanged.

Sparse arithmetic removal. In Theorem 3.2, we proved an arithmetic removal lemma for linear patterns such as arithmetic progressions. More generally, an arithmetic removal lemma claims that if a system of linear equations $Ma = b$ over the integers has a small number of solutions $a = (a_1, a_2, \dots, a_n)$ with $a_i \in A_i$ for all $i = 1, 2, \dots, n$ then one may remove a small number of elements from each A_i to find subsets A'_i such that there are no solutions $a' = (a'_1, a'_2, \dots, a'_n)$ to $Ma' = b$ with $a'_i \in A'_i$ for all $i = 1, 2, \dots, n$. Such a result was conjectured by Green [21] and proved by Král', Serra, and Vena [26] and, independently, Shapira [40]. Both of these proofs are based upon representing a system of linear equations by a hypergraph and deducing the arithmetic removal lemma from a hypergraph removal lemma. Such an idea was first used by Král', Serra, and Vena [25] with graphs (instead of hypergraphs). In [7], we adapted the arguments of [25] to sparse pseudorandom subsets of the integers using the removal lemma in sparse pseudorandom graphs. Likewise, our results on hypergraph removal in this paper may be used to prove a sparse pseudorandom generalization of the arithmetic removal lemma [26, 40] for all systems of linear equations.

Acknowledgements. We would like to thank Tom Bohman and Ben Green for helpful discussions.

REFERENCES

- [1] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. arXiv:1204.6530.
- [2] P. Bennett. In preparation.
- [3] B. Bollobás and O. Riordan. Metrics for sparse graphs. In *Surveys in combinatorics 2009*, volume 365 of *London Math. Soc. Lecture Note Ser.*, pages 211–287. Cambridge Univ. Press, Cambridge, 2009.
- [4] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztegombi. Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.
- [5] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.
- [6] A. Coja-Oghlan, C. Cooper, and A. Frieze. An efficient sparse regularity concept. *SIAM J. Discrete Math.*, 23(4):2000–2034, 2009/10.
- [7] D. Conlon, J. Fox, and Y. Zhao. Extremal results in sparse pseudorandom graphs. arXiv:1204.6645.
- [8] D. Conlon, J. Fox, and Y. Zhao. Linear forms from the Gowers uniformity norm. Unpublished companion note.
- [9] D. Conlon, J. Fox, and Y. Zhao. The sparse regularity method. In preparation.
- [10] D. Conlon and W. T. Gowers. Combinatorial theorems in sparse random sets. arXiv:1011.4310.
- [11] D. Conlon, W. T. Gowers, W. Samotij, and M. Schacht. On the KLR conjecture in random graphs. arXiv:1305.2516.
- [12] B. Cook and A. Magyar. Constellations in \mathbb{P}^d . *Int. Math. Res. Not. IMRN*, 2012(12):2794–2816, 2012.
- [13] P. Frankl and V. Rödl. Extremal problems on set systems. *Random Structures Algorithms*, 20(2):131–164, 2002.
- [14] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.
- [15] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.*, 34:275–291 (1979), 1978.
- [16] D. A. Goldston and C. Y. Yıldırım. Higher correlations of divisor sums related to primes. I. Triple correlations. *Integers*, 3:A5, 66, 2003.
- [17] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.
- [18] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math.*, 166(3):897–946, 2007.
- [19] W. T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.*, 42(4):573–606, 2010. arXiv:0811.3103.
- [20] B. Green. Personal communication.
- [21] B. Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005.
- [22] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, 167(2):481–547, 2008.
- [23] B. Green and T. Tao. Linear equations in primes. *Ann. of Math.*, 171(3):1753–1850, 2010.

- [24] Y. Kohayakawa. Szemerédi’s regularity lemma for sparse graphs. In *Foundations of computational mathematics (Rio de Janeiro, 1997)*, pages 216–230. Springer, Berlin, 1997.
- [25] D. Král’, O. Serra, and L. Vena. A combinatorial proof of the removal lemma for groups. *J. Combin. Theory Ser. A*, 116(4):971–978, 2009.
- [26] D. Král’, O. Serra, and L. Vena. A removal lemma for systems of linear equations over finite fields. *Israel J. Math.*, 187:193–207, 2012.
- [27] T. H. Lê. Green-Tao theorem in function fields. *Acta Arith.*, 147(2):129–152, 2011.
- [28] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.
- [29] L. Lovász and B. Szegedy. Szemerédi’s lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.
- [30] L. Matthiesen. Correlations of the divisor function. *Proc. Lond. Math. Soc.*, 104(4):827–858, 2012.
- [31] L. Matthiesen. Linear correlations amongst numbers represented by positive definite binary quadratic forms. *Acta Arith.*, 154(3):235–306, 2012.
- [32] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k -uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.
- [33] O. Reingold, L. Trevisan, M. Tulsiani, and S. Vadhan. Dense subsets of pseudorandom sets. In *49th Annual IEEE Symposium on Foundations of Computer Science*, pages 76–85. IEEE Computer Society, 2008.
- [34] V. Rödl and J. Skokan. Regularity lemma for k -uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.
- [35] V. Rödl and J. Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures Algorithms*, 28(2):180–194, 2006.
- [36] 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*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam, 1978.
- [37] D. Saxton and A. Thomason. Hypergraph containers. arXiv:1204.6595.
- [38] M. Schacht. Extremal results for random discrete structures. Submitted.
- [39] A. Scott. Szemerédi’s regularity lemma for matrices and sparse graphs. *Combin. Probab. Comput.*, 20(3):455–466, 2011.
- [40] A. Shapira. A proof of Green’s conjecture regarding the removal properties of sets of linear equations. *J. Lond. Math. Soc.*, 81(2):355–373, 2010.
- [41] J. Solymosi. Note on a generalization of Roth’s theorem. In *Discrete and computational geometry*, volume 25 of *Algorithms Combin.*, pages 825–827. Springer, Berlin, 2003.
- [42] J. Solymosi. A note on a question of Erdős and Graham. *Combin. Probab. Comput.*, 13(2):263–267, 2004.
- [43] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.
- [44] T. Tao. A remark on Goldston-Yıldırım correlation estimates. Unpublished.
- [45] T. Tao. The Gaussian primes contain arbitrarily shaped constellations. *J. Anal. Math.*, 99:109–176, 2006.
- [46] T. Tao. A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A*, 113(7):1257–1280, 2006.
- [47] T. Tao and T. Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Math.*, 201(2):213–305, 2008.
- [48] H. Towsner. An analytic approach to sparse hypergraphs: hypergraph removal. arXiv:1204.1884.
- [49] L. Trevisan, M. Tulsiani, and S. Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *24th Annual IEEE Conference on Computational Complexity*, pages 126–136. IEEE Computer Society, 2009.

MATHEMATICAL INSTITUTE, OXFORD OX1 3LB, UNITED KINGDOM
E-mail address: david.conlon@maths.ox.ac.uk

DEPARTMENT OF MATHEMATICS, MIT, CAMBRIDGE, MA 02139-4307
E-mail address: fox@math.mit.edu

DEPARTMENT OF MATHEMATICS, MIT, CAMBRIDGE, MA 02139-4307
E-mail address: yufeiz@math.mit.edu