

ERGODIC-THEORETIC IMPLEMENTATIONS OF THE ROTH DENSITY-INCREMENT ARGUMENT

TIM AUSTIN

ABSTRACT. We exhibit proofs of Furstenberg’s Multiple Recurrence Theorem and of a special case of Furstenberg and Katznelson’s multidimensional version of this theorem, using an analog of the density-increment argument of Roth and Gowers. The second of these results requires also an analog of some recent finitary work by Shkredov.

Many proofs of these multiple recurrence theorems are already known. However, the approach of this paper sheds some further light on the well-known heuristic correspondence between the ergodic-theoretic and combinatorial aspects of multiple recurrence and Szemerédi’s Theorem. Focusing on the density-increment strategy highlights several close points of connection between these settings.

CONTENTS

1. Introduction	1
1.1. Ergodic Ramsey Theory	2
1.2. The density-increment argument	3
1.3. Outline of this note	5
2. Powers of a single transformation	6
2.1. Preliminary discussion	6
2.2. The density-increment proof	11
3. Two commuting transformations	16
3.1. The density increment in higher dimensions	16
3.2. A closer look at the characteristic factors and the main estimate	19
3.3. The main estimate	22
3.4. Shkredov’s version of the density increment	24
4. Further discussion	28
References	31

1. INTRODUCTION

In 1975 Szemerédi published the first proof of a long-standing conjecture of Erdős and Turán concerning arithmetic progressions in dense arithmetic sets.

Work supported by fellowships from Microsoft Corporation and from the Clay Mathematics Institute.

Theorem 1.1 (Szemerédi’s Theorem). *If $E \subset \mathbb{Z}$ admits some $\delta > 0$ for which there are arbitrarily long intervals $[M, N]$ with*

$$|E \cap [M, N]| \geq \delta(N - M)$$

(that is, E has ‘upper Banach density’ equal to at least δ), then E also contains for every $k \geq 1$ a nondegenerate arithmetic progression of length k :

$$E \supset \{a, a + n, a + 2n, \dots, a + (k - 1)n\} \quad \text{for some } a \in \mathbb{Z}, n \geq 1.$$

Separate proofs for various special cases were given earlier by Roth and by Szemerédi himself. The thirty-five years subsequent to Szemerédi’s breakthrough have seen the emergence of a host of alternative approaches to this theorem and several generalizations.

The many techniques that have been brought to bear in this investigation are loosely drawn from three areas of mathematics:

- graph and hypergraph theory (in work of Szemerédi, Solymosi, Nagle, Rödl, Schacht, Skokan, Gowers and others),
- ergodic theory (largely building on ideas of Furstenberg and Katznelson),
- harmonic analysis (following Roth, Bourgain, Gowers, Green, Tao and Shkredov).

The alternative arguments constructed from these three bodies of theory sometimes correspond much more closely than is initially apparent, owing to many differences in technical detail that turn out to be quite superficial. No really comprehensive overview of the relations among these approaches is yet available, but fragments of the picture can be found in the papers [Kra07, GT12, Tao06a] and in Chapters 10 and 11 of Tao and Vu’s book [TV06].

The purpose of the present note is to extract one aspect of the harmonic analytic approach — the ‘density-increment argument’, originating in the early work of Roth [Rot53] — and present a natural analog of it in the rather different setting of ergodic theory. No new theorems will be proved except for some technical results needed on route, but I hope that this alternative presentation of existing ideas will contribute to enhancing the toolkits of those working on this class of problems, and also shed some light on the open questions that remain concerning the density-increment approach.

1.1. Ergodic Ramsey Theory. Two years after Szemerédi’s proof of Theorem 1.1 appeared, Furstenberg offered in [Fur77] a very different approach to the same result based on a conversion to a problem in ergodic theory, using what is now referred to as ‘Furstenberg’s correspondence principle’.

A precise formulation of the general correspondence principle can be found, for example, in Bergelson [Ber96]. Here we simply recall that Furstenberg proved the equivalence of Szemerédi’s Theorem to the following:

Theorem 1.2 (Multiple Recurrence Theorem). *If $T : \mathbb{Z} \curvearrowright (X, \mu)$ is a probability-preserving action on a standard Borel probability space and $A \subset X$ is measurable*

and has $\mu(A) > 0$, then also

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-(k-1)n}A) > 0 \quad \forall k \geq 1.$$

Furstenberg's proof of Theorem 1.2 relied on a powerful structural classification of probability-preserving dynamical systems developed independently by Furstenberg and by Zimmer ([Zim76b, Zim76a]).

Shortly after that proof appeared, Furstenberg and Katznelson realized that only a modest adaptation yields a significantly stronger result.

Theorem 1.3 (Multidimensional Multiple Recurrence Theorem). *If $T_1, T_2, \dots, T_d : \mathbb{Z} \curvearrowright (X, \mu)$ are commuting probability-preserving actions on a standard Borel probability space and $A \subset X$ has $\mu(A) > 0$ then also*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T_1^{-n}A \cap \dots \cap T_d^{-n}A) > 0.$$

This appeared in [FK78]. Theorem 1.2 follows from Theorem 1.3 by setting $d := k - 1$ and $T_i := T^i$ for $i \leq k - 1$. On the other hand, Theorem 1.3 also has a combinatorial consequence that strengthens Szemerédi's Theorem:

Theorem 1.4. *If $E \subset \mathbb{Z}^d$ admits some $\delta > 0$ for which there are cuboids $\prod_{i \leq d} [M_i, N_i]$ with $\min_{i \leq d} |N_i - M_i|$ arbitrarily large and*

$$\left| E \cap \prod_{i \leq d} [M_i, N_i] \right| \geq \delta \prod_{i \leq d} (N_i - M_i),$$

then E also contains the set of vertices of a nondegenerate upright right-angled isosceles simplex:

$$E \supset \{\mathbf{a}, \mathbf{a} + n\mathbf{e}_1, \dots, \mathbf{a} + n\mathbf{e}_d\} \quad \text{for some } \mathbf{a} \in \mathbb{Z}^d, n \geq 1.$$

Interestingly, this result went unproven by purely combinatorial means until the development of hypergraph analogs of Szemerédi's famous Regularity Lemma by Nagle, Rödl and Schacht [NRS06], Gowers [Gow07] and [Tao06b], more than twenty years later. In addition, several other purely combinatorial assertions have now been accessed though 'Ergodic Ramsey Theory', the subject that emerged from Furstenberg and Katznelson's early developments, including a density version of the Hales-Jewett Theorem [FK91] and a density Ramsey Theorem for subtrees of trees [FW03].

Within ergodic theory, a great deal of energy has now been spent on obtaining the most precise possible understanding of the averages whose limit infima are the subjects of Theorems 1.2 and 1.3; we will return to some of these developments later.

1.2. The density-increment argument. The 'density-increment argument' was first used by Roth for his early proof of the case $k = 3$ of Theorem 1.1. Much more recently, Gowers developed in [Gow98, Gow01] an extremely sophisticated

extension of Roth's approach, and using this was able to give a density-increment proof of the full Szemerédi Theorem.

We will not spend time here on the many technical accomplishments involved in Gowers' work, requiring a call to tools from yet other parts of arithmetic combinatorics such as Freiman's Theorem. Rather we record just a simple statement of the density-increment proposition that lies at its heart.

Proposition 1.5. *Suppose that $\delta > 0$, that N is sufficiently large and that $E \subset \{1, 2, \dots, N\}$ has $|E| \geq \delta N$ but contains no k -term arithmetic progression. Then there is an arithmetic progression $P \subset \{1, 2, \dots, N\}$ of size at least $N^{((\delta/2)^{k2^k})^{2^{k+8}}}$ such that*

$$|E \cap P| \geq (\delta + ((\delta/2)^{k2^k})^{2^{k+8}})|P|.$$

□

This proposition is implicit in [Gow01], but does not appear in the above form because Gowers presents his argument in terms of the crucial auxiliary notion of 'higher-degree uniformity', and splits the above result into several pieces that are connected via this auxiliary notion.

This kind of uniformity is defined in terms of the Gowers uniformity norms (Section 3 of [Gow01]; see also Chapter 11 of Tao and Vu [TV06]) that have since become widely used in additive combinatorics. Uniformity of degree 1 can be described simply in terms of the presence of some large values among the Fourier coefficients of 1_E , regarded as a function on the group $\mathbb{Z}/N\mathbb{Z}$; this is essentially the notion that Roth uses in his approach for $k = 3$. Higher-degree uniformity extends this property, although it is not so easily described using Fourier analysis. In his more general setting, Gowers proves on the one hand that if E is sufficiently uniform of degree $k - 2$ then it contains a k -term arithmetic progression (Corollary 3.6 in [Gow01]), and on the other that if E is not sufficiently uniform of degree $k - 2$ then we may partition $\{1, 2, \dots, N\}$ into long arithmetic subprogressions such that E has a relative density inside some of these subprogressions that is substantially larger than δ (Theorem 18.1 in [Gow01]). This fact can then be used to pick out one such subprogression satisfying the above conclusion (Lemma 5.15 in [Gow01]). Proposition 1.5 amounts to the conjunction of these facts.

Our proof of Theorem 1.2 below takes a similar form (although we should stress that our task is much simpler than Gowers'), using an ergodic-theoretic analog of the notion of 'uniformity' arising in work of Host and Kra. Similarly to the presentation in [Gow01], we will find that handling the consequences of non-uniformity is the more complicated of the two steps involved.

From Proposition 1.5 a proof of Szemerédi's Theorem follows quickly by contradiction. If E is a counterexample of density δ and N is sufficiently large, then for a subprogression P as given by Proposition 1.5 we see that $E \cap P$, identified with a subset of $\{1, 2, \dots, |P|\}$ by the obvious affine map, is another counterexample with density that exceeds δ by an amount depending only on δ and k . It is contained in a discrete interval of length $|P| = N^{\kappa(\delta, k)}$ for some small fixed $\kappa(\delta, k) > 0$. Therefore, provided N was sufficiently large to begin with, iterating

this construction must eventually turn a counterexample of density at least δ into a counterexample of density greater than 1: an obvious contradiction.

In addition to its aesthetic value, Gowers' new proof of Szemerédi's Theorem gives much the best known bound on how large N must be taken in order that a k -term arithmetic progression is certain to be found in a density- δ subset $E \subset \{1, 2, \dots, N\}$. In view of this, it was natural to ask whether this approach could also be brought to bear on the multidimensional Theorem 1.4 in order to give a similarly striking improvement to the bounds available there. Gowers poses this problem explicitly and offers some discussion of it in his survey [Gow00]. Recently Shkredov has made the first serious progress on this problem by essentially solving the case $d = 2$ in [Shk06b], applying some important new technical ideas that are needed to prove and then use a relative of Proposition 1.5. However, a further enhancement of these ideas that will yield a density-increment proof of the full Theorem 1.4, with or without improved bounds, still seems relatively distant.

1.3. Outline of this note. The centrepieces of this note are 'density-increment' proofs of the Multiple Recurrence Theorem 1.2 and the case $d = 2$ of Theorem 1.3, corresponding to Gowers' and Shkredov's combinatorial implementations of the density-increment argument respectively.

The main steps taken by Gowers and Shkredov do have counterparts in these proofs, but we need different structural results from within ergodic theory to enable them. These will largely be drawn from recent studies of the 'nonconventional ergodic averages' whose limit infima appear in Theorems 1.2 and 1.3. In particular we rely on the method of 'characteristic factors', which has emerged through the works of several researchers since Furstenberg's original paper [Fur77], and especially on some of the technical steps in Host and Kra's proof ([HK05]) of convergence for the averages of 1.2 and in the work of Conze and Lesigne [CL84, CL88a, CL88b] and the subsequent works [Aus09, Aus10a] on the multi-dimensional case. Many other researchers have contributed to this story within ergodic theory, including Rudolph, Zhang, Katznelson, Weiss, Ziegler, Frantzikinakis and Chu, and the reader is referred to [Aus10b] for a more complete discussion.

The basic ergodic theoretic version of the density-increment argument for Theorem 1.2 will be introduced in Subsection 2.2 and then used to complete the proof of that theorem later in Section 2. Although a density increment is central to Shkredov's proof as well, he uses it in a slightly more complicated way, and so in Section 3 we introduce the ergodic theoretic analog of this separately and then use it to prove the case $d = 2$ of Theorem 1.3.

On the one hand, I hope that these proofs shed some light on the nature of the density-increment argument. On the other, it seems that recent progress in ergodic theory is beginning to address some of the problems of extending this approach to give a density-increment proof of the whole of Theorem 1.3 (and so, one might hope, also to give a finitary density-increment proof of Theorem 1.4, as requested by Gowers). In the final Section 4 we will draw on results from [Aus10b, AUSA, Ausb] to sketch some of the further developments suggested by this progress.

2. POWERS OF A SINGLE TRANSFORMATION

2.1. Preliminary discussion. In this section we show how the density-increment strategy can be used to give a proof of Theorem 1.2, building on two important ergodic-theoretic ingredients. Let us first recall a convenient definition.

Definition 2.1 (Process). *We will refer to a probability-preserving \mathbb{Z} -system (X, μ, T) together with a distinguished subset A as a **process** and denote it by $(X \supset A, \mu, T)$.*

Definition 2.2. *An ergodic process $(X \supset A, \mu, T)$ has **no k -APs in its return times** if*

$$\mu(A \cap T^{-n}A \cap \dots \cap T^{-(k-1)n}A) = 0 \quad \text{for all } n \in \mathbb{Z} \setminus \{0\}.$$

Clearly if $\mu(A) > 0$ then the above property is stronger than being a counterexample to the Multiple Recurrence Theorem, which requires that the relevant intersections have positive measure on average, not just for a single nonzero n . Since that theorem turns out to be true, the above definition is essentially vacuous, but it will be a convenient handle at various points during the proofs that follow.

The first ingredient we need is a corollary of the recent result of Host and Kra [HK05] that the limiting values of the multiple recurrence averages are precisely controlled by certain special nilrotation factors of a system (X, μ, T) .

Definition 2.3 (Nilrotations). *For any $k \geq 1$ a **k -step nilrotation** is a \mathbb{Z} -system on a homogeneous space G/Γ for G a k -step nilpotent Lie group and $\Gamma \leq G$ a cocompact discrete subgroup, where G/Γ is endowed with its normalized Haar measure m and the transformation is given by*

$$R_g : h\Gamma \mapsto gh\Gamma$$

for some $g \in G$.

Theorem 2.4 (Host-Kra Theorem). *For each $k \geq 2$, any ergodic \mathbb{Z} -system $\mathbf{X} = (X, \mu, T)$ has a factor map $\pi_{k-2} : \mathbf{X} \rightarrow \mathbf{Z}_{k-2}$ onto a system generated by an inverse sequence of $(k-2)$ -step nilrotations such that*

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int_X f_0 \cdot (f_1 \circ T^n) \cdot \dots \cdot (f_{k-1} \circ T^{(k-1)n}) \, d\mu \\ & \sim \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}_\mu(f_0 | \pi_{k-2}) \cdot (\mathbb{E}_\mu(f_1 | \pi_{k-2}) \circ T^n) \cdot \dots \cdot (\mathbb{E}_\mu(f_{k-1} | \pi_{k-2}) \circ T^{kn}) \, d\mu \end{aligned}$$

as $N \rightarrow \infty$ for any $f_0, f_1, \dots, f_{k-1} \in L^\infty(\mu)$, where the notation asserts that the difference between these two sequences of averages tends to 0 as $N \rightarrow \infty$. \square

Remark. The above result is often expressed by asserting that the factor π_{k-2} is **characteristic** for the averages in question. This theorem first appears in [HK05], where its proof invokes a family of seminorms on $L^\infty(\mu)$ that Host and Kra introduce for this purpose and that are closely analogous to Gowers' uniformity seminorms from [Gow01], so offering another point of proximity between the ergodic theoretic and quantitative approaches. Another proof of Theorem 2.4 has now been

given by Ziegler in [Zie07], who also shows that the maximal factor of \mathbf{X} generated by $(k - 2)$ -step nilrotations is also the unique minimal factor that is characteristic in the above sense. \triangleleft

With the Host-Kra Theorem in mind, the second result that we use is simply the fact that multiple recurrence does hold for nilrotations.

Theorem 2.5 (Multiple recurrence for nilrotations). *If $R_g \curvearrowright G/\Gamma$ is an ergodic nilrotation, $A \subset G/\Gamma$ has positive measure and $K \geq 1$ then there is some $r \geq 1$ such that*

$$m(g^{-Kr} A \cap g^{-(K-1)r} A \cap \dots \cap g^{Kr} A) > 0.$$

\square

In fact, this result is considerably simpler than Theorem 2.4, which really does the heavy lifting in what follows. The point is that the orbit of the diagonal $\Delta := \{(x, x, \dots, x) : x \in G/\Gamma\} \subset (G/\Gamma)^{2K+1}$ (or rather, its normalized surface measure m_Δ) under the off-diagonal transformation $R_{(g^{-K}, g^{-K+1}, \dots, g^K)}$ (which is clearly still a nilrotation acting on $(G/\Gamma)^{2K+1}$) can be shown to equidistribute in some finite union of closed connected nilsubmanifolds of $(G/\Gamma)^{2K+1}$ that contains the whole of this diagonal set. This follows from strong results classifying all ergodic invariant measures for nilrotations. From this point a fairly elementary argument gives the positivity of

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N m(g^{-Kn} A \cap g^{-(K-1)n} A \cap \dots \cap g^{Kn} A) \\ &= \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int 1_{A \times A \times \dots \times A} \circ R_{(g^{-K}, g^{-K+1}, \dots, g^K)}^n \, dm_\Delta, \end{aligned}$$

and also the fact that these averages actually converge, so this limit infimum is really a limit. A related instance of this argument can be found presented in detail in Section 2 of the work [BLL08] by Bergelson, Leibman and Lesigne, who use it for the related end of proving multiple recurrence along certain families of polynomials. Equidistribution results for nilrotations on which this reasoning can be founded are available in either Ziegler [Zie05] or Bergelson, Host and Kra [BHK05], which in turn build on older works of Parry [Par69, Par70, Par73], Lesigne [Les91] and Leibman [Lei98, Lei05].

With Theorems 2.4 and 2.5 at our disposal, it is relatively easy to lay out a density-increment proof of the full Multiple Recurrence Theorem. However, it is important to observe right away that this is a rather perverse thing to do, because the above two ingredients also imply that theorem through the following even quicker argument:

- given our process $(X \supset A, \mu, T)$, we wish to prove that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n} A \cap \dots \cap T^{-(k-1)n} A) > 0,$$

so by Theorem 2.4 it suffices to prove instead that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X \prod_{i=0}^{k-1} (\mathbb{E}(A | \pi_{k-2}) \circ T^{in}) d\mu > 0$$

with $\pi_{k-2} : \mathbf{X} \rightarrow \mathbf{Z}_{k-2}$ the inverse limit of nilrotation factors from that theorem (and where we write $\mathbb{E}(A | \pi_{k-2})$ as short for $\mathbb{E}(1_A | \pi_{k-2})$);

- this, in turn, will follow if we prove that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(B \cap T^{-n}B \cap \dots \cap T^{-(k-1)n}B) > 0$$

where $B := \{\mathbb{E}(A | \pi_{k-2}) > \varepsilon\}$ for any positive ε chosen so small that $\mu(B) > 0$ (for example, $\varepsilon \leq \mu(A)/2$ will do);

- finally, importing a simple trick from [FK78], this follows by choosing a further factor $\alpha : Z_{k-2} \rightarrow G/\Gamma$ onto a finite-dimensional nilrotation such that

$$\|\mathbb{E}(A | \pi_{k-2}) - \mathbb{E}(A | \alpha \circ \pi_{k-2})\|_1 < \frac{\varepsilon}{100(k+1)}$$

(which is possible because π_{k-2} is generated by an inverse sequence of such further factors α) and letting

$$C := \left\{ \mathbb{E}(B | \alpha \circ \pi_{k-2}) > 1 - \frac{1}{k+1} \right\},$$

for which we now easily deduced that

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(B \cap T^{-n}B \cap \dots \cap T^{-(k-1)n}B) \\ \geq \frac{1}{2} \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(C \cap T^{-n}C \cap \dots \cap T^{-(k-1)n}C), \end{aligned}$$

which Theorem 2.5 shows is strictly positive.

(This proof is also essentially that used in [BLL08] for their instance of polynomial recurrence.)

Therefore the point of this section is not to provide a serious new approach to the Multiple Recurrence Theorem, but rather to exhibit the density-increment strategy in a setting familiar to ergodic theorists.

The reason why the approach to multiple recurrence just sketched does not have a clear analog among quantitative proofs of Szemerédi's Theorem is hidden in our appeal to Theorem 2.4. In fact, the technical result that drives Gowers' work is really more analogous to the following easy corollary of Theorem 2.4 than to Theorem 2.4 itself:

Corollary 2.6. *If an ergodic system (X, μ, T) and measurable functions $f_0, f_1, \dots, f_{k-1} : X \rightarrow [-1, 1]$ are such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f_0 \cdot (f_1 \circ T^n) \cdots (f_{k-1} \circ T^{(k-1)n}) \, d\mu \geq \gamma > 0,$$

then there is a factor map $\pi : (X, \mu, T) \rightarrow (G/\Gamma, m, R_g)$ onto an ergodic $(k-2)$ -step nilrotation such that

$$\|\mathbb{E}(f_i | \pi)\|_2 \geq \frac{1}{2}\gamma \quad \text{for each } i = 0, 1, \dots, k-1.$$

In particular, if $(X \supset A, \mu, T)$ is a process with $\mu(A) \geq \delta$ but no k -APs in its return times then there is such a factor map π for which

$$\|\mathbb{E}(A | \pi) - \mu(A)\|_2 \geq \frac{1}{2k}\delta^k.$$

Proof. By Theorem 2.4, the averages in question have the same asymptotic behaviour as the averages

$$\frac{1}{N} \sum_{n=1}^N \int_X f'_0 \cdot (f'_1 \circ T^n) \cdots (f'_{k-1} \circ T^{(k-1)n}) \, d\mu$$

with $f'_i := \mathbb{E}(f_i | \pi_{k-2})$, and now all these functions still lie in the unit ball of $L^\infty(\mu)$, and so for any i we can apply the Cauchy-Schwartz inequality to $f'_i \circ T^{in}$ and the product of the remaining factors to deduce that the above average is bounded in absolute value by $\|f'_i\|_2$. Since the sum of these averages must be greater than $\gamma/2$ infinitely often, this requires that $\|f'_i\|_2 > \gamma/2$ for each i ; finally, since π_{k-2} is generated by further factor maps onto finite-dimensional $(k-2)$ -step nilrotations, letting π be a large enough one of these gives the first conclusion.

To derive the second conclusion, first use Theorem 2.4 to obtain

$$\frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A) \sim \frac{1}{N} \sum_{n=1}^N \int_X \prod_{i=0}^{k-1} \mathbb{E}(A | \pi_{k-2}) \circ T^{in} \, d\mu$$

as $N \rightarrow \infty$, so that if A contains no k -APs in its return times then both of these expressions must vanish as $N \rightarrow \infty$. Now use the decomposition

$$\mathbb{E}(A | \pi_{k-2}) = (\mathbb{E}(A | \pi_{k-2}) - \mu(A)) + \mu(A)$$

to form the telescoping sum

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int_X \prod_{i=0}^{k-1} \mathbb{E}(A | \pi_{k-2}) \circ T^{in} \, d\mu \\ &= \sum_{i=0}^{k-1} \frac{1}{N} \sum_{n=1}^N \int_X \mu(A)^i \cdot ((\mathbb{E}(A | \pi_{k-2}) \circ T^{in} - \mu(A)) \\ & \quad \cdot \prod_{i < j \leq k-1} (\mathbb{E}(A | \pi_{k-2}) \circ T^{jn}) \, d\mu \\ & \quad + \mu(A)^k. \end{aligned}$$

Since this tends to 0, the first k terms of the sum must asymptotically cancel the term $\mu(A)^k \geq \delta^k$, and hence at least one of these first terms must have magnitude arbitrarily close to $\frac{1}{k}\delta^k$ for infinitely many N . The first part of the corollary therefore gives some $(k-2)$ -step nilrotation factor π for which

$$\|\mathbb{E}(A | \pi) - \mu(A)\|_2 \geq \frac{1}{2k}\delta^k,$$

as required. \square

Within Roth's and Gowers' works lie quantitative analogs of the above result: this is what drives Gowers' proof that a failure of uniformity of degree $k-2$ for a density- δ set $E \subset \{1, 2, \dots, N\}$ gives a partition of $\{1, 2, \dots, N\}$ into fairly long subprogressions on which E enjoys an enlarged relative density (Theorem 18.1 in [Gow01]).

Heuristically, Gowers shows first that a failure of uniformity of degree $k-2$ implies a nontrivial correlation between $1_E - \delta$ and a function on $\{1, 2, \dots, N\}$ which behaves like the exponential of i times a real polynomial of degree $(k-2)$ on many large subprogressions of $\{1, 2, \dots, N\}$. He then converts this correlation into the desired partition of $\{1, 2, \dots, N\}$ into long subprogressions. This correlation with a function that behaves 'locally' like a degree- $(k-2)$ polynomial is the analog of having a nontrivial conditional expectation onto a $(k-2)$ -step nilsystem. The exact formulation of the finitary 'inverse theorem' for the failure of higher-degree uniformity is rather complicated, and we omit it here, but again a gentle introduction with many further references can be found in the book [TV06] of Tao and Vu.

In the infinitary ergodic-theoretic setting the implication of Corollary 2.6 by Theorem 2.4 can easily be reversed: given any indicator function 1_A we can decompose it as $1_A = (1_A - \mathbb{E}(A | \pi_{k-2})) + \mathbb{E}(A | \pi_{k-2})$, and now if we form a telescoping sum for the expression $\frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-(k-1)n}A)$ similar to the above then (the contrapositive of) Corollary 2.6 implies that all the terms involving $1_A - \mathbb{E}(A | \pi_{k-2})$ must vanish as $N \rightarrow \infty$, leaving us with Theorem 2.4:

$$\frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-(k-1)n}A) \sim \frac{1}{N} \sum_{n=1}^N \int_X \prod_{i=0}^{k-1} \mathbb{E}(A | \pi_{k-2}) \circ T^{in} d\mu.$$

However, difficulties emerge when one tries to develop a quantitative analog of this reverse implication, and so provide a truer analog of Theorem 2.4 in the finitary setting. In order to make sense of either conditional expectations such as $\mathbb{E}(A | \pi_{k-2})$, or of the structure of π_{k-2} as an inverse limit of a possibly-infinite collection of nilrotation factors, one needs a quantitative analog of taking a limit in $L^2(\mu)$. In practice this leads to an explosion in the bounds obtained. Although something in this vein is possible (see Tao [Tao06a]), in general it is much less efficient than the density-increment strategy, for which (the finitary analog of) Corollary 2.6 suffices. On the other hand, our quick presentation above of the deduction of multiple recurrence from Theorems 2.4 and 2.5 clearly uses the full strength of Theorem 2.4, and so if instead we started from Corollary 2.6 it would require us

to prove Theorem 2.4 first before proceeding as above. In the next subsection we will see that the density-increment strategy, by contrast, uses only the conjunction of Corollary 2.6 and Theorem 2.5, and it is this feature that accounts for its greater efficiency (leading to better bounds) in the finitary world.

2.2. The density-increment proof. The strategy here is to prove Theorem 1.2 or 1.3 by ‘induction on $\mu(A)$ ’. The technical result underlying this is an ergodic-theoretic analog of Proposition 1.5.

Proposition 2.7 (Ergodic-theoretic density-increment). *For each $k \geq 1$ there is a function $c_k : (0, 1] \rightarrow (0, 1]$ that is bounded away from 0 on compact subsets such that the following holds: if $(X \supset A, \mu, T)$ is a process with $\mu(A) = \delta > 0$ but no k -APs in its return times, then for every $\varepsilon > 0$ and $N \geq 1$ there are some non-negligible $B \subset X$ and integer $r \geq 1$ such that*

$$\mu(B \Delta T^r B) < \varepsilon \mu(B)$$

and

$$\mu(A | T^{-rn} B) \geq \delta + c_k(\delta) \quad \text{for all } -N \leq n \leq N.$$

Before proving this result let us see why it implies Theorem 1.2.

Corollary 2.8. *With c_k as in Proposition 2.7 the following holds: if there exists a process $(X \supset A, \mu, T)$ having $\mu(A) = \delta > 0$ but no k -APs in its return times, then there is another process $(Y \supset B, \nu, S)$ having $\nu(B) \geq \delta + c_k(\delta)$ but no k -APs in its return times.*

Proof of Corollary from Proposition 2.7. By Proposition 2.7, for any N we can find a non-negligible $B_N \subset X$ and $r_N \geq 1$ such that

$$\mu(B_N \Delta T^{r_N} B_N) < \frac{\mu(B_N)}{N}$$

and

$$\mu(A | T^{-r_N n} B_N) \geq \delta + c_k(\delta) \quad \text{for all } -N \leq n \leq N.$$

Let ν_N be the probability measure on $Y = \{0, 1\}^{\mathbb{Z}}$ that is the law of the random sequence

$$\varphi_N : x \mapsto (1_A(T^{r_N n}(x)))_{n \in \mathbb{Z}}$$

for x drawn at random from $\mu(\cdot | B_N)$ (that is, x is chosen ‘uniformly from B_N ’).

Let $S : Y \rightarrow Y$ be the coordinate left-shift and

$$A_a := \{(\omega_i)_{i \in \mathbb{Z}} : \omega_a = 1\} \subset Y \quad \text{for } a \in \mathbb{Z},$$

so $A_a = S^{-a}(A_0)$. The lower bound on the measures $\mu(A | T^{-r_N n} B_N)$ for $-N \leq n \leq N$ implies that any vague accumulation point ν of the sequence ν_N , say $\nu = \lim_{i \rightarrow \infty} \nu_{N_i}$, must satisfy

$$\nu(A_a) = \lim_{i \rightarrow \infty} \mu(A | T^{r_{N_i} a} B_{N_i}) \geq \delta + c_k(\delta) \quad \forall a.$$

On the other hand, the assumption that there are no k -APs in the return times of A implies that

$$\begin{aligned} \nu_N(A_a \cap A_{a+r} \cap \cdots \cap A_{a+(k-1)r}) \\ = \nu_N\{(\omega_i)_{i \in \mathbb{Z}} : \omega_a = \omega_{a+r} = \cdots = \omega_{a+(k-1)r} = 1\} = 0 \end{aligned}$$

for all $a \in \mathbb{Z}$, $r \geq 1$ and all N , and so the same is true for ν .

Finally, the inequality $\mu(B_N \Delta T^{rN} B_N) < \mu(B_N)/N$ implies for any Borel $C \subset Y$ that

$$\begin{aligned} |\nu_N(C) - \nu_N(S^{-1}C)| &= |\mu(\varphi_N^{-1}C | B_N) - \mu(T^{-rN}\varphi_N^{-1}C | B_N)| \\ &= |\mu(\varphi_N^{-1}C | B_N) - \mu(\varphi_N^{-1}C | T^{rN}B_N)| \\ &\leq \frac{\mu(\varphi_N^{-1}C \cap (B_N \Delta T^{rN}B_N))}{\mu(B_N)} < 1/N, \end{aligned}$$

so the vague limit ν is also S -invariant. Letting $B := A_0$, this gives a process $(Y \supset B, \nu, S)$ with no k -APs in its return times and the desired improved bounds. \square

Proof of Theorem 1.2 from Corollary 2.8. Step 1 If $(X \supset A, \mu, T)$ is any counterexample to Theorem 1.2 with $\delta_0 := \mu(A) > 0$, then a simple vague limit argument can enhance it to an example with the same density value δ_0 but no k -APs in its return times. This construction forms the bulk of this proof.

We first transfer our initially-given example onto the space $Y := \{0, 1\}^{\mathbb{Z}}$ with the left-shift S . Let

$$B := \{(\omega_i)_{i \in \mathbb{Z}} \in Y : \omega_0 = 1\},$$

and now consider the map

$$f_A : X \rightarrow Y : x \mapsto (1_A(T^i(x)))_{i \in \mathbb{Z}}.$$

This intertwines T with S , so the pushforward $\nu_1 := (f_A)_\# \mu$ is an S -invariant Borel measure on Y for which $\nu_1(B) = \mu(A)$ and

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \nu_1(B \cap S^{-n}B \cap \cdots \cap S^{-(k-1)n}B) \\ = \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \cdots \cap T^{-(k-1)n}A) = 0. \end{aligned}$$

This implies that the corresponding averages along any subset $a \cdot \mathbb{N} \subset \mathbb{N}$, $a \neq 1$, also tend subsequentially to zero:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \nu_1(B \cap S^{-an}B \cap \cdots \cap S^{-a(k-1)n}B) = 0,$$

because for large N the terms corresponding to $n \in a \cdot \mathbb{N}$ account for about $1/a$ of the full average.

Now let ν_k be the image measure of ν_1 under the coordinate-dilation

$$\text{dil}_k : Y \rightarrow Y : (\omega_i)_{i \in \mathbb{Z}} \mapsto (\omega_{ki})_{i \in \mathbb{Z}},$$

so that each ν_k is still shift-invariant and satisfies $\nu_k(B) = \nu_1(B)$, since $B = \text{dil}_k^{-1}(B)$ because it depends only on the zeroth coordinate. Letting ν be any limit point of the averages $\frac{1}{K} \sum_{k=1}^K \nu_k$ in the vague topology, the above convergence tells us that

$$\nu(B \cap S^{-a}B \cap \dots \cap S^{-(k-1)a}B) = 0$$

whenever $a \neq 0$.

Step 2 Having made this simplification, Corollary 2.8 gives a new counterexample with density $\delta_1 \geq \delta_0 + c_k(\delta_0)$. Since c_k is bounded away from 0 on the subinterval $[\delta_0, 1] \subset (0, 1]$, after finitely many iterations this procedure gives a counterexample with density greater than 1, and hence a contradiction. \square

Before presenting the proof of Proposition 2.7 we need one further enabling result, for which we will make our appeal to Theorem 2.5. From that theorem we need the consequence that one can approximately decompose an arbitrary positive-measure $U \subset G/\Gamma$ into a collection of almost-invariant sets for different powers of R_g .

Proposition 2.9. *If $R_g \curvearrowright (G/\Gamma, m)$ is an ergodic nilrotation, $U \subset G/\Gamma$ is measurable and of positive measure and $K \geq 1$, then there is a countable set of pairs $\{(V_1, r_1), (V_2, r_2), \dots\}$ (which could be finite or infinite) such that*

- (i) *each V_i has positive measure;*
- (ii) *if $i \neq i'$ then the unions $\bigcup_{k=-K}^K g^{r_i k} V_i$ and $\bigcup_{k=-K}^K g^{r_{i'} k} V_{i'}$ are disjoint;*
- (iii) *$U \supset \bigcup_{i \geq 1} \bigcup_{k=-K}^K g^{r_i k} V_i$;*
- (iv) *and $m(U \setminus \bigcup_{i \geq 1} \bigcup_{k=-K}^K g^{r_i k} V_i) = 0$.*

Proof. The proof given here invokes Zorn's Lemma, although a more careful argument shows that it only really needs the ability to induct transfinitely below ω_1 . Frustratingly, I have not been able to find a proof that avoids this kind of induction entirely, although in the finitary analog of this step all sets are finite and so the issue does not arise.

Let \mathcal{A} be the set of all countable families of pairs $\{(V_1, r_1), (V_2, r_2), \dots\}$ that have properties (i–iii) above (but possibly not (iv)), and order \mathcal{A} by inclusion of families. If $\mathcal{F} = \{(V_1, r_1), (V_2, r_2), \dots\} \in \mathcal{A}$ then set

$$m(\mathcal{F}) := m\left(\bigcup_{i \geq 1} \bigcup_{k=-K}^K g^{r_i k} V_i\right).$$

Since $m(U) > 0$, Theorem 2.5 promises some r such that

$$m(g^{-Kr}U \cap g^{-(K-1)r}U \cap \dots \cap g^{Kr}U) > 0.$$

Therefore the set $V := g^{-Kr}U \cap g^{-(K-1)r}U \cap \dots \cap g^{Kr}U$ has positive measure and satisfies $g^{rk}V \subset U$ for every $-K \leq k \leq K$, so $\{(V, r)\} \in \mathcal{A}$ and hence \mathcal{A} is nonempty.

Now suppose that $(\mathcal{F}_\alpha)_\alpha$ is a totally ordered family in \mathcal{A} . Since $m(V) > 0$ for any $(V, r) \in \mathcal{F}_\alpha$, the values of the measures $m(\mathcal{F}_\alpha)$ are totally ordered, are all

distinct and are bounded by 1. We may therefore extract a non-decreasing sequence $\mathcal{F}_{\alpha_1} \subset \mathcal{F}_{\alpha_2} \subset \dots$ such that $m(\mathcal{F}_{\alpha_i}) \rightarrow \sup_{\alpha} m(\mathcal{F}_{\alpha})$ as $i \rightarrow \infty$.

Since each \mathcal{F}_{α} is countable and they are totally ordered, it follows that $\mathcal{G} := \bigcup_{i \geq 1} \mathcal{F}_{\alpha_i}$ is still countable, and in fact is still a member of \mathcal{A} . Moreover, if $(V, r) \in \mathcal{F}_{\alpha}$ for some α , then this pair must actually appear in some \mathcal{F}_{α_i} , for otherwise we would have $m(\mathcal{F}_{\alpha_i}) \leq m(\mathcal{F}_{\alpha}) - m(\bigcup_{k=-K}^K g^{rk}V)$ for every i , contradicting our construction. Hence $\mathcal{G} \supset \mathcal{F}_{\alpha}$ for all α , and so \mathcal{G} is an upper bound for the chain $(\mathcal{F}_{\alpha})_{\alpha}$.

Therefore by Zorn's Lemma the whole family \mathcal{A} has a maximal element, say $\mathcal{F} = \{(V_1, r_1), (V_2, r_2), \dots\}$. Now we need simply observe that this must have $m(\mathcal{F}) = m(A)$ (which implies property (iv)), since otherwise another appeal to Theorem 2.5 would give $V' \subset U \setminus \bigcup \mathcal{F}$ and $r' \geq 1$ such that $\mathcal{F} \cup \{(V', r')\} \in \mathcal{A}$, contradicting the maximality of \mathcal{F} . Therefore \mathcal{F} has all the desired properties, and the proof is complete. \square

Remark. The finitary analog of this result in [Gow01] (see his Corollary 5.6) is very elementary and quantitative. I suspect that a version of Gowers' proof could be adapted to the present setting (perhaps with some additional assumptions on U , such as that it be open with piecewise-smooth boundary), but that this would require the use of a Mal'cev basis for G and the ability to study orbits of R_g in terms of 'explicit' generalized polynomials using the resulting coordinate system. Such a more quantitative argument would probably be considerably longer than the proof given above. \triangleleft

We can now complete the density-increment proof of multiple recurrence using the above proposition and Corollary 2.6.

Proof of Proposition 2.7. Suppose that $(X \supset A, \mu, T)$ is a process having $\mu(A) =: \delta > 0$ but no k -APs in its return times. Then Corollary 2.6 gives

$$\|\mathbb{E}(A | \pi) - \mu(A)\|_2 \geq \frac{1}{2k} \delta^k$$

for some factor map $\pi : (X, \mu, T) \rightarrow (G/\Gamma, m, R_g)$ onto a $(k-2)$ -step nilrotation.

Since G/Γ is compact and its Borel σ -algebra is generated by its open sets, we can find a finite Borel partition \mathcal{U} of G/Γ into small-diameter positive-measure pieces such that

$$\frac{1}{m(U)} \int_U |\mathbb{E}(A | \pi) - \mu(A | \pi^{-1}U)| dm < \frac{1}{20k} \delta^k$$

for all $U \in \mathcal{U} \setminus \mathcal{U}_{\text{bad}}$, where \mathcal{U}_{bad} is a subcollection such that $m(\bigcup \mathcal{U}_{\text{bad}}) < \delta^k/20k$. Combined with the preceding inequality, this implies that there is some $U \in \mathcal{U}$ for which

$$\mu(A | \pi^{-1}U) > \delta + \frac{1}{10k} \delta^k.$$

Now given $N \geq 1$ choose $K := LN \geq 1$ with $L \geq 1$ so large that $1/L \leq \delta^k/20k$ and $L > 2/\varepsilon + 1$. Apply Proposition 2.9 to the set U to obtain pairs $(V_1, r_1), (V_2, r_2), \dots$ with each V_i having positive measure and such that the unions

$\bigcup_{k=-K}^K g^{r_i k} V_i$ for $i \geq 1$ are pairwise disjoint, all contained in U and together fill up m -almost all of U . In view of the convex combination

$$\mu(A | \pi^{-1}U) = \sum_{i \geq 1} \frac{m\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)}{m(U)} \mu\left(A \mid \pi^{-1}\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)\right),$$

there is some i for which

$$\mu\left(A \mid \pi^{-1}\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)\right) \geq \delta + \frac{1}{10k} \delta^k.$$

Letting $B' = \bigcup_{k=-K}^{K-2N-1} g^{r_i k} V_i$ and

$$C := \bigcup_{k=-K}^K g^{r_i k} V_i \setminus B',$$

the shifts $C, g^{-r_i(2N+1)}C, g^{-2r_i(2N+1)}C, \dots, g^{-(L-1)r_i(2N+1)}C$ are pairwise disjoint and contained in $\bigcup_{k=-K}^K g^{r_i k} V_i$, so each has measure at most $\frac{1}{L}m\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)$ and therefore

$$m(B') \geq \frac{L-1}{L}m\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right).$$

In addition, the set difference $g^{r_i}B' \setminus B'$ is contained in C and so has measure at most

$$\frac{1}{L}m\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right) \leq \frac{1}{L-1}m(B'),$$

and a symmetrical argument controls the measure of $B' \setminus g^{r_i}B'$ so together we obtain

$$m(B' \triangle g^{r_i}B') \leq \frac{2}{L-1}m(B') < \varepsilon m(B').$$

Finally, letting $B := g^{-r_i N}B'$, it follows that $g^{r_i n}B \subset \bigcup_{k=-K}^K g^{r_i k} V_i$ for all $-N \leq n \leq N$, that

$$\begin{aligned} \mu(A | \pi^{-1}g^{r_i n}B) &\geq \frac{\mu(A \cap \pi^{-1}g^{r_i n}B)}{m\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)} \\ &\geq m\left(A \mid \pi^{-1}\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)\right) - \frac{m(C)}{m\left(\bigcup_{k=-K}^K g^{r_i k} V_i\right)} \\ &\geq \delta + \frac{1}{10k} \delta^k - \frac{1}{L} \\ &\geq \delta + \frac{1}{20k} \delta^k, \end{aligned}$$

and that B enjoys the same approximate g^{r_i} -invariance as B' , completing the proof of Proposition 2.7 with $c_k(\delta) := \frac{1}{20k} \delta^k$. \square

3. TWO COMMUTING TRANSFORMATIONS

3.1. The density increment in higher dimensions. With the appearance of Gowers' density-increment proof of Szemerédi's Theorem, it became natural to ask whether a similar approach can yield improved upper bounds for any cases of the multidimensional Szemerédi Theorem. Gowers discusses this question explicitly in [Gow00]. It poses significant new challenges, and remains mostly open. For the analogous ergodic-theoretic study of multiple recurrence we will see that the difficulty arises from the nature of the characteristic factors in multiple dimensions, which are rather more complicated than the pro-nilsystems that give the complete picture for powers of a single ergodic transformation.

In the context of finitary proofs, it is still possible to set up a 'directional' variant of the norms (actually now just seminorms) that Gowers introduced to define uniformity, and to show that the resulting new notion of uniformity does control the count of the desired patterns in a subset $E \subset \{1, 2, \dots, N\}^d$. The difficulty is in handling those sets, or more generally functions $f : \{1, 2, \dots, N\}^d \rightarrow [-1, 1]$, which are *not* uniform in the sense of this seminorm. Extending the approach of Roth and Gowers requires one to find the appropriate class of functions against which an arbitrary function must see a large correlation if it is not uniform. For uniformity of degree k in the one-dimensional setting, these were the functions which on many long arithmetic subprogressions of $\{1, 2, \dots, N\}$ agree with the exponential of i times some degree- k real polynomials (see the discussion following Corollary 2.6), but in the multi-dimensional setting they are much more complicated. Part of the difficulty in extending Gowers' approach lies in the problem of identifying the most appropriate class of functions to use here, and part of it lies in establishing some necessary properties of those functions once they have been found (properties which are fairly classical in the case of the one-dimensional 'local' polynomial functions).

However, in spite of these difficulties, Gowers-like bounds have now been obtained in the following special case of Theorem 1.4 by Shkredov:

Theorem 3.1. *There is some absolute constant $C > 0$ such that if $\delta > 0$, $N \geq 2^{2^{1/\delta^C}}$ and $A \subset \{1, 2, \dots, N\}^2$ has $|A| \geq \delta N^2$, then A contains a corner:*

$$A \supseteq \{\mathbf{a}, \mathbf{a} + r\mathbf{e}_1, \mathbf{a} + r\mathbf{e}_2\}$$

for some $\mathbf{a} \in \{1, 2, \dots, N\}^2$ and $r \geq 1$, where $\mathbf{e}_1, \mathbf{e}_2$ are the standard basis vectors in \mathbb{Z}^2 .

In fact, since the appearance of his original article [Shk06b], in [Shk06a] Shkredov has improved the above bound further to the form $2^{2^{1/\delta^C}}$, effectively by replacing a repeated descent to arithmetic subprogressions with a descent through a nested sequence of Bohr sets, following Bourgain's use of these for his improved bounds in Roth's Theorem [Bou99, Bou08]. In addition, Shkredov has shown in [Shk09] how this latter argument can also be implemented in the setting of arbitrary finite Abelian groups (see also Section 5 of Green's survey [Gre05] for a treatment of the case of high-dimensional vector spaces over a finite field).

However, for the sake of simplicity this note will focus on analogs of the original paper [Shk06b], and where appropriate make comparisons to the steps taken there.

Thus, we here present a new proof of the following special case of Theorem 1.3:

Theorem 3.2. *If $T_1, T_2 : \mathbb{Z} \curvearrowright (X, \mu)$ commute and $A \subset X$ has $\mu(A) > 0$ then*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T_1^{-n} A \cap T_2^{-n} A) > 0.$$

Henceforth we will generally refer to the quadruple (X, μ, T_1, T_2) as a \mathbb{Z}^2 -**system**, in reference to the action of the whole group generated by T_1 and T_2 .

In contrast with our work in the previous section, the analog of Theorem 2.4 that will appear in this setting does not reduce our study to a class of systems for which multiple recurrence can simply be proved directly, as was the case using Theorem 2.5. For this reason, although Theorem 3.2 has of course been known since Furstenberg and Katznelson's work, the proof presented here is not quite so redundant as is the density-increment proof in one dimension (recall the discussion following the statement of Theorem 2.5).

An important aspect of Shkredov's proof is the introduction, in addition to $E \subset \{1, 2, \dots, N\}^2$, of a superset of it which is a product set $F_1 \times F_2$ which must also be manipulated as the proof proceeds. We will employ a similar idea in the following, where for a system (X, μ, T_1, T_2) the structure of a 'product set' is replaced by that of an intersection of sets which are invariant under either T_1 or T_2 . The importance of these special sets corresponds to the emergence of the factor generated by the T_1 - or T_2 -invariant sets within the structure of the characteristic factors. With this in mind, we make the following analog of Definition 2.1.

Definition 3.3 (Augmented process). *An **augmented process** is a \mathbb{Z}^2 -system (X, μ, T_1, T_2) together with distinguished measurable subsets A, E_1 and E_2 satisfying $A \subset E_1 \cap E_2$ and such that E_i is T_i -invariant. We shall sometimes denote these data by $(X \supset E_1 \cap E_2 \supset A, \mu, T_1, T_2)$.*

Definition 3.4. *An augmented process has **no corners in its return set** if*

$$\mu(A \cap T_1^{-n} A \cap T_2^{-n} A) = 0 \quad \forall n \neq 0.$$

In addition, the following notation will be used throughout the sequel.

Definition 3.5 (Partially invariant sets). *If (X, μ, T_1, T_2) is a \mathbb{Z}^2 -system, then a subset $A \subset X$ is **partially invariant** if it is invariant under $T_1^{n_1} T_2^{n_2}$ for some (n_1, n_2) . The σ -algebra of $(T_1^{n_1} T_2^{n_2})$ -invariant measurable sets is denoted by $\Sigma^{(n_1, n_2)}$, and in addition we let $\zeta_0^{(n_1, n_2)}$ be some factor map $X \rightarrow Z_0^{(n_1, n_2)}$ onto an auxiliary system where the transformation in direction (n_1, n_2) is trivial and which generates $\Sigma^{(n_1, n_2)}$.*

(This correspondence between globally invariant σ -subalgebras of Σ and factor maps onto other systems is standard in ergodic theory; see, for instance, Chapter 2 of [Aus10b] and the references given there.)

Definition 3.6 (Kronecker factors). *If (X, μ, T_1, T_2) is an ergodic \mathbb{Z}^2 -system then ζ_1^T will denote some choice of a factor map from X onto an action by rotations on a compact Abelian group which generates the Kronecker factor of (X, μ, T_1, T_2) , and similarly for \mathbb{Z} -systems.*

Definition 3.7 (Arithmetic of factors). *Given two factor maps $\pi_i : (X, \mu, T_1, T_2) \rightarrow (Y_i, \nu_i, S_{1,i}, S_{2,i})$ of a \mathbb{Z}^2 -system, we let $\pi_1 \vee \pi_2$ denote a factor map which generates the same σ -algebra as π_1 and π_2 together (for example, the Cartesian product map $(\pi_1, \pi_2) : X \rightarrow Y_1 \times Y_2$) will do), and $\pi_1 \wedge \pi_2$ denote a factor map which generates the σ -algebra of all sets that are both π_1 - and π_2 -measurable.*

In his setting, Shkredov considered nested inclusions

$$E \subset F_1 \times F_2 \subset \{1, 2, \dots, N\}^2.$$

His main innovation is the result that in order to count approximately the number of corners in E it suffices to control the non-uniformity of E relative to its superset $F_1 \times F_2$, and crucially *to an extent which depends only on the relative density $\frac{|E|}{|F_1||F_2|}$* , provided the sets F_1 and F_2 have some uniformity properties of their own. He effectively formulated this latter uniformity condition in terms of a uniform bound on the one-dimensional Fourier coefficients of the F_i , but for our sets $E_i \in \Sigma^{T_i}$ it turns out that a stronger condition is more convenient, formulated in terms of the independence of their shifts under T_j for $j \neq i$; this condition will appear shortly.

The need for the E_i below becomes natural upon understanding the analog of Theorem 2.4 for the averages of Theorem 3.2. However, in the ergodic theoretic world this involves another new twist, which has no real analog in the finitary setting. It turns out that simply-described characteristic factors for the averages of Theorem 3.2 may be obtained only after ascending to some extension of the initially-given system. (The original system will certainly *have* characteristic factors, but they may be much more complicated to describe.) Of course, it suffices to prove multiple recurrence for such an extension, and so this is quite adequate for our proof strategy. The following result is specialized from the construction of so-called ‘pleasant and isotropized extensions’ in [Aus09, Aus10a].

Theorem 3.8. *Any \mathbb{Z}^2 -system $(X^\circ, \mu^\circ, T_1^\circ, T_2^\circ)$ has an extension*

$$\pi : (X, \mu, T_1, T_2) \rightarrow (X^\circ, \mu^\circ, T_1^\circ, T_2^\circ)$$

with the property that

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \int_X f_0 \cdot (f_1 \circ T_2^n) \cdot (f_2 \circ T_2^n) \, d\mu \\ \sim \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}_\mu(f_0 \mid \pi_0) \cdot \mathbb{E}_\mu(f_1 \mid \pi_1) \cdot \mathbb{E}_\mu(f_2 \mid \pi_2) \, d\mu \end{aligned}$$

as $N \rightarrow \infty$ for any $f_0, f_1, f_2 \in L^\infty(\mu)$, where

$$\begin{aligned}\pi_0 &:= \zeta_0^{(1,0)} \vee \zeta_0^{(0,1)} \\ \pi_1 &:= \zeta_0^{(1,0)} \vee \zeta_0^{(1,-1)} \\ \pi_2 &:= \zeta_0^{(1,-1)} \vee \zeta_0^{(0,1)}.\end{aligned}$$

□

Definition 3.9 (Pleasant system). *Essentially following the nomenclature of [Aus09], we will refer to a system having the property of the extension constructed above as **pleasant**.*

Replacing an initially-given \mathbb{Z}^2 -system with an extension if necessary, we may henceforth concentrate on pleasant systems.

With this description of the characteristic factors in hand, we can now offer our ergodic theoretic translation of Shkredov's main estimate (Theorem 7 in [Shk06b]).

Proposition 3.10. *Suppose that $(X \supset E_1 \cap E_2 \supset A, \mu, T_1, T_2)$ is a pleasant augmented process with $\mu(A) > 0$, that*

- *the return-set of A contains no corners, and*
- *$E_1 \perp T_2^n(E_1)$ and $E_2 \perp T_1^n(E_2)$ for all $n \neq 0$, where \perp denotes independence,*

and let $\pi_0 := \zeta_0^{(1,0)} \vee \zeta_0^{(0,1)}$. Then

$$\|\mathbb{E}_\mu(A | \pi_0) - \mu(A | E_1 \cap E_2)\|_{L^2(\mu(\cdot | E_1 \cap E_2))} \geq \mu(A | E_1 \cap E_2)^3.$$

The benefit of working with the conditions $E_1 \perp T_2^n(E_1)$ is that they will be relatively easy to recover for the new process that we construct during the coming density increment. We will see shortly (Corollary 3.13) that this condition implies that E_1 is orthogonal to the Kronecker factor ζ_1^T , and this orthogonality is a truer ergodic-theoretic analog of Shkredov's condition that they be degree-1 uniformity.

Proposition 3.10 will be proved in Subsection 3.3.

3.2. A closer look at the characteristic factors and the main estimate. Before proving Proposition 3.10 we need some simple auxiliary results about the factors appearing in Theorem 3.8.

Lemma 3.11. *If (X, μ, T_1, T_2) is ergodic as a \mathbb{Z}^2 -system, then any two of the factors $\zeta_0^{(1,0)}$, $\zeta_0^{(0,1)}$, $\zeta_0^{(1,-1)}$ are independent, and the three together are relatively independent over their intersections with the Kronecker factor:*

$$\zeta_1^T \wedge \zeta_0^{(1,0)}, \zeta_1^T \wedge \zeta_0^{(0,1)}, \zeta_1^T \wedge \zeta_0^{(1,-1)}.$$

Proof. The first assertion is an immediate consequence of the commutativity of T_1 and T_2 . We prove it for $\zeta_0^{(1,0)}$ and $\zeta_0^{(0,1)}$, the other pairs being similar: since T_1 and T_2 commute, if A_1 is T_1 -invariant then the conditional expectation $\mathbb{E}(A_1 | \zeta_0^{(0,1)})$ is invariant under both T_1 and T_2 and hence constant, by ergodicity, and must therefore simply equal $\mu(A_1)$.

Handling the three factors together is only a little trickier. If $A_1 \in \zeta_0^{(1,0)}$, $A_2 \in \zeta_0^{(0,1)}$ and $A_{12} \in \zeta_0^{(1,-1)}$, then by the first assertion the target of the factor map $\zeta_0^{(1,0)} \vee \zeta_0^{(0,1)}$ can simply be identified with a Cartesian product system

$$(Y_1 \times Y_2, \nu_1 \otimes \nu_2, S_2 \times \text{id}, \text{id} \times S_1)$$

where S_2 is an ergodic transformation of the first coordinate alone and S_1 an ergodic transformation of the second. The fact that the invariant measure of this target system is a product $\nu_1 \otimes \nu_2$ corresponds to the independence of $\zeta_0^{(1,0)}$ and $\zeta_0^{(0,1)}$. In this picture the set A_i is lifted from some subset $A'_i \subset Y_i$ under the further coordinate projection $Y_1 \times Y_2 \rightarrow Y_i$. Since A_{12} is $\zeta_0^{(1,-1)}$ -measurable one has

$$\mu(A_1 \cap A_2 \cap A_{12}) = \int_X \mathbb{E}(A_1 \cap A_2 | \zeta_0^{(1,-1)}) \cdot 1_{A_{12}} d\mu,$$

and on $Y_1 \times Y_2$ the conditional expectation $\mathbb{E}(A_1 \cap A_2 | \zeta_0^{(1,-1)})$ is identified with the conditional expectation of $A'_1 \times A'_2$ onto the sets invariant under $S_2^{-1} \times S_1$.

It is standard that the invariant sets of a product of ergodic systems depend only on the product of their Kronecker factors (see, for instance, the more general Theorem 7.1 in Furstenberg's original paper [Fur77]), and so our conditional expectation of $A'_1 \times A'_2$ is actually onto the invariant sets of $\zeta_1^{S_2} \times \zeta_1^{S_1}$, whose lifts back up to X must all be measurable with respect to ζ_1^T . Therefore $\mathbb{E}(A_1 \cap A_2 | \zeta_0^{(1,-1)})$ is actually ζ_1^T -measurable, and so the above integral is equal to

$$\int_X \mathbb{E}(A_1 \cap A_2 | \zeta_0^{(1,-1)}) \cdot \mathbb{E}(A_{12} | \zeta_1^T \wedge \zeta_0^{(1,-1)}) d\mu = \int_X 1_{A_1 \cap A_2} \cdot \mathbb{E}(A_{12} | \zeta_1^T \wedge \zeta_0^{(1,-1)}) d\mu.$$

Applying a symmetric argument to the other sets A_i now shows that this equals

$$\int_X \mathbb{E}(A_1 | \zeta_1^T \wedge \zeta_0^{(1,0)}) \cdot \mathbb{E}(A_2 | \zeta_1^T \wedge \zeta_0^{(0,1)}) \cdot \mathbb{E}(A_{12} | \zeta_1^T \wedge \zeta_0^{(1,-1)}) d\mu,$$

which is the desired assertion of relative independence. \square

Remark. The second part of the above lemma, although a very simple consequence of classical results in ergodic theory, has an important counterpart in Lemma 1 (4) of [Shk06b]. It corresponds to the assertion that if sets $F_1, F_2, F_{12} \subseteq \mathbb{Z}/N\mathbb{Z}$ are lifted through the coordinate projections

$$(n_1, n_2) \mapsto n_1, n_2, n_1 + n_2 \quad \text{respectively}$$

and if in addition they are all linearly uniform (meaning that their Fourier coefficients are all small), then their lifts are approximately independent. In his paper Shkredov phrases this in terms of the approximate constancy of a certain convolution of two functions that are lifted from $\mathbb{Z}/N\mathbb{Z}$ in this way. \triangleleft

Lemma 3.12. *Suppose that (Y, ν, S) is an ergodic \mathbb{Z} -system and let $\zeta_1^S : (Y, \nu, S) \rightarrow (Z, m_Z, R)$ be its Kronecker factor. Then for any $f, g \in L^\infty(\nu)$, any $B \subset X$ with $\nu(B) > 0$ that is ζ_1^S -measurable, and any $\varepsilon > 0$, the set*

$$\left\{ n \in \mathbb{Z} : \left| \int_B f \cdot (g \circ S^n) d\nu - \int_B \mathbb{E}_\nu(f | \zeta_1^S) \cdot \mathbb{E}_\nu(g \circ S^n | \zeta_1^S) d\nu \right| \leq \varepsilon \right\}$$

has density 1 in \mathbb{Z} .

Remark. The conclusion of this lemma may be re-phrased as asserting that

$$\mathbb{E}_\nu(f \cdot (g \circ S^n) \mid \zeta_1^S) \sim \mathbb{E}_\nu(f \mid \zeta_1^S) \cdot \mathbb{E}_\nu(g \circ S^n \mid \zeta_1^S)$$

weakly in $L^2(m_Z) \circ \zeta_1^S \subset L^2(\nu)$ as $n \rightarrow \infty$ along some full-density subset of \mathbb{Z} . Strong convergence here for all f and g , rather than weak convergence, would be equivalent to (Y, ν, S) being relatively weakly mixing over its Kronecker factor, which is not always the case. \triangleleft

Proof. On the one hand

$$\int_B f \cdot (g \circ S^n) \, d\nu = \int_Y (f1_B) \cdot (g \circ S^n) \, d\nu$$

and on the other $\mathbb{E}_\nu(f1_B \mid \zeta_1^S) = \mathbb{E}_\nu(f \mid \zeta_1^S)1_B$, because B is already ζ_1^S -measurable, so after replacing f with $f1_B$ if necessary it suffices to treat the case $B = Y$. The desired assertion is now simply that

$$\langle f, g \circ S^n \rangle \sim \langle \mathbb{E}(f \mid \zeta_1^S), \mathbb{E}(g \mid \zeta_1^S) \circ S^n \rangle$$

as $n \rightarrow \infty$ outside some zero-density set of ‘exceptional times’ in \mathbb{Z} , and this is a well-known property of the Kronecker factor (see, for instance, Furstenberg [Fur81]). \square

Corollary 3.13. *If (Y, ν, S) is an ergodic \mathbb{Z} -system and $E \subset Y$ is such that $E \perp S^n(E)$ for all $n \neq 0$ then E is independent from the σ -algebra generated by ζ_1^S under μ .*

Proof. The degenerate case $B = Y$ of the preceding lemma shows that asymptotically for most n we have

$$\nu(E \cap S^{-n}E) \approx \int_Y \mathbb{E}_\nu(E \mid \zeta_1^S) \cdot (\mathbb{E}_\nu(E \mid \zeta_1^S) \circ S^n) \, d\nu.$$

Since the Kronecker factor (Z, m_Z, R) is a compact system, for any $\varepsilon > 0$ there is some nonempty Bohr set in \mathbb{Z} along which the right-hand values above return within ε of

$$\int_Y \mathbb{E}_\mu(E \mid \zeta_1^S)^2 \, d\nu = \|\mathbb{E}_\mu(E \mid \zeta_1^S)\|_2^2.$$

This Bohr set must have positive density and therefore contain a further subset of values of n where our first approximation above is also good. This implies that for any $\varepsilon > 0$ there are infinitely many n for which

$$|\nu(E \cap S^n(E)) - \|\mathbb{E}_\mu(E \mid \zeta_1^S)\|_2^2| < \varepsilon,$$

but on the other hand our assumption on E implies that

$$\nu(E \cap S^n(E)) = \nu(E)^2 = \|\mathbb{E}_\mu(E \mid \zeta_1^S)\|_1^2 \quad \forall n \neq 0.$$

This is possible only if $\|\mathbb{E}_\mu(E \mid \zeta_1^S)\|_1 = \|\mathbb{E}_\mu(E \mid \zeta_1^S)\|_2$, which in turn requires that $\mathbb{E}_\mu(E \mid \zeta_1^S)$ be constant, as required. \square

Lemma 3.14. *If (X, μ, T_1, T_2) is an ergodic \mathbb{Z}^2 -system and $E_1 \in \Sigma^{(1,0)}$, $E_2 \in \Sigma^{(0,1)}$ are such that $E_i \perp T_j^n(E_i)$ for all $n \neq 0$ whenever $\{i, j\} = \{1, 2\}$, then also E_1 (resp. E_2) is independent from $\zeta_0^{(0,1)} \vee \zeta_0^{(1,-1)}$ (resp. $\zeta_0^{(1,0)} \vee \zeta_0^{(1,-1)}$).*

Remark. For us this is analogous to the way Shkredov uses his Lemma 1 to estimate the second term in equation (21) in his Theorem 7. \triangleleft

Proof. The second part of Lemma 3.11 implies

$$\mathbb{E}_\mu(E_1 | \zeta_0^{(0,1)} \vee \zeta_0^{(1,-1)}) = \mathbb{E}_\mu(\mathbb{E}_\mu(E_1 | \zeta_1^T \wedge \zeta_0^{(1,0)}) | \zeta_0^{(0,1)} \vee \zeta_0^{(1,-1)}).$$

Corollary 3.13 now gives that $\mathbb{E}_\mu(E_1 | \zeta_1^T \wedge \zeta_0^{(1,0)})$ is constant, and hence so is the conditional expectation of interest. The proof for E_2 is exactly similar. \square

3.3. The main estimate.

Proof of Proposition 3.10. Define the trilinear form Λ on $L^\infty(\mu)^3$ by

$$\Lambda(f_0, f_1, f_2) := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X f_0 \cdot (f_1 \circ T_1^n) \cdot (f_2 \circ T_2^n) d\mu.$$

(In fact this is the integral of the function $f_0 \otimes f_1 \otimes f_2$ against a certain three-fold self-joining of the system (X, μ, T_1, T_2) called the ‘Furstenberg self-joining’. We will not use that more elaborate formalism here, but refer the reader to [Aus10b] and the references given there for a detailed explanation, as well as a proof that the limit exists.)

Our assumptions include that $\Lambda(A, A, A) = 0$ (where we have simply written A in place of 1_A), but on the other hand by Theorem 3.8 we have

$$\begin{aligned} \Lambda(A, A, A) &= \Lambda(\mathbb{E}(A | \pi_0), A, A) \\ &= \Lambda(\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}, A, A) \\ &\quad + \mu(A | E_1 \cap E_2) \cdot \Lambda(E_1 \cap E_2, A, A). \end{aligned}$$

We now estimate these two terms separately.

First term Directly from the definition of Λ we deduce that

$$\begin{aligned} &|\Lambda(\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}, A, A)| \\ &\leq \Lambda(|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}|, A, A) \\ &\leq \Lambda(|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}|, E_1 \cap E_2, E_1 \cap E_2), \end{aligned}$$

where the second inequality uses that these three functions are non-negative and that $1_A \leq 1_{E_1 \cap E_2}$. Now another appeal to Theorem 3.8 shows that this last upper bound is equal to

$$\Lambda(|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}|, \mathbb{E}(E_1 \cap E_2 | \pi_1), \mathbb{E}(E_1 \cap E_2 | \pi_2)).$$

From our hypothesis that $E_1 \perp T_2^n(E_1)$ for all nonzero n and Lemma 3.14 it follows that E_2 is π_2 -measurable whereas E_1 is independent from π_2 , and hence that

$$\mathbb{E}(E_1 \cap E_2 | \pi_2) = \mu(E_1)1_{E_2},$$

and similarly with the two indices reversed. Given this we can re-write the above term as

$$\begin{aligned}
 & \mu(E_1)\mu(E_2)\Lambda(|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}|, E_1, E_2) \\
 &= \mu(E_1)\mu(E_2) \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int_X |\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}| \\
 & \quad \cdot 1_{T_1^{-n}(E_1)} \cdot 1_{T_2^{-n}(E_2)} d\mu \\
 &= \mu(E_1)\mu(E_2) \int_X |\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}| d\mu,
 \end{aligned}$$

where for the second equality we have now used that E_i is T_i -invariant and that

$$|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}| \cdot 1_{E_1} \cdot 1_{E_2} = |\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}|,$$

which in turn holds because $E_1 \cap E_2$ is π_0 -measurable while $A \subset E_1 \cap E_2$, so that both $\mathbb{E}(A | \pi_0)$ and $1_{E_1 \cap E_2}$ are still supported on $E_1 \cap E_2$.

This integral (which no longer involves the trilinear form Λ) may now be identified as

$$\begin{aligned}
 & \mu(E_1 \cap E_2)^2 \|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}\|_{L^1(\mu(\cdot | E_1 \cap E_2))} \\
 & \leq \mu(E_1 \cap E_2)^2 \|\mathbb{E}(A | \pi_0) - \mu(A | E_1 \cap E_2)1_{E_1 \cap E_2}\|_{L^2(\mu(\cdot | E_1 \cap E_2))},
 \end{aligned}$$

using the fact that $\Sigma^{(1,0)}$ and $\Sigma^{(0,1)}$ are independent to write $\mu(E_1)\mu(E_2) = \mu(E_1 \cap E_2)$ and using Hölder's inequality for the final upper bound.

Second term This is much simpler: since $A \subset E_1 \cap E_2$ and E_i is T_i -invariant we have

$$\begin{aligned}
 \Lambda(E_1 \cap E_2, A, A) &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu((E_1 \cap E_2) \cap T_1^{-n}A \cap T_2^{-n}A) \\
 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(T_1^{-n}E_1 \cap T_1^{-n}A \cap T_2^{-n}E_2 \cap T_2^{-n}A) \\
 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(T_1^{-n}A \cap T_2^{-n}A) \\
 &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap (T_2T_1^{-1})^{-n}A) \\
 &= \|\mathbb{E}_\mu(A | \zeta_0^{(1,-1)})\|_2^2
 \end{aligned}$$

(using the Mean Ergodic Theorem for the last equality), and by another appeal to Hölder's inequality this is bounded below by

$$\|\mathbb{E}_\mu(A | \zeta_0^{(1,-1)})\|_1^2 = \mu(A)^2 = \mu(A | E_1 \cap E_2)^2 \mu(E_1 \cap E_2)^2.$$

Combining the estimates Using the inequalities just obtained in our original decomposition of $\Lambda(A, A, A)$ we find that

$$0 = \Lambda(A, A, A) \geq \mu(A | E_1 \cap E_2)^3 \mu(E_1 \cap E_2)^2 \\ - \|E(A | \pi) - \mu(A | E_1 \cap E_2)\|_{L^2(\mu(\cdot | E_1 \cap E_2))} \cdot \mu(E_1 \cap E_2)^2,$$

so re-arranging gives the desired result. \square

3.4. Shkredov's version of the density increment. We can now present Shkredov's main increment result (which corresponds roughly to the conjunction of Proposition 2.7 and Corollary 2.8 in the one-dimensional setting):

Proposition 3.15. *There is a nondecreasing function $c : (0, 1] \rightarrow (0, 1]$ which is bounded away from 0 on compact subsets of $(0, 1]$ and has the following property. If $(X \supset E_1 \cap E_2 \supset A, \mu, T_1, T_1)$ is such that*

- (i) $\mu(A) > 0$,
- (ii) *the return-set of A contains no nontrivial corners, and*
- (iii) $E_1 \perp T_2^{-n}(E_1)$ *for all $n \neq 0$ and similarly for E_2 ,*

and if we set $\delta := \mu(A | E_1 \cap E_2)$, then there exists another augmented process

$$(X' \supset E'_1 \cap E'_2 \supset A', \mu', T'_1, T'_2)$$

having the analogous properties (i-iii) and such that

$$\mu'(A' | E'_1 \cap E'_2) \geq \delta + c(\delta).$$

Remark. Shkredov's argument does *not* give any effective control over the size of the sets E'_i in terms of the E_i — in particular, it could happen that they are very much smaller — but the point is that this is not needed. \triangleleft

Proof. This breaks naturally into two steps.

Step 1 Extending (X, μ, T_1, T_2) and lifting A and the E_i if necessary, we may assume the system is pleasant. Now by Proposition 3.10 conditions (i) and (ii) imply that

$$\|E(A | \pi_0) - \mu(A | E_1 \cap E_2)\|_{L^2(\mu(\cdot | E_1 \cap E_2))} \geq \delta^3,$$

and hence there is some non-negligible π_0 -measurable set F such that

$$\mu(A | F) > \delta + \delta^3/2.$$

Moreover, since π_0 is generated by $\zeta_0^{(1,0)}$ and $\zeta_0^{(0,1)}$, after approximating this F by a disjoint union of intersections of T_1 - or T_2 -invariant sets we may assume that it is itself of the form $F_1 \cap F_2$ for some $F_1 \in \zeta_0^{(1,0)}$, $F_2 \in \zeta_0^{(0,1)}$.

Naively we should like to replace $A \subset E_1 \cap E_2$ with $A \cap F_1 \cap F_2 \subset F_1 \cap F_2$, but these sets F_i may not satisfy $F_i \perp T^n F_i$ for $n \neq 0$. We resolve this by another conditioning and a vague limit construction. Note at this point that this selection of the sets F_i will be responsible for our lack of control over $\mu'(E'_i)$ in terms of $\mu(E_i)$.

Step 2 Let

$$X' := (\{0, 1\} \times \{0, 1\} \times \{0, 1\})^{\mathbb{Z}^2}$$

with its product Borel space structure, let T'_1, T'_2 be the two coordinate-shifts on this space, and let E'_1, E'_2 and A' be the three obvious time-zero cylinder sets of X' :

$$E'_1 := \{(\omega_{\mathbf{n}}^1, \omega_{\mathbf{n}}^2, \omega_{\mathbf{n}}^3) \in X' : \omega_{\mathbf{0}}^1 = 1\} \quad \text{and similarly.}$$

We will show that for any $\varepsilon > 0$ and $K \geq 1$ there is a probability measure ν on X' such that

- ν is approximately invariant: $|\nu(C) - \nu((T'_i)^{-1}C)| < \varepsilon$ for any $C \subset X'$ and $i = 1, 2$,
- $\nu(E'_1 \cap E'_2) \geq (\delta^3/20)\mu(F_1 \cap F_2)$,
- $\nu(E'_i \Delta T'_i E'_i) = 0$ for $i = 1, 2$,
- $|\nu(E'_i \Delta (T'_j)^{-k} E'_i) - \nu(E'_i)^2| < \varepsilon$ for all nonzero $-K \leq k \leq K$ for $\{i, j\} = \{1, 2\}$, and
- $\nu(A' | E'_i \cap E'_2) \geq \delta + \delta^3/2$.

Given this, we may take a sequence of such measures as $\varepsilon \downarrow 0$ and $K \rightarrow \infty$ and let μ' be a vague limit of some subsequence to obtain an augmented process

$$(X' \supset E'_1 \cap E'_2 \supset A', \mu', T'_1, T'_2)$$

having all the desired properties. The T'_i -invariance of μ' follows from the approximate invariance of the measures ν , and the T'_i -invariance of E'_i is only up to a μ' -negligible set, but this may then be repaired by replacing E'_i with $\bigcup_n (T'_i)^n E'_i$, which differs from E'_i only by a μ' -negligible set. The second of the above points ensures that the limit μ' is non-trivial insofar as $\mu'(A'), \mu'(E'_1 \cap E'_2) > 0$.

Now fix ε and K . To obtain such a ν , let (Z, m, R_1, R_2) be a compact group rotation isomorphic to the Kronecker factor of (X, μ, T_1, T_2) with factor map $\zeta_1^T =: \zeta : X \rightarrow Z$, and let \mathcal{U} be a Borel partition of Z into sufficiently small pieces that

$$\|\mathbb{E}_\mu(F_i | \zeta)|_U - \mu(F_i | U)\|_{L^2(\mu(\cdot|U))} < \varepsilon/4$$

for all $U \in \mathcal{U} \setminus \mathcal{U}_{\text{bad}}$ where $m(\bigcup \mathcal{U}_{\text{bad}}) < (\delta^3/20)\mu(F_1 \cap F_2)$.

Considering the convex combination

$$\mu(A | F_1 \cap F_2) = \sum_{U \in \mathcal{U}} \frac{\mu(F_1 \cap F_2 \cap \zeta^{-1}U)}{\mu(F_1 \cap F_2)} \mu(A \cap \zeta^{-1}U | F_1 \cap F_2 \cap \zeta^{-1}U),$$

the terms indexed by \mathcal{U}_{bad} must contribute very little (because their sum cannot be more than $\delta^3/20$ if we estimate by simply ignoring the factors of $\mu(A \cap \zeta^{-1}U | F_1 \cap F_2 \cap \zeta^{-1}U) \leq 1$). Similarly, the terms for which

$$\mu(F_1 \cap F_2 \cap \zeta^{-1}U | F_1 \cap F_2) < (\delta^3/20)m(U)$$

must also contribute very little (their sum is also less than $\delta^3/20$). Therefore there must be some $U \in \mathcal{U} \setminus \mathcal{U}_{\text{bad}}$ for which

$$\mu(F_1 \cap F_2 \cap \zeta^{-1}U | F_1 \cap F_2) \geq (\delta^3/20)m(U)$$

and

$$\mu(A \cap \zeta^{-1}U | F_1 \cap F_2 \cap \zeta^{-1}U) \geq \delta + \delta^3/4.$$

Using Bayes' formula, the first of these inequalities implies that

$$\begin{aligned}\mu(F_1 \cap F_2 | \zeta^{-1}U) &= \mu(F_1 \cap F_2 \cap \zeta^{-1}U | F_1 \cap F_2) \cdot \frac{\mu(F_1 \cap F_2)}{m(U)} \\ &\geq (\delta^3/20)\mu(F_1 \cap F_2).\end{aligned}$$

Now let $V \subset \mathbb{Z}$ be the Bohr set

$$\{n \in \mathbb{Z} : m(U \triangle R_1^n U) < \varepsilon m(U)/2 \text{ and } m(U \triangle R_2^n U) < \varepsilon m(U)/2\}.$$

This is nontrivial because the rotation orbit $z \mapsto 1_{z+U}$ is continuous from Z to $L^2(m)$, and so V has some (perhaps very small) positive density in \mathbb{Z} . In view of this positive density, Lemma 3.12 implies that each of the sets

$$\begin{aligned}V_{j,k} := \\ \left\{ n \in V : \left| \mu(F_i \cap T_j^{-kn} F_i | \zeta^{-1}U) \right. \right. \\ \left. \left. - \frac{1}{m(U)} \int_U \mathbf{E}_\mu(F_i | \zeta) \cdot \mathbf{E}_\mu(T_j^{-kn} F_i | \zeta) \, dm \right| \leq \varepsilon/2 \right\}\end{aligned}$$

still has relative density 1 inside V for any $k \neq 0$ and $j = 1$ or 2 , because the whole set $\mathbb{Z} \setminus V_{j,k}$ has density zero. Hence we may choose some $r \in V$, $r \geq 1$ that lies in every $V_{j,k}$ for $j = 1, 2$ and $k \in \{-K, -K+1, \dots, K\} \setminus \{0\}$. On the other hand, the approximation that defines the members of $\mathcal{U} \setminus \mathcal{U}_{\text{bad}}$ and the approximate return of U to itself under R_j^n for $n \in V$ imply that

$$\begin{aligned}\frac{1}{m(U)} \int_U \mathbf{E}_\mu(F_i | \zeta) \cdot \mathbf{E}_\mu(T_j^{-kr} F_i | \zeta) \, dm \\ \approx \frac{1}{m(U)} \int_U \mu(F_i | \zeta^{-1}U) \cdot \mathbf{E}_\mu(T_j^{-kr} F_i | \zeta) \, dm \\ = \mu(F_i | \zeta^{-1}U) \cdot \mu(F_i | T_j^{kr} \zeta^{-1}U) \approx \mu(F_i | \zeta^{-1}U)^2\end{aligned}$$

for all nonzero $-K \leq k \leq K$, where the error incurred is at most $\varepsilon/4 + \varepsilon/4 = \varepsilon/2$.

Now consider the map

$$\varphi : X \rightarrow X' : x \mapsto (1_{F_1}(T_2^{rn_2}x), 1_{F_2}(T_1^{rn_1}x), 1_{A \cap F_1 \cap F_2}(T_1^{rn_1}T_2^{rn_2}x))_{(n_1, n_2) \in \mathbb{Z}^2}$$

and let ν be the image measure $\varphi_{\#}\mu(\cdot | \zeta^{-1}U)$ on X' . We will show that this has the five desired properties:

- approximate invariance of ν follows from approximate invariance of U along V :

$$\begin{aligned}|\nu(C) - \nu((T_i^r)^{-1}C)| &= |\mu(\varphi^{-1}C | \zeta^{-1}U) - \mu(T_i^{-r}\varphi^{-1}C | \zeta^{-1}U)| \\ &= \frac{\mu(\varphi^{-1}C \cap (\zeta^{-1}U \triangle T_i^r \zeta^{-1}U))}{\mu(\zeta^{-1}U)} \\ &\leq \varepsilon/2 < \varepsilon\end{aligned}$$

for any $C \subset X'$;

- a simple calculation gives

$$\nu(E'_1 \cap E'_2) = \mu(F_1 \cap F_2 | \zeta^{-1}U),$$

and this is at least $(\delta^3/20)\mu(F_1 \cap F_2)$ by our choice of U ;

- similarly,

$$\nu(E'_i \Delta T'_i E'_i) = \mu(F_i \Delta T_i F_i | \zeta^{-1}U) = 0$$

for $i = 1, 2$;

- for any nonzero $-K \leq k \leq K$ we have

$$\nu(E'_i \cap (T'_j)^{-k} E'_i) = \mu(F_i \cap T_j^{-kr} F_i | \zeta^{-1}U),$$

and by our selection of r this is within $\varepsilon/2$ of

$$\frac{1}{m(U)} \int_U \mathbb{E}_\mu(F_i | \zeta) \cdot \mathbb{E}_\mu(T_j^{-kn} F_i | \zeta) dm,$$

which in turn is within $\varepsilon/2$ of

$$\mu(F_i | \zeta^{-1}U)^2 = \nu(E'_i)^2,$$

giving the required estimate;

- lastly, our choice of U also guarantees that

$$\nu(A' | E'_1 \cap E'_2) = \mu(A \cap \zeta^{-1}U | F_1 \cap F_2 \cap \zeta^{-1}U) \geq \delta + \delta^3/4,$$

as required.

This completes the proof with $c(\delta) := \delta^3/4$. □

Remark. The two steps above can also be loosely identified with two steps in Shkredov's work. The first is similar to the conjunction of Lemma 11 and Proposition 3 in Section 3 of [Shk06b], whose use appears at the beginning of the proof of Theorem 4. The second, rather more involved, amounts to Corollary 1 and the various auxiliary results needed to reach it in Section 4 of [Shk06b], which then underpin the second step of each increment in the proof of Shkredov's Theorem 4. ◁

Proof of Theorem 3.2. This now proceeds almost exactly as for Theorem 1.2.

Suppose there exists an augmented process $(X \supset E_1 \cap E_2 \supset A, \mu, T_1, T_2)$ such that $\mu(A) > 0$ and hence $\mu(A | E_1 \cap E_2) =: \delta_0 > 0$, $E_i \perp T_j^n(E_i)$ for all $n \geq 0$, and for which

$$\frac{1}{N} \sum_{n=1}^N \mu(A \cap T_1^{-n} A \cap T_2^{-n} A) \rightarrow 0.$$

In particular, if $(X \supset A, \mu, T_1, T_2)$ is a process violating Theorem 3.2, then $(X \supset X \cap X \supset A, \mu, T_1, T_2)$ is an augmented process with these properties.

From these data one can construct another augmented process $(Y \supset G_1 \cap G_2 \supset B, \nu, S_1, S_2)$ such that $\nu(B) = \mu(A)$, $\nu(G_i) = \mu(E_i)$ and this new process actually has no corners in its return set. This construction proceeds in exact analogy

with Step 1 in the proof Theorem 1.3 from Corollary 2.8: the initial process is transferred to the symbolic space

$$Y := (\{0, 1\} \times \{0, 1\} \times \{0, 1\})^{\mathbb{Z}^2},$$

where now the three copies of $\{0, 1\}$ above the coordinate (n_1, n_2) receive the indicator functions of $1_{E_1} \circ T_1^{n_1} T_2^{n_2}$, $1_{E_2} \circ T_1^{n_1} T_2^{n_2}$ and $1_A \circ T_1^{n_1} T_2^{n_2}$ respectively; and then averaging over dilations constructs a new shift-invariant measure on this symbolic space that retains the properties of the original system but actually has no corners in its return set. A quick check shows that if G_1, G_2 and B denote the one-dimensional cylinder sets defined by the three different $\{0, 1\}$ -valued coordinates above $(0, 0)$ in Y , then the G_i retain the property of ν -a.s. invariance under S_i and also the property that $G_i \perp S_j^n(G_i)$ for all $n \neq 0$ (because the measure $\nu(G_i \cap S_j^n(G_i))$ is obtained as an average over m of $\mu(E_i \cap T_j^{nm}(E_i))$, and these are all equal to $\mu(E_i)^2 = \nu(G_i)^2$ by assumption).

Now implementing Proposition 3.15, one can construct from $(Y \supset G_1 \cap G_2 \supset B, \nu, S_1, S_2)$ a new augmented process $(X' \supset E'_1 \cap E'_2 \supset A', \mu', T'_1, T'_2)$ which still has all the properties (i–iii) and for which $\mu(A' | E'_1 \cap E'_2) \geq \delta_0 + c(\delta_0)$. Since c is uniformly positive on $[\delta_0, 1]$, after iterating this construction finitely many times we obtain an example of an augmented process for which this relative density is greater than 1, a contradiction. \square

Remark. The above treatment bears comparison with how Shkredov assembles the various components of the proof of his main result, Theorem 4, in [Shk06b]. \triangleleft

4. FURTHER DISCUSSION

Theorem 3.2 remains the most elaborate higher-dimensional case of Theorem 1.4 to be successfully proved using a density-increment argument, or to be given bounds that improve over the hypergraph-regularity proofs of the general theorem obtained in [Gow07] and [NRS06]. Perhaps the most obvious obstruction to further progress is that the various ‘inverse theorems’ that are known for the relevant notions of uniformity remain incomplete. However, in the ergodic-theoretic world these correspond to ‘characteristic factor’ theorems such as Theorem 3.8, and recent work has in fact taken these a little further. The following result appears (in a slightly more general form) as Theorem 1.1 in [Ausb], where it is used for a different purpose.

Theorem 4.1. *Any ergodic \mathbb{Z}^2 -system $(X^\circ, \mu^\circ, T_1^\circ, T_2^\circ)$ admits an ergodic extension*

$$\pi : (X, \mu, T_1, T_2) \rightarrow (X^\circ, \mu^\circ, T_1^\circ, T_2^\circ)$$

with the property that

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int_X f_0 \cdot (f_1 \circ T_1^n) \cdot (f_2 \circ T_2^n) \cdot (f_3 \circ T_1^n T_2^n) \, d\mu \\ & \sim \frac{1}{N} \sum_{n=1}^N \int_X \mathbf{E}(f_0 | \pi_0) \cdot (\mathbf{E}(f_1 | \pi_1) \circ T_1^n) \cdot (\mathbf{E}(f_2 | \pi_2) \circ T_2^n) \cdot (\mathbf{E}(f_3 | \pi_3) \circ T_1^n T_2^n) \, d\mu \end{aligned}$$

in $L^2(\mu)$ as $N \rightarrow \infty$ for any $f_0, f_1, f_2, f_3 \in L^\infty(\mu)$, where

$$\begin{aligned}\pi_0 = \pi_3 &:= \zeta_0^{(1,0)} \vee \zeta_0^{(0,1)} \vee \zeta_0^{(1,1)} \vee \zeta_{2,\text{nil}}^T \\ \pi_1 = \pi_2 &:= \zeta_0^{(1,0)} \vee \zeta_0^{(1,-1)} \vee \zeta_0^{(0,1)} \vee \zeta_{2,\text{nil}}^T,\end{aligned}$$

and where $\zeta_{2,\text{nil}}^T$ denotes a factor generated by an inverse limit of a sequence of actions of \mathbb{Z}^2 by two-step nilrotations.

Once again, these π_i are referred to as the ‘characteristic’ factors for these multiple averages.

Moreover, a relatively simple extension of Lemma 3.11 shows that the four factors $\zeta_0^{(1,0)}$, $\zeta_0^{(0,1)}$, $\zeta_0^{(1,1)}$ and $\zeta_0^{(1,-1)}$ that appear above are relatively independent over their further intersections with $\zeta_{2,\text{nil}}^T$ (see Proposition 5.3 in [Ausb]). Theorem 4.1 and this second result are both known special cases of a general conjecture on the joint distributions of partially invariant factors of \mathbb{Z}^d -systems, which may be found formulated carefully in Section 6 of [Aus10b] and which suggests that an inverse theory for all higher-dimensional notions of uniformity generalizing the Gowers norms will ultimately be available.

Theorem 4.1 itself bears on the special case of multiple recurrence asserting that

$$\mu(A) > 0 \quad \Rightarrow \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T_1^{-n}A \cap T_2^{-n}A \cap T_1^{-n}T_2^{-n}A) > 0,$$

which in the finitary world corresponds to finding squares in dense subsets of \mathbb{Z}^2 . The above structural results offer hope that some analog of Shkredov’s density-increment approach may be possible through the study of pleasant augmented processes of the form

$$(X \supset E_1 \cap E_2 \cap E_3 \cap E_4 \supset A, \mu, T_1, T_2)$$

where E_1, E_2, E_3 and E_4 are measurable with respect to $\zeta_0^{(1,0)}$, $\zeta_0^{(0,1)}$, $\zeta_0^{(1,1)}$ and $\zeta_0^{(1,-1)}$ respectively. Of course, more ideas would still be needed to give a new density-increment proof of this instance of multiple recurrence, even in the infinitary setting of ergodic theory. For example, Proposition 3.10 must be replaced with some more complicated estimate, and then arguments in the previous section which used some conditioning on the Kronecker factor would presumably be replaced by conditioning on $\zeta_{2,\text{nil}}^T$, which can have much more complicated behaviour.

Another interesting issue on which ergodic theory can shed some light concerns the difference between the problems of proving multiple recurrence for the above averages and for the averages

$$\frac{1}{N} \sum_{n=1}^N \mu(A \cap T_1^{-n}A \cap T_2^{-n}A \cap T_3^{-n}A)$$

arising from a \mathbb{Z}^3 -system (X, μ, T_1, T_2, T_3) . We offer only a very informal discussion of this here, since precise results on these more complex problems are still in their infancy. In the finitary world, these latter averages correspond to finding

three-dimensional corners in dense subsets of \mathbb{Z}^3 , rather than squares in \mathbb{Z}^2 . Since a triple of the form (T_1, T_2, T_1T_2) does formally generate an action of \mathbb{Z}^3 , it is clear that multiple recurrence for these \mathbb{Z}^3 -system averages is at least as strong as its counterpart for the averages of Theorem 4.1. However, the identification of characteristic factors for the case of \mathbb{Z}^3 -systems is also apparently simpler: the main result of [Aus10a] shows that, after passing to a suitable extension if necessary, one has

$$\begin{aligned} & \frac{1}{N} \sum_{n=1}^N \int_X f_0 \cdot (f_1 \circ T_1^n) \cdot (f_2 \circ T_2^n) \cdot (f_3 \circ T_3^n) d\mu \\ & \sim \frac{1}{N} \sum_{n=1}^N \int_X \mathbb{E}(f_0 | \pi'_0) \cdot (\mathbb{E}(f_1 | \pi'_1) \circ T_1^n) \cdot (\mathbb{E}(f_2 | \pi'_2) \circ T_2^n) \cdot (\mathbb{E}(f_3 | \pi'_3) \circ T_3^n) d\mu \end{aligned}$$

with

$$\begin{aligned} \pi'_0 & := \zeta_0^{(1,0,0)} \vee \zeta_0^{(0,1,0)} \vee \zeta_0^{(0,0,1)}, & \pi'_1 & := \zeta_0^{(1,0,0)} \vee \zeta_0^{(1,-1,0)} \vee \zeta_0^{(1,0,-1)}, \\ \pi'_2 & := \zeta_0^{(1,-1,0)} \vee \zeta_0^{(0,1,0)} \vee \zeta_0^{(0,1,-1)} & \text{and} & \pi'_3 := \zeta_0^{(1,0,-1)} \vee \zeta_0^{(0,1,-1)} \vee \zeta_0^{(0,0,1)}, \end{aligned}$$

and these are the minimal factors with this property. These factors are ‘simpler’ in that they involve only partially-invariant factors, and not compact group rotations or nilsystems. The fact that some of the ingredients needed in Theorem 4.1 no longer appear here does not contradict the fact that a triple such as (T_1, T_2, T_1T_2) generates a \mathbb{Z}^3 -system, because after passing to a suitable extension the algebraic relations among the generators of this \mathbb{Z}^3 -system will usually be lost.

It thus appears that the analysis of the more general averages might actually be easier, and in fact for deploying some of the methods at our disposal this is true. The ergodic theoretic proof of convergence of these averages in [Aus09] (reproving a result of Tao from [Tao08]) implicitly needs the linear independence of the group elements corresponding to T_1 , T_2 and T_3 . In the finitary world, the hypergraph-regularity proofs of the multidimensional Szemerédi Theorem must first lift the problem into a group \mathbb{Z}^d for d large enough so that one is looking for the corners of a d -dimensional simplex (rather than any more complicated d -dimensional constellations) before this search can be correctly recast in the language of extremal hypergraph theory.

However, both of these arguments use only the most basic, ‘rough’ structure for the data being studied, and by contrast the more refined density-increment approach *is* simpler in the case of two-dimensional squares than that of three-dimensional corners. Each of the superficially-simpler characteristic factors π'_i for the three-dimensional problem is assembled from ingredients of the form $\zeta_0^{\mathbf{v}}$ for some $\mathbf{v} \in \mathbb{Z}^3$, and each of these is a factor map onto a factor of (X, μ, T_1, T_2, T_3) on which the acting group is essentially $\mathbb{Z}^3/\mathbb{Z}\mathbf{v} \cong \mathbb{Z}^2$ (owing to the partial invariance). In order to mimic Shkredov’s approach to these results, it is then necessary to know how all of these essentially two-dimensional systems are jointly distributed as factors of (X, μ, T_1, T_2, T_3) (in order to generalize our use of Lemma 3.11 in the proof of Proposition 3.10, for example). It turns out that to understand this joint

distribution one needs the same kind of machinery as for the identification of a tuple of characteristic factors in the first place (the reason why these are essentially equivalent problems is discussed in detail in Chapter 4 of [Aus10b]); and the particular problem of describing the joint distribution of these ‘two-dimensional’ factors turns out to be of a similar level of complexity to the problem of describing characteristic factors for multiple recurrence across squares in a \mathbb{Z}^2 -action.

So the finer information required for the density-increment strategy forces one to understand not only the ‘top-level’ structural result that is contained in the identification of a characteristic tuple of factors, but also how all the ingredients appearing in those characteristic factors are jointly distributed. This can be of similar difficulty to a lower-dimensional problem of identifying characteristic factors. For understanding multiple recurrence across translates and dilates of some complicated constellation in \mathbb{Z}^d , one might need to work with a large partially ordered family of factors of a given system, where the characteristic factors appear as the maximal elements, and several layers of smaller factors (including group rotations, nilsystems, or possibly something else) must also be identified in order to describe all the necessary joint distributions well enough to implement a density increment. For a density-increment proof such as in Section 3 above, this would presumably entail working with a much richer analog of the augmented processes that appear there.

These speculations notwithstanding, serious problems surround the status of finitary analogs of Theorem 4.1 or its generalizations. I believe such analogs are expected by many researchers in this field, but formulating a precise conjecture is already tricky, and at this writing I know of no higher-dimensional results beyond Shkredov’s. It is not clear what methods (extending Shkredov’s or others) are needed to establish such structural results. Without them, the prospect of a density-increment proof of the presence of prescribed constellations in dense subsets of \mathbb{Z}^d seems rather remote.

REFERENCES

- [Ausa] Tim Austin. Pleasant extensions retaining algebraic structure, I. Preprint, available online at [arXiv.org](https://arxiv.org/abs/0905.0518): 0905.0518.
- [Ausb] Tim Austin. Pleasant extensions retaining algebraic structure, II. Preprint, available online at [arXiv.org](https://arxiv.org/abs/0910.0907): 0910.0907.
- [Aus09] Tim Austin. On the norm convergence of nonconventional ergodic averages. *Ergodic Theory Dynam. Systems*, 30(2):321–338, 2009.
- [Aus10a] Tim Austin. Deducing the multidimensional Szemerédi theorem from an infinitary removal lemma. *J. Anal. Math.*, 111:131–150, 2010.
- [Aus10b] Tim Austin. *Multiple recurrence and the structure of probability-preserving systems*. ProQuest LLC, Ann Arbor, MI, 2010. Thesis (Ph.D.)—University of California, Los Angeles.
- [Ber96] Vitaly Bergelson. Ergodic Ramsey Theory – an Update. In M. Pollicott and K. Schmidt, editors, *Ergodic Theory of \mathbb{Z}^d -actions: Proceedings of the Warwick Symposium 1993-4*, pages 1–61. Cambridge University Press, Cambridge, 1996.
- [BHK05] Vitaly Bergelson, Bernard Host, and Bryna Kra. Multiple recurrence and nilsequences. *Invent. Math.*, 160(2):261–303, 2005. With an appendix by Imre Ruzsa.
- [BLL08] V. Bergelson, A. Leibman, and E. Lesigne. Intersective polynomials and the polynomial Szemerédi theorem. *Adv. Math.*, 219(1):369–388, 2008.

- [Bou99] J. Bourgain. On triples in arithmetic progression. *Geom. Funct. Anal.*, 9(5):968–984, 1999.
- [Bou08] Jean Bourgain. Roth’s theorem on progressions revisited. *J. Anal. Math.*, 104:155–192, 2008.
- [CL84] Jean-Pierre Conze and Emmanuel Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, 112(2):143–175, 1984.
- [CL88a] Jean-Pierre Conze and Emmanuel Lesigne. Sur un théorème ergodique pour des mesures diagonales. In *Probabilités*, volume 1987 of *Publ. Inst. Rech. Math. Rennes*, pages 1–31. Univ. Rennes I, Rennes, 1988.
- [CL88b] Jean-Pierre Conze and Emmanuel Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris Sér. I Math.*, 306(12):491–493, 1988.
- [FK78] Hillel Furstenberg and Yitzhak Katznelson. An ergodic Szemerédi Theorem for commuting transformations. *J. d’Analyse Math.*, 34:275–291, 1978.
- [FK91] Hillel Furstenberg and Yitzhak Katznelson. A Density Version of the Hales-Jewett Theorem. *J. d’Analyse Math.*, 57:64–119, 1991.
- [Fur77] Hillel Furstenberg. Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d’Analyse Math.*, 31:204–256, 1977.
- [Fur81] Hillel Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, Princeton, 1981.
- [FW03] Hillel Furstenberg and Benjamin Weiss. Markov processes and Ramsey theory for trees. *Combin. Probab. Comput.*, 12(5-6):547–563, 2003. Special issue on Ramsey theory.
- [Gow98] W. T. Gowers. A new proof of Szemerédi’s theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.
- [Gow00] W. T. Gowers. Rough structure and classification. *Geom. Funct. Anal.*, (Special Volume, Part I):79–117, 2000. GAFA 2000 (Tel Aviv, 1999).
- [Gow01] W. T. Gowers. A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.
- [Gow07] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007.
- [Gre05] Ben Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005.
- [GT12] Ben Green and Terence Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.
- [HK05] Bernard Host and Bryna Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. Math.*, 161(1):397–488, 2005.
- [Kra07] Bryna Kra. Ergodic methods in additive combinatorics. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 103–143. Amer. Math. Soc., Providence, RI, 2007.
- [Lei98] A. Leibman. Polynomial sequences in groups. *J. Algebra*, 201(1):189–206, 1998.
- [Lei05] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):201–213, 2005.
- [Les91] Emmanuel Lesigne. Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques. *Ergodic Theory Dynam. Systems*, 11(2):379–391, 1991.
- [NRS06] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht. The counting lemma for regular k -uniform hypergraphs. *Random Structures Algorithms*, 28(2):113–179, 2006.
- [Par69] William Parry. Ergodic properties of affine transformations and flows on nilmanifolds. *Amer. J. Math.*, 91:757–771, 1969.
- [Par70] William Parry. Dynamical systems on nilmanifolds. *Bull. London Math. Soc.*, 2:37–40, 1970.
- [Par73] William Parry. Dynamical representations in nilmanifolds. *Compositio Math.*, 26:159–174, 1973.
- [Rot53] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

- [Shk06a] I. D. Shkredov. On a generalization of Szemerédi's theorem. *Proc. London Math. Soc.* (3), 93(3):723–760, 2006.
- [Shk06b] I. D. Shkredov. On a problem of Gowers. *Izv. Ross. Akad. Nauk Ser. Mat.*, 70(2):179–221, 2006.
- [Shk09] I. D. Shkredov. On a two-dimensional analogue of Szemerédi's theorem in abelian groups. *Izv. Ross. Akad. Nauk Ser. Mat.*, 73(5):181–224, 2009.
- [Tao06a] Terence Tao. A quantitative ergodic theory proof of Szemerédi's theorem. *Electron. J. Combin.*, 13(1):Research Paper 99, 49 pp. (electronic), 2006.
- [Tao06b] Terence Tao. A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A*, 113(7):1257–1280, 2006.
- [Tao08] Terence Tao. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory and Dynamical Systems*, 28:657–688, 2008.
- [TV06] Terence Tao and Van Vu. *Additive combinatorics*. Cambridge University Press, Cambridge, 2006.
- [Zie05] T. Ziegler. A non-conventional ergodic theorem for a nilsystem. *Ergodic Theory Dynam. Systems*, 25(4):1357–1370, 2005.
- [Zie07] Tamar Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20(1):53–97 (electronic), 2007.
- [Zim76a] Robert J. Zimmer. Ergodic actions with generalized discrete spectrum. *Illinois J. Math.*, 20(4):555–588, 1976.
- [Zim76b] Robert J. Zimmer. Extensions of ergodic group actions. *Illinois J. Math.*, 20(3):373–409, 1976.

COURANT INSTITUTE, NEW YORK UNIVERSITY, NEW YORK, NY 10012, USA

E-mail address: tim@cims.nyu.edu

URL: <http://www.cims.nyu.edu/~tim>