Centre de Recherches Mathématiques
CRM Proceedings and Lecture Notes
Volume **??**, 2007

Ergodic Methods in Additive Combinatorics

Bryna Kra

Abstract. Shortly after Szemerédi's proof that a set of positive upper density contains arbitrarily long arithmetic progressions, Furstenberg gave a new proof of this theorem using ergodic theory. This gave rise to the field of combinatorial ergodic theory, in which problems motivated by additive combinatorics are addressed kwith ergodic theory. Combinatorial ergodic theory has since produced combinatorial results, some of which have yet to be obtained by other means, and has also given a deeper understanding of the structure of measure preserving systems. We outline the ergodic theory background needed to understand these results, with an emphasis on recent developments in ergodic theory and the relation to recent developments in additive combinatorics.

These notes are based on four lectures given during the School on Additive Combinatorics at the Centre de recherches mathématiques, Montreal in April, 2006. The talks were aimed at an audience without background in ergodic theory. No attempt is made to include complete proofs of all statements and often the reader is referred to the original sources. Many of the proofs included are classic, included as an indication of which ingredients play a role in the developments of the past ten years.

1. Combinatorics to ergodic theory

1.1. Szemerédi's theorem. Answering a long standing conjecture of Erdős and Turán [11], Szemerédi [54] showed that a set $E \subset \mathbb{Z}$ with positive upper density[1] contains arbitrarily long arithmetic progressions. Soon thereafter, Furstenberg [16] gave a new proof of Szemerédi's Theorem using ergodic theory, and this has lead to the rich field of combinatorial ergodic theory. Before describing some of the results in this subject, we motivate the use of ergodic theory for studying combinatorial problems.

We start with the finite formulation of Szemerédi's theorem:

Theorem 1.1 (Szemerédi [54]). *Given $\delta > 0$ and $k \in \mathbb{N}$, there is a function $N(\delta, k)$ such that if $N > N(\delta, k)$ and $E \subset \{1, \dots, N\}$ is a subset with $|E| \geq \delta N$, then E contains an arithmetic progression of length k.*

1991 *Mathematics Subject Classification.* Primary 37A30; Secondary 11B25, 27A45.

Key words and phrases. Ergodic theory, additive combinatorics.

The author was supported in part by NSF Grant DMS-#0555250.

This is the final form of the paper.

[1]Given a set $E \subset \mathbb{Z}$, its *upper density* $d^*(E)$ is defined by $d^*(E) = \limsup_{N\to\infty} |E \cap \{1, \dots, N\}|/N$.

It is clear that this statement immediately implies the first formulation of Szemerédi's theorem, and a compactness argument gives the converse implication.

1.2. Translation to a probability system. Starting with Szemerédi's theorem, one gains insight into the intersection of sufficiently many sets with positive measure in an arbitrary probability system.[2] Note that $N(\delta, k)$ denotes the quantity in Theorem 1.1.

Corollary 1.2. *Let $\delta > 0$, $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu)$ be a probability space and $A_1, \dots, A_N \in \mathcal{X}$ with $\mu(A_i) \geq \delta$ for $i = 1, \dots, N$. If $N > N(\delta, k)$, then there exist $a, d \in \mathbb{N}$ such that*

$$A_a \cap A_{a+d} \cap A_{a+2d} \cap \cdots \cap A_{a+kd} \neq \varnothing.$$

PROOF. For $A \in \mathcal{X}$, let $\mathbf{1}_A(x)$ denote the characteristic function of A (meaning that $\mathbf{1}_A(x)$ is 1 for $x \in A$ and is 0 otherwise). Let $N > N(\delta, k)$. Then

$$\int_X \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{1}_{A_n} \, \mathrm{d}\mu \geq \delta.$$

Thus there exists $x \in X$ such that

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathbf{1}_{A_n}(x) \geq \delta.$$

Then $E = \{n : x \in A_n\}$ satisfies $|E| \geq \delta N$, and so Szemerédi's theorem implies that E contains an arithmetic progression of length k. By the definition of E, we have a sequence of sets with the desired property. □

1.3. Measure preserving systems. A *probability measure preserving system* is a quadruple $(X, \mathcal{X}, \mu, T)$, where $(X, \mathcal{X}, \mu)$ is a probability space and $T \colon X \to X$ is a bijective, measurable, measure preserving transformation. This means that for all $A \in \mathcal{X}$, $T^{-1}A \in \mathcal{X}$ and

$$\mu(T^{-1}A) = \mu(A).$$

In general, we refer to a probability measure preserving system as a *system.*

Without loss of generality, we can place several simplifying assumptions on our systems. We assume that $\mathcal{X}$ is countably generated; thus for $1 \leq p < \infty$, $L^p(\mu)$ is separable. We implicitly assume that all sets and functions are measurable with respect to the appropriate σ-algebra, even when this is not explicitly stated. Equality between sets or functions is always meant up to sets of measure 0.

[2] A *σ-algebra* is a collection $\mathcal{X}$ of subsets of X satisfying: (i) $X \in \mathcal{X}$ (ii) for any $A \in \mathcal{X}$ we also have $X \setminus A \in \mathcal{X}$ (iii) for any countable collection $A_n \in \mathcal{X}$, we also have $\bigcup_{n=1}^{\infty} A_n \in \mathcal{X}$. A σ-algebra is endowed with operations $\bigvee$, $\bigwedge$, and c, which correspond to union, intersection, and taking complements. By a *probability system*, we mean a triple $(X, \mathcal{X}, \mu)$ where X is a measure space, $\mathcal{X}$ is a σ-algebra of measurable subsets of X, and μ is a probability measure. In general, we use the convention of denoting the σ-algebra $\mathcal{X}$ by the associated calligraphic version of the measure space X.

1.4. Furstenberg multiple recurrence. In a system, one can use Szemerédi's theorem to derive a bit more information about intersections of sets. If $(X,\mathcal{X},\mu,T)$ is a system and $A\in\mathcal{X}$ with $\mu(A)\geq\delta>0$, then

$$A, T^{-1}A, T^{-2}A, \ldots, T^{-n}A, \ldots$$

are all sets of the same measure, and so all have measure $\geq\delta$. Applying Corollary 1.2 to this sequence of sets, we have the existence of $a,d\in\mathbb{N}$ with

$$T^{-a}A\cap T^{-(a+d)}A\cap T^{-(a+2d)}\cap\cdots\cap T^{-(a+kd)}A\neq\varnothing.$$

Furthermore, the measure of this intersection must be positive. If not, we could remove from A a subset of measure zero containing all the intersections and obtain a subset of measure at least δ without this property. In this way, starting with Szemerédi's Theorem, we have derived Furstenberg's multiple recurrence theorem:

Theorem 1.3 (Furstenberg [16]). *Let $(X,\mathcal{X},\mu,T)$ be a system and let $A\in\mathcal{X}$ with $\mu(A)>0$. Then for any $k\geq 1$, there exists $n\in\mathbb{N}$ such that*

$$\mu\big(A\cap T^{-n}A\cap T^{-2n}A\cap\cdots\cap T^{-kn}A\big)>0. \tag{1.1}$$

2. Ergodic theory to combinatorics

2.1. Strong form of multiple recurrence. We have seen that Furstenberg multiple recurrence can be easily derived from Szemerédi's theorem. More interesting is the converse implication, showing that one can use ergodic theory to prove regularity properties of subsets of the integers, and in particular derive Szemerédi's theorem. This is what Furstenberg did in his landmark paper [16], and the techniques introduced in this paper have been used subsequently to deduce other patterns in subsets of integers with positive upper density. (See Section 9.) Moreover, Furstenberg's proof lead to new questions within ergodic theory, about the structure of measure preserving systems. In turn, this finer analysis of measure preserving systems has had implications in additive combinatorics. We return to these questions in Section 3.

Furstenberg's approach to Szemerédi's theorem has two major components. The first is proving a certain recurrence statement in ergodic theory, like that of Theorem 1.3. The second is showing that this statement implies a corresponding statement about subsets of the integers. We now make this more precise.

To use ergodic theory to show that some intersection of sets has positive measure, it is natural to average the expression under consideration. This leads us to the strong form of Furstenberg's multiple recurrence:

Theorem 2.1 (Furstenberg [16]). *Let $(X,\mathcal{X},\mu,T)$ be a system and let $A\in\mathcal{X}$ with $\mu(A)>0$. Then for any $k\geq 1$,*

$$\liminf_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\mu(A\cap T^{-n}A\cap T^{-2n}A\cap\cdots\cap T^{-kn}A) \tag{2.1}$$

is positive.

In particular, this implies the existence of infinitely many $n\in\mathbb{N}$ such that the intersection in (1.1) is positive and Theorem 1.3 follows. In Section 3, we discuss how to prove Theorem 2.1.

2.2. The correspondence principle. The second major component in Furstenberg's proof is using this multiple recurrence statement to derive a statement about integers, such as Szemerédi's theorem. This is the content of Furstenberg's correspondence principle:

Theorem 2.2 (Furstenberg [16,17]**).** *Let $E \subset \mathbb{Z}$ have positive upper density. There exist a system $(X, \mathcal{X}, \mu, T)$ and a set $A \in \mathcal{X}$ with $\mu(A) = d^*(E)$ such that*

$$\mu(T^{-m_1}A \cap \cdots \cap T^{-m_k}A) \le d^*\big((E+m_1)\cap\cdots\cap(E+m_k)\big)$$

for all $k \in \mathbb{N}$ and all $m_1, \ldots, m_k \in \mathbb{Z}$.

Proof. Let $X = \{0,1\}^{\mathbb{Z}}$ be endowed with the product topology and the shift map T given by $Tx(n) = x(n+1)$ for all $n \in \mathbb{Z}$. A point of X is thus a sequence $\mathbf{x} = \{x(n)\}_{n\in\mathbb{Z}}$, and the distance between two points $\mathbf{x} = \{x(n)\}_{n\in\mathbb{Z}}$, $\mathbf{y} = \{y(n)\}_{n\in\mathbb{Z}}$ is defined to be 0 if $\mathbf{x} = \mathbf{y}$ and to be 2^{-k} if $\mathbf{x} \ne \mathbf{y}$ and $k = \min\{|n| : x(n) \ne y(n)\}$. Define $\mathbf{a} = \{a(n)\}_{n\in\mathbb{Z}} \in \{0,1\}^{\mathbb{Z}}$ by

$$a(n) = \begin{cases} 1 & \text{if } n \in E \\ 0 & \text{otherwise} \end{cases}$$

and let $A = \{\mathbf{x} \in X : x(0) = 1\}$. Thus A is a clopen (closed and open) set.

The set $A \in \mathcal{X}$ plays the same role as the set $E \subset \mathbb{Z}$: for all $n \in \mathbb{Z}$,

$$T^n\mathbf{a} \in A \text{ if and only if } n \in E.$$

By definition of $d^*(E)$, there exist sequences $\{M_i\}$ and $\{N_i\}$ of integers with $N_i \to \infty$ such that

$$\lim_{i\to\infty} \frac{1}{N_i}\big|E \cap [M_i, M_i+N_i-1]\big| \to d^*(E).$$

It follows that

$$\lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_A(T^n\mathbf{a}) = \lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_E(n) = d^*(E).$$

Let $\mathcal{C}$ be the countable algebra generated by cylinder sets, meaning sets that are defined by specifying finitely many coordinates of each element and leaving the others free. We can define an additive measure μ on $\mathcal{C}$ by

$$\mu(B) = \lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_B(T^n\mathbf{a}),$$

where we pass, if necessary, to subsequences $\{N_i\}$, $\{M_i\}$ such that this limit exists for all $B \in \mathcal{C}$. (Note that $\mathcal{C}$ is countable and so by diagonalization we can arrange it such that this limit exists for all elements of $\mathcal{C}$.)

We can extend the additive measure to a σ-additive measure μ on all Borel sets $\mathcal{X}$ in X, which is exactly the σ-algebra generated by $\mathcal{C}$. Then μ is an invariant measure, meaning that for all $B \in \mathcal{C}$,

$$\mu(T^{-1}B) = \lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_B(T^{n-1}\mathbf{a}) = \mu(B).$$

Furthermore,

$$\mu(A) = \lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_A(T^n\mathbf{a}) = d^*(E).$$

If $m_1, \dots, m_k \in \mathbb{Z}$, then the set $T^{-m_1}A \cap \dots \cap T^{-m_k}A$ is a clopen set, its indicator function is continuous, and

$$\begin{aligned}\mu(T^{-m_1}A \cap \dots \cap T^{-m_k}A) &= \lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_{T^{-m_1}A\cap\dots\cap T^{-m_k}A}(T^n\mathbf{a}) \\ &= \lim_{i\to\infty} \frac{1}{N_i} \sum_{n=M_i}^{M_i+N_i-1} \mathbf{1}_{(E+m_1)\cap\dots\cap(E+m_k)}(n) \\ &\le d^*\big((E+m_1)\cap\dots\cap(E+m_k)\big). \qquad \square\end{aligned}$$

We use this to deduce Szemerédi's theorem from Theorem 1.3. As in the proof of the correspondence principle, define $\mathbf{a} \in \{0,1\}^{\mathbb{Z}}$ by

$$a(n) = \begin{cases} 1 & \text{if } n \in E \\ 0 & \text{otherwise,} \end{cases}$$

and set $A = \{\mathbf{x} \in \{0,1\}^{\mathbb{Z}}\ x(0) = 1\}$. Thus $T^n\mathbf{a} \in A$ if and only if $n \in E$.

By Theorem 1.3, there exists $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \dots \cap T^{-kn}A) > 0.$$

Therefore for some $m \in \mathbb{N}$, $T^m\mathbf{a}$ enters this multiple intersection and so

$$a(m) = a(m+n) = a(m+2n) = \dots = a(m+kn) = 1.$$

But this means that

$$m, m+n, m+2n, \dots, m+kn \in E$$

and so we have found an arithmetic progression of length $k+1$ in E.

3. Convergence of multiple ergodic averages

3.1. Convergence along arithmetic progressions. Furstenberg's multiple recurrence theorem left open the question of the existence of the limit in (2.1). More generally, one can ask if given a system $(X, \mathcal{X}, \mu, T)$ and $f_1, f_2, \dots, f_k \in L^\infty(\mu)$, does

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) \cdot f_2(T^{2n}x) \cdot \dots \cdot f_k(T^{kn}x) \tag{3.1}$$

exist? Moreover, we can ask in what sense (in $L^2(\mu)$ or pointwise) does this limit exist, and if it does exist, what can be said about the limit? Setting each function f_i to be the indicator function of a measurable set A, we are back in the context of Furstenberg's theorem.

For $k = 1$, existence of the limit in $L^2(\mu)$ is the mean ergodic theorem of von Neumann. In Section 4.2, we give a proof of this statement. For $k = 2$, existence of the limit in $L^2(\mu)$ was proven by Furstenberg [16] as part of his proof of Szemerédi's theorem. Furthermore, in the same paper he showed the existence of the limit in $L^2(\mu)$ in a weak mixing system for arbitrary k; we define weak mixing in Section 5.5 and outline the proof for this case.

For $k \geq 3$, the proof of existence of the limit in 3.1 requires a more subtle understanding of measure preserving systems, and we begin discussing this case in Section 5.8. Under some technical hypotheses, the existence of the limit in $L^2(\mu)$ for $k = 3$ was first proven by Conze and Lesigne (see [8, 9]), then by Furstenberg and Weiss [22], and in the general case by Host and Kra [32]. More generally, we showed the existence of the limit for all $k \in \mathbb{N}$:

Theorem 3.1 (Host and Kra [34]**).** *Let $(X, \mathcal{X}, \mu, T)$ be a system, let $k \in \mathbb{N}$, and let $f_1, f_2, \ldots, f_k \in L^\infty(\mu)$. Then the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^n x) \cdot f_2(T^{2n} x) \cdot \cdots \cdot f_k(T^{kn} x)$$

converge in $L^2(\mu)$ as $N \to \infty$.

Such a convergence result for a finite system is trivial. For example, if $X = \mathbb{Z}/N\mathbb{Z}$, then $\mathcal{X}$ consists of all partitions of X and μ is the uniform probability measure, meaning that the measure of a set is proportional to the cardinality of the set. The transformation T is given by $Tx = x + 1 \mod N$. It is then trivial to check the convergence of the average in (3.1). However, although the ergodic theory is trivial in this case, there are common themes to be explored. Throughout these notes, an effort is made to highlight the connection with recent advances in additive combinatorics (see [39] for more on this connection). Of particular interest is the role played by nilpotent groups, and homogeneous spaces of nilpotent groups, in the proof of the ergodic statement.

Much of the present notes is devoted to understanding the ingredients in the proof of Theorem 3.1, and the role of nilpotent groups in this proof. Other expository accounts of this proof can be found in [31, 40]. In this context, 2-step nilpotent groups first appeared in the work of Conze-Lesigne in their proof of convergence for $k = 3$, and a $(k-1)$-step nilpotent group plays a similar role in convergence for the average in (3.1). Nilpotent groups also play some role in the combinatorial setup, and this has been recently verified by Green and Tao (see [26–28]) for progressions of length 4 (which corresponds to the case $k = 3$ in (3.1)). For more on this connection, see the lecture notes of Ben Green in this volume.

3.2. Other results. Using ergodic theory, other patterns have been shown to exist in sets of positive upper density and we discuss these results in Section 9. We briefly summarize these results. A striking example is the theorem of Bergelson and Leibman [6] showing the existence of polynomial patterns in such sets. Analogous to the linear average corresponding to arithmetic progressions, existence of the associated polynomial averages was shown in [35, 45]. One can also average along "cubes"; existence of these averages and a corresponding combinatorial statement was shown in [34]. For commuting transformations, little is known and these partial results are summarized in Section 9.1. An explicit formula for the limit in (3.1) was given by Ziegler [56], who also has recently given a second proof [57] of Theorem 3.1.

4. Single convergence (the case $k = 1$)

4.1. Poincaré recurrence. The case $k = 1$ in Furstenberg's multiple recurrence (Theorem 1.3) is Poincaré recurrence:

Theorem 4.1 (Poincaré [49]**).** *If $(X, \mathcal{X}, \mu, T)$ is a system and $A \in \mathcal{X}$ with $\mu(A) > 0$, then there exist infinitely many $n \in \mathbb{N}$ such that $\mu(A \cap T^{-n}A) > 0$.*

Proof. Let $F = \{x \in A : T^{-n}x \notin A \text{ for all } n \geq 1\}$. Thus $F \cap T^{-n}F = \varnothing$ for all $n \geq 1$, and so for all integers $n \neq m$,

$$T^{-m}A \cap T^{-n}A = \varnothing.$$

In particular, $F, T^{-1}F, T^{-2}F, \dots$ are all pairwise disjoint sets and each set in this sequence has measure equal to $\mu(F)$. If $\mu(F) > 0$, then

$$\mu\Big(\bigcup_{n\geq 0} T^{-n}F\Big) = \sum_{n\geq 0} \mu(F) = \infty,$$

a contradiction of μ being a probability measure.

Therefore $\mu(F) = 0$ and the statement is proven. □

In fact the same proof shows a bit more: by a simple modification of the definition of F, we have that μ-almost every $x \in A$ returns to A infinitely often.

4.2. The von Neumann ergodic theorem. Although the proof of Poincaré recurrence is simple, unfortunately there seems to be no way to generalize it to show multiple recurrence. In order to find a method that generalizes for multiple recurrence, we prove a stronger statement than Poincaré recurrence, taking the average of the expression under consideration and showing that the lim inf of this average is positive. It is not any harder (for $k = 1$ only!) to show that the limit of this average exists (and is positive). This is the content of the von Neumann mean ergodic theorem. We first give the statement in a general Hilbert space:

Theorem 4.2 (von Neumann [55]**).** *If U is an isometry of a Hilbert space $\mathcal{H}$ and P is orthogonal projection onto the U-invariant subspace $\mathcal{I} = \{f \in \mathcal{H} : Uf = f\}$, then for all $f \in \mathcal{H}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} U^n f = Pf. \tag{4.1}$$

Thus the case $k = 1$ in Theorem 3.1 is an immediate corollary of the von Neumann ergodic theorem.

Proof. If $f \in \mathcal{I}$, then

$$\frac{1}{N} \sum_{n=0}^{N-1} U^n f = f$$

for all $N \in \mathbb{N}$ and so obviously the average converges to f. On the other hand, if $f = g - Ug$ for some $g \in \mathcal{H}$, then

$$\sum_{n=0}^{N-1} U^n f = g - U^N g$$

and so the average converges to 0 as $N \to \infty$. Set $\mathcal{J} = \{g - Ug : g \in \mathcal{H}\}$. If $f_k \in \mathcal{J}$ and $f_k \to f \in \overline{\mathcal{J}}$, then

$$\begin{aligned}\left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n f\right\| &\leq \left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n(f - f_k)\right\| + \left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n(f_k)\right\| \\ &\leq \left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n\right\| \cdot \|f - f_k\| + \left\|\frac{1}{N}\sum_{n=0}^{N-1} U^n(f_k)\right\|.\end{aligned}$$

Thus for $f \in \bar{\mathcal{J}}$, the average $(1/N)\sum_{n=0}^{N-1} U^n f$ also converges to 0 as $N \to \infty$.

We now show that an arbitrary $f \in \mathcal{H}$ can be written as a combination of functions which exhibit these behaviors, meaning that any $f \in \mathcal{H}$ can be written as $f = f_1 + f_2$ for some $f_1 \in \mathcal{I}$ and $f_2 \in \bar{\mathcal{J}}$. If $h \in \mathcal{J}^\perp$, then for all $g \in \mathcal{H}$,

$$0 = \langle h, g - Ug\rangle = \langle h, g\rangle - \langle h, Ug\rangle = \langle h, g\rangle - \langle U^* h, g\rangle = \langle h - U^* h, g\rangle$$

and so $h = U^* h$ and $h = Uh$. Conversely, reversing the steps we have that if $h \in \mathcal{I}$, then $h \in \mathcal{J}^\perp$.

Since $\bar{\mathcal{J}}^\perp = \mathcal{J}^\perp$, we have that

$$\mathcal{H} = \mathcal{I} \oplus \bar{\mathcal{J}}.$$

Thus writing $f = f_1 + f_2$ with $f_1 \in \mathcal{I}$ and $f_2 \in \bar{\mathcal{J}}$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} U^n f = \frac{1}{N}\sum_{n=0}^{N-1} U^n f_1 + \frac{1}{N}\sum_{n=0}^{N-1} U^n f_2.$$

As $N \to \infty$, the first sum converges to the identity and the second sum to 0. $\square$

The idea behind the proof of von Neumann's Theorem is simple: decompose an arbitrary function into two pieces and then show that the limit exists for each of these pieces. This sort of decomposition is used (in some sense) in Furstenberg's proof of Theorem 2.1, the original proof of Szemerédi's theorem, the convergence result of Theorem 3.1, and in the recent results of Green and Tao on patterns in the prime numbers.

Under a mild hypothesis on the system, we have an explicit formula for the limit (4.1). Let $(X, \mathcal{X}, \mu, T)$ be a system. A subset $A \subset X$ is said to be *invariant* if $T^{-1}A = A$. The invariant sets form a sub-σ-algebra $\mathcal{I}$ of $\mathcal{X}$. The system $(X, \mathcal{X}, \mu, T)$ is said to be *ergodic* if $\mathcal{I}$ is trivial, meaning that every invariant set has either measure 0 or measure 1.

A measure preserving transformation $T\colon X \to X$ defines a linear operator $U_T\colon L^2(\mu) \to L^2(\mu)$ by

$$(U_T f)(x) = f(Tx).$$

It is easy to check that the operator U_T is a unitary operator (meaning its adjoint is equal to its inverse). In a standard abuse of notation, we use the same letter to denote the operator and the transformation, writing $Tf(x) = f(Tx)$ instead of the more cumbersome $U_T f(x) = f(Tx)$.

Applying von Neumann's ergodic theorem in a measure preserving system, we have:

Corollary 4.3. *If $(X, \mathcal{X}, \mu, T)$ is a system and $f \in L^2(\mu)$, then*

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$$

converges in $L^2(\mu)$, as $N \to \infty$, to a T-invariant function $\tilde{f}$. If the system is ergodic, then the limit is the constant function $\int f \,\mathrm{d}\mu$.

Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system and let $A, B \in \mathcal{X}$. Taking $f = \mathbf{1}_A$ in Corollary 4.3 and integrating with respect to μ over a set B, we have that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int_B \mathbf{1}_A(T^n x)\,\mathrm{d}\mu(x) = \int_B \Big(\int \mathbf{1}_A(y)\,\mathrm{d}\mu(y)\Big)\,\mathrm{d}\mu(x).$$

Rewriting this, we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{-n}B) = \mu(A)\mu(B).$$

In fact, one can check that this condition holds for all $A, B \in \mathcal{X}$ if and only if the system is ergodic.

As already discussed, convergence in the case of the finite system $\mathbb{Z}/N\mathbb{Z}$ with the transformation of adding $1 \mod N$ is trivial. Furthermore this system is ergodic. More generally, any permutation on $\mathbb{Z}/N\mathbb{Z}$ can be expressed as a product of disjoint cyclic permutations. These permutations are the "indecomposable" invariant subsets of an arbitrary transformation on $\mathbb{Z}/N\mathbb{Z}$ and the restriction of the transformation to one of these subsets is ergodic.

This idea of dividing a space into indecomposable components generalizes: an arbitrary measure preserving system can be decomposed into, perhaps continuously many, indecomposable components, and these are exactly the ergodic ones. Using this *ergodic decomposition* (see, for example, [10]), instead of working with an arbitrary system, we reduce most of the recurrence and convergence questions we consider here to the same problem in an ergodic system.

5. Double convergence (the case $k = 2$)

5.1. A model for double convergence. We now turn to the case of $k = 2$ in Theorem 3.1, and study convergence of the double average

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^n x) \cdot f_2(T^{2n} x) \tag{5.1}$$

for bounded functions f_1 and f_2. Our goal is to explain how a simple class of systems, the rotations, suffice to understand convergence for the double average.

First we explicitly define what is meant by a rotation. Let G be a compact abelian group, with Borel σ-algebra $\mathcal{B}$, Haar measure m, and fix some $\alpha \in G$. Define $T\colon G \to G$ by

$$Tx = x + \alpha.$$

The system $(G, \mathcal{B}, m, T)$ is called a *group rotation.* It is ergodic if and only if $\mathbb{Z}\alpha$ is dense in G. For example, when X is the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ and $\alpha \notin \mathbb{Q}$, the rotation by α is ergodic.

The double average is the simplest example of a *nonconventional* ergodic average: even for an ergodic system, the limit is not necessarily constant. This sort of behavior does not occur for the single average of von Neumann's theorem, where we have seen that the limit is constant in an ergodic system. Even for the simple example of an an ergodic rotation, the limit of the double average is not constant:

Example 5.1. Let $X = \mathbb{T}$, with Borel σ-algebra and Haar measure, and let $T\colon X \to X$ be the rotation $Tx = x + \alpha \mod 1$. Setting $f_1(x) = \exp(4\pi \mathrm{i} x)$ and $f_2(x) = \exp(-2\pi \mathrm{i} x)$, then for all $n \in \mathbb{N}$,

$$f_1(T^n x) \cdot f_2(T^{2n} x) = \overline{f_2(x)}.$$

In particular, the double average (5.1) for these functions converges to a nonconstant function.

More generally, if $\alpha \notin \mathbb{Q}$ and $f_1, f_2 \in L^\infty(\mu)$, the double average converges to

$$\int_{\mathbb{T}} f_1(x+t) \cdot f_2(x+2t)\, \mathrm{d}t.$$

We shall see that Fourier analysis suffices to understand this average. By taking both functions to be the indicator function of a set with positive measure and integrating over this set, we then have that Fourier analysis suffices for the study of arithmetic progressions of length 3. This gives a complete proof of Roth's Theorem via ergodic theory. Later we shall see that more powerful methods are needed to understand the average along longer progressions. In a similar vein, rotations are the model for an ergodic average with 3 terms, but are not sufficient for more terms. We introduce some terminology to make these notions more precise.

5.2. Factors. For the remainder of this section, we assume that $(X, \mathcal{X}, \mu, T)$ is an ergodic system.

A *factor* of a system $(X, \mathcal{X}, \mu, T)$ can be defined in one of several equivalent ways. It is a T-invariant sub-σ-algebra $\mathcal{Y}$ of $\mathcal{X}$. A second characterization is that a factor is a system $(Y, \mathcal{Y}, \nu, S)$ and a measurable map $\pi\colon X \to Y$, the *factor map*, such that $\mu \circ \pi^{-1} = \nu$ and $S \circ \pi = \pi \circ T$ for μ-almost every $x \in X$. A third characterization is that a factor is a T-invariant subalgebra $\mathcal{F}$ of $L^\infty(\mu)$. One can check that the first two definitions agree by identifying $\mathcal{Y}$ with $\pi^{-1}(\mathcal{Y})$, and that the first and third agree by identifying $\mathcal{F}$ with $L^\infty(\mathcal{Y})$. When any of these conditions holds, we say that Y, or the appropriate sub-σ-algebra, is a factor of X and write $\pi\colon X \to Y$ for the factor map. We usually make use of a slight (and standard) abuse of notation, using the same letter T to denote both the transformation in the original system and the transformation in the factor system. If the factor map $\pi\colon X \to Y$ is also injective, we say that the two systems $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are *isomorphic*.

For example, if $(X, \mathcal{X}, \mu, T)$ and $(Y, \mathcal{Y}, \nu, S)$ are systems, then each is a factor of the product system $(X \times Y, \mathcal{X} \times \mathcal{Y}, \mu \times \nu, T \times S)$ and the associated factor map for each is projection onto the appropriate coordinate.

A more interesting example can be given in the system $X = \mathbb{T} \times \mathbb{T}$, with Borel σ-algebra and Haar measure, and transformation $T\colon X \to X$ given by

$$T(x, y) = (x + \alpha, y + x).$$

Then $\mathbb{T}$ with the rotation $x \mapsto x + \alpha$ is a factor of X.

5.3. Conditional expectation. If $\mathcal{Y}$ is a T-invariant sub-σ-algebra of $\mathcal{X}$ and $f \in L^2(\mu)$, the *conditional expectation* $\mathbb{E}(f \mid \mathcal{Y})$ *of* f *with respect to* $\mathcal{Y}$ is the function on Y defined by $\mathbb{E}(f \mid Y) \circ \pi = \mathbb{E}(f \mid \mathcal{Y})$. It is characterized as the $\mathcal{Y}$-measurable function on X such that

$$\int_X f(x) \cdot g(\pi(x))\,\mathrm{d}\mu(x) = \int_Y \mathbb{E}(f \mid \mathcal{Y})(y) \cdot g(y)\,\mathrm{d}\nu(y)$$

for all $g \in L^\infty(\nu)$ and it satisfies the identities

$$\int \mathbb{E}(f \mid \mathcal{Y})\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu$$

and

$$T\mathbb{E}(f \mid \mathcal{Y}) = \mathbb{E}(Tf \mid \mathcal{Y}).$$

As an example, take $X = \mathbb{T} \times \mathbb{T}$ endowed with the transformation $(x, y) \mapsto (x + \alpha, y + x)$. We have a factor $Z = \mathbb{T}$ endowed with the map $x \mapsto x + \alpha$. Considering $f(x, y) = \exp(x) + \exp(y)$, we have that $\mathbb{E}(f \mid \mathcal{Z}) = \exp(x)$. The factor σ-algebra $\mathcal{Z}$ is the σ-algebra of sets that depend only on the x coordinate.

5.4. Characteristic factors. For $f_1, \ldots, f_k \in L^\infty(\mu)$, we are interested in convergence in $L^2(\mu)$ of:

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \cdots \cdot T^{kn} f_k. \tag{5.2}$$

Instead of working with the whole system $(X, \mathcal{X}, \mu, T)$, it turns out that it is easier to find some factor of the system that characterizes this average, meaning find some well chosen factor such that we can prove convergence of the average in this factor and this convergence suffices to understand convergence of the same average in the original system. This motivates the following definition.

A factor Y of X is *characteristic* for the average (5.2) if the difference between (5.2) and

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n \mathbb{E}(f_1 \mid \mathcal{Y}) \cdot T^{2n} \mathbb{E}(f_2 \mid \mathcal{Y}) \cdot \cdots \cdot T^{kn} \mathbb{E}(f_k \mid \mathcal{Y})$$

(the same average with $\mathbb{E}(f_i \mid \mathcal{Y})$ substituted for f_i for $i = 1, 2, \ldots, k$) converges to 0 in $L^2(\mu)$ as $N \to \infty$. Rewriting the average (5.2) in terms of $f_i - \mathbb{E}(f_i \mid \mathcal{Y})$ for $i = 1, 2, \ldots, k$, it follows that the factor Y is characteristic for the average (5.2) if and only if the average in (5.2) converges to 0 as $N \to \infty$ when $\mathbb{E}(f_i \mid \mathcal{Y}) = 0$ for some $i \in \{1, 2, \ldots, k\}$.

The idea of a characteristic factor is that the limiting behavior of the average under study can be reduced to that of a factor of the system. We have already seen an example of a characteristic factor in the von Neumann ergodic theorem: the trivial factor, consisting only of the constants, is characteristic. (Recall that we have assumed that the system is ergodic.)

By definition, the whole system is always a characteristic factor. Of course nothing is gained by using such a characteristic factor, and the notion only becomes useful when we can find a characteristic factor that has useful geometric and/or algebraic properties. A very short outline of the proof of convergence of the average (5.2) is as follows: find a characteristic factor that has sufficient structure so as to allow one to prove convergence. We return to this idea later.

The definition of a characteristic factor can be extended for any other average under consideration, with the obvious changes: the limit remains unchanged when each function is replaced by its conditional expectation on this factor. This notion has been implicit in the literature since Furstenberg's proof of Szemerédi's theorem, but the terminology we now use was only introduced more recently in [22].

5.5. Weak mixing systems. The system $(X,\mathcal{X},\mu,T)$ is *weak mixing* if for all $A,B\in\mathcal{X}$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu(T^{-n}A\cap B)-\mu(A)\mu(B)\right|=0.$$

Any weak mixing system is ergodic, and the example of an irrational circle rotation shows that converse does not hold. There are many equivalent formulations of weak mixing, and we give a few (see, for example [10]):

Proposition 5.2. *Let $(X,\mathcal{X},\mu,T)$ be a system. The following are equivalent:*

(1) *$(X,\mathcal{X},\mu,T)$ is weak mixing.*
(2) *There exists $J\subset\mathbb{N}$ of density zero such that for all $A,B\in\mathcal{X}$*

$$\mu(T^{-n}A\cap B)\to\mu(A)\mu(B)\quad\textit{as } n\to\infty \textit{ and } n\notin J.$$

(3) *For all $A,B,C\in\mathcal{X}$ with $\mu(A)\mu(B)\mu(C)>0$, there exists $n\in\mathbb{N}$ such that*

$$\mu(A\cap T^{-n}B)\mu(A\cap T^{-n}C)>0.$$

(4) *The system $(X\times X,\mathcal{X}\times\mathcal{X},\mu\times\mu,T\times T)$ is ergodic.*

Any system exhibiting rotational behavior (for example a rotation on a circle, or a system with a nontrivial circle rotation as a factor) is not weak mixing. We have already seen in Example 5.1 that weak mixing, or lack thereof, has an effect on multiple averages. We give a second example to highlight this effect:

Example 5.3. Suppose that $X=X_1\cup X_2\cup X_3$ with $T(X_1)=X_2$, $T(X_2)=X_3$ and $T(X_3)=X_1$, and further suppose that T^3 restricted to X_i, for $i=1,2,3$, is weak mixing. For the double average

$$\frac{1}{N}\sum_{n=0}^{N-1}f_1(T^nx)\cdot f_2(T^{2n}x),$$

where $f_1,f_2\in L^\infty(\mu)$, if $x\in X_1$, this average converges to

$$\frac{1}{3}\Big(\int_{X_1}f_1\,\mathrm{d}\mu\int_{X_1}f_2\,\mathrm{d}\mu+\int_{X_2}f_1\,\mathrm{d}\mu\int_{X_3}f_2\,\mathrm{d}\mu+\int_{X_3}f_1\,\mathrm{d}\mu\int_{X_2}f_2\,\mathrm{d}\mu\Big).$$

A similar expression with obvious changes holds for $x\in X_2$ or $x\in X_3$.

The main point is that in both Example 5.1 and in Example 5.3 (for the double average) the limit depends on the rotational behavior of the system. Example 5.3 lacks weak mixing and so has a nontrivial rotation factor. We now formalize this notion.

5.6. Kronecker factor. The *Kronecker factor* $(Z_1, \mathcal{Z}_1, m, T)$ of $(X, \mathcal{X}, \mu, T)$ is the sub-σ-algebra of $\mathcal{X}$ spanned by the eigenfunctions. (Recall that there is a unitary operator U_T associated to the measure preserving transformation T. By eigenfunctions, we refer to the eigenfunctions of this unitary operator.) A classical result is that the Kronecker factor can be given the structure of a group rotation:

Theorem 5.4 (Halmos and von Neumann [30]). *The Kronecker factor of a system is isomorphic to a system $(Z_1, \mathcal{Z}_1, m, T)$, where Z_1 is a compact abelian group, $\mathcal{Z}_1$ is its Borel σ-algebra, m is the Haar measure, and $Tx = x + \alpha$ for some fixed $\alpha \in Z_1$.*

We use $\pi_1 \colon X \to Z_1$ to denote the factor map from the system $(X, \mathcal{X}, \mu, T)$ to its Kronecker factor $(Z_1, \mathcal{Z}_1, m, T)$. Then any eigenfunction f of X takes the form

$$f(x) = c\gamma\big(\pi_1(x)\big),$$

where c is a constant and $\gamma \in \widehat{Z_1}$ is a character of Z_1.

We give two examples of Kronecker factors:

Example 5.5. If $X = \mathbb{T} \times \mathbb{T}$, $\alpha \in \mathbb{T}$, and $T \colon X \to X$ is the map

$$T(x, y) = (x + \alpha, y + x),$$

then the rotation $x \mapsto x + \alpha$ on $\mathbb{T}$ is the Kronecker factor of X. It corresponds to the pure point spectrum. (The spectrum in the orthogonal complement of the Kronecker factor is countable Lebesgue.)

Example 5.6. If $X = \mathbb{T}^3$, $\alpha \in \mathbb{T}$, and $T \colon X \to X$ is the map

$$T(x, y, z) = (x + \alpha, y + x, z + y),$$

then again the rotation $x \mapsto x + \alpha$ on $\mathbb{T}$ is the Kronecker factor of X. This example has the same pure point spectrum as the first example, but the system in the first example is a factor of the system in the second example.

The Kronecker factor can be used to give another characterization of weak mixing:

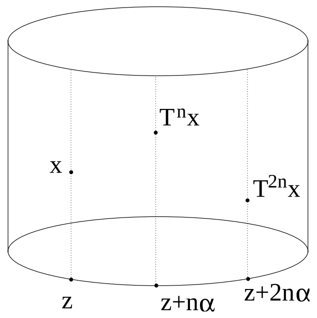
Theorem 5.7 (Koopman and von Neumann [38]). *A system is not weak mixing if and only if it has a nontrivial factor which is a rotation on a compact abelian group.*

The largest of these factors is exactly the Kronecker factor.

5.7. Convergence for $\boldsymbol{k = 2}$. If we take into account the rotational behavior in a system, meaning the existence of a nontrivial Kronecker factor, then we can understand the limit of the double average

$$\frac{1}{N} \sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2. \tag{5.3}$$

An obvious constraint is that for μ-almost every x, the triple $(x, T^n x, T^{2n} x)$ projects to an arithmetic progression in the Kronecker factor $\mathcal{Z}_1$. Assuming that the Kronecker factor is a circle with rotation by some α, we can think of each point in the progression $(x, T^n x, T^{2n} x)$ as located on the fiber above the corresponding point in the progression $(z, z + \alpha, z + 2\alpha)$:



Furstenberg proved that this obvious restriction is the only restriction, showing that to prove convergence of the double average, one can assume that the system is an ergodic rotation on a compact abelian group:

Theorem 5.8 (Furstenberg [16]**).** *If $(X, \mathcal{X}, \mu, T)$ is an ergodic system, $(Z_1, \mathcal{Z}_1, m, T)$ is its Kronecker factor, and $f_1, f_2, \in L^\infty(\mu)$, then the limit*

$$\Big\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 - \frac{1}{N} \sum_{n=0}^{N-1} T^n \mathbb{E}(f_1 \mid \mathcal{Z}_1) \cdot T^{2n} \mathbb{E}(f_2 \mid \mathcal{Z}_1) \Big\|_{L^2(\mu)}$$

tends to 0 *as* $N \to \infty$.

In our terminology, this theorem can be quickly summarized: the Kronecker factor is characteristic for the double average. To prove the theorem, we use a standard trick for averaging, which is an iterated use of a variation of the van der Corput lemma on differences. (See [41] for uses of the van der Corput Lemma in number theory and [2] for its introduction to uses in ergodic theory.)

Lemma 5.9 (van der Corput). *Let $\{u_n\}$ be a sequence in a Hilbert space with $\|u_n\| \leq 1$ for all $n \in \mathbb{N}$. For $h \in \mathbb{N}$, set*

$$\gamma_h = \limsup_{N \to \infty} \Big| \frac{1}{N} \sum_{n=0}^{N-1} \langle u_{n+h}, u_n \rangle \Big|.$$

Then

$$\limsup_{N \to \infty} \Big\| \frac{1}{N} \sum_{n=0}^{N-1} u_n \Big\|^2 \leq \limsup_{H \to \infty} \frac{1}{H} \sum_{h=0}^{H-1} \gamma_h.$$

Proof. Given $\varepsilon > 0$ and $H \in \mathbb{N}$, for N sufficiently large we have that

$$\Big| \frac{1}{N} \sum_{n=0}^{N-1} u_n - \frac{1}{N} \frac{1}{H} \sum_{n=0}^{N-1} \sum_{h=0}^{H-1} u_{n+h} \Big| < \varepsilon.$$

By convexity,

$$\begin{aligned} \Big\| \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{H} \sum_{h=0}^{H-1} u_{n+h} \Big\|^2 &\leq \frac{1}{N} \sum_{n=0}^{N-1} \Big\| \frac{1}{H} \sum_{h=0}^{H-1} u_{n+h} \Big\|^2 \\ &= \frac{1}{N} \frac{1}{H^2} \sum_{n=0}^{N-1} \sum_{h_1, h_2=0}^{H-1} \langle u_{n+h_1}, u_{n+h_2} \rangle \end{aligned}$$

and this approaches

$$\frac{1}{H^2}\sum_{h_1,h_2}^{H-1}\gamma_{h_1-h_2}$$

as $N\to\infty$. But the assumption implies that this approaches 0 as $H\to\infty$. □

We now use this in the proof of Furstenberg's theorem:

PROOF OF THEOREM 5.8. By replacing f_1 by $f_1-\mathbb{E}(f_1\mid\mathcal{Z}_1)$ and f_2 by $f_2-\mathbb{E}(f_2\mid\mathcal{Z}_1)$, it suffices to show that if some f_i, for $i=1,2$ satisfies $\mathbb{E}(f_i\mid\mathcal{Z}_1)=0$, then the doubule average converges to 0. Without loss, we assume that $\mathbb{E}(f\mid\mathcal{Z}_1)=0$.

Set $u_n=T^nf_1\cdot T^{2n}f_2$. Then

$$\begin{aligned}\langle u_n,u_{n+h}\rangle &= \int T^nf_1\cdot T^{2n}f_2\cdot T^{n+h}\overline{f_1}\cdot T^{2n+2h}\overline{f_2}\,\mathrm{d}\mu\\ &= \int (f_1\cdot T^h\overline{f_1})\cdot T^n(f_2\cdot T^{2h}\overline{f_2})\,\mathrm{d}\mu.\end{aligned}$$

Thus

$$\frac{1}{N}\sum_{n=0}^{N-1}\langle u_n,u_{n+h}\rangle = \left(\int f_1\cdot T^h\overline{f_1}\,\mathrm{d}\mu\right)\frac{1}{N}\sum_{n=0}^{N-1}T^n(f_2\cdot T^{2h}\overline{f_2})\,\mathrm{d}\mu.$$

By the von Neumann ergodic theorem (Theorem 4.2) applied to the second term, the limit

$$\gamma_h=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\langle u_n,u_{n+h}\rangle$$

exists. Moreover, it is equal to

$$\gamma_h=\int f_1\cdot T^h\overline{f_1}\cdot\mathbb{P}(f_2\cdot T^{2h}\overline{f_2})\,\mathrm{d}\mu, \tag{5.4}$$

where $\mathbb{P}$ is projection onto the T-invariant functions of $L^2(\mu)$. Since T is ergodic, $\mathbb{P}$ is projection onto the constant functions. But since $\mathbb{E}(f_1\mid\mathcal{Z}_1)=0$, f_1 is orthogonal to the constant functions and so by averaging over h, we have that

$$\lim_{H\to\infty}\frac{1}{H}\sum_{h=0}^{H-1}\gamma_h=0.$$

By the van der Corput lemma, it follows that the double average also converges to 0. □

Furstenberg used a similar argument combined with induction to show that in a weak mixing system, the average (5.2) converges to the product of the integrals in $L^2(\mu)$ for all $k\geq 1$. This is one of the (simpler) steps in the proof of the Furstenberg's multiple recurrence theorem (Theorem 1.3) and gives a proof of multiple recurrence for weakly mixing systems. However, much more is needed to prove Theorem 1.3 in an arbitrary system; this is carried out by showing that for any function, the average along arithmetic progressions can be decomposed into two pieces, one of which has a generalized weak mixing property and the other of which is rigid in some sense. We have already seen a simple example of such a decomposition, in the proof of the von Neumann ergodic theorem. Some sort of decomposition is behind all of the multiple recurrence and convergence results we discuss.

We now return to showing that a set of integers with positive upper density contains arithmetic progressions of length three (Roth's theorem). By Furstenberg's correspondence principle it suffices to show double recurrence:

Theorem 5.10 (Theorem 1.3 for $k = 2$). *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system, and let $A \in \mathcal{X}$ with $\mu(A) > 0$. There exists $n \in \mathbb{N}$ with*

$$\mu(A \cap T^{-n}A \cap T^{-2n}A) > 0.$$

PROOF. Let $f = \mathbf{1}_A$. Then

$$\mu(A \cap T^{-n}A \cap T^{-2n}A) = \int f \cdot T^n f \cdot T^{2n} f \, d\mu.$$

It suffices to show that

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f \cdot T^n f \cdot T^{2n} f \, d\mu$$

is positive. However, we will show the stronger statement that the limit exists and is positive, rather than just the lim sup is positive.[3]

By Theorem 5.8, the limiting behavior of the double average $(1/N)\sum_{n=0}^{N-1} T^n f \cdot T^{2n} f$ is unchanged if f is replaced by $\mathbb{E}(f \mid \mathcal{Z}_1)$. Multiplying by f and integrating, it thus suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int f \cdot T^n \mathbb{E}(f \mid \mathcal{Z}_1) \cdot T^{2n}\mathbb{E}(f \mid \mathcal{Z}_1) \, \mathrm{d}\mu \tag{5.5}$$

exists and is positive. Since $\mathcal{Z}_1$ is T-invariant, $T^n\mathbb{E}(f \mid \mathcal{Z}_1) \cdot T^{2n}\mathbb{E}(f \mid \mathcal{Z}_1)$ is measurable with respect to $\mathcal{Z}_1$ and so we can replace (5.5) by

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \mathbb{E}(f \mid \mathcal{Z}_1) \cdot T^n \mathbb{E}(f \mid \mathcal{Z}_1) \cdot T^{2n}\mathbb{E}(f \mid \mathcal{Z}_1) \, \mathrm{d}\mu.$$

This means that we can assume that the first term is also measurable with respect to the Kronecker factor, and so we can assume that f is a nonnegative function that is measurable with respect to the Kronecker. Thus the system X can be assumed to be Z_1 and the transformation T is rotation by some irrational α. Thus it suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int_{Z_1} f(s) \cdot f(s + n\alpha) \cdot f(s + 2n\alpha) \, \mathrm{d}m(s)$$

exists and is positive. But the convergence of this last expression is immediate using Fourier analysis. Since $\{n\alpha\}$ is equidistributed in Z_1, this limit approaches

$$\iint_{Z_1 \times Z_1} f(s) \cdot f(s+t) \cdot f(s+2t) \, \mathrm{d}m(s) \, \mathrm{d}m(t). \tag{5.6}$$

[3]In Furstenberg's proof of Szemerédi's theorem via Theorem 1.3, he showed that the analogous lim sup for $k \geq 2$ is positive and only showed the existence of the associated limit for $k = 2$. The positivity of the lim sup suffices for proving Szemerédi's theorem. As we are interested in the existence of the limit for $k > 2$ and the finer combinatorial information that can be gleaned from this, we prove the deeper statement here.

But

$$\lim_{t\to 0}\int_{Z_1} f(s)\cdot f(s+t)\cdot f(s+2t)\,\mathrm{d}m(s) = \int_{Z_1} f(s)^3\,\mathrm{d}m(s),$$

which is clearly positive. In particular, the double integral in (5.6) is positive. □

In the proof we have actually proven a stronger statement than needed to obtain Roth's theorem: we have shown the existence of the limit of the double average in $L^2(\mu)$. Letting $\tilde{f} = \mathbb{E}(f \mid \mathcal{Z}_1)$ for $f \in L^\infty(\mu)$, we have show that the double average (5.3) converges to

$$\int_{Z_1} \widetilde{f}_1(\pi_1(x)+s)\cdot \widetilde{f}_2(\pi_1(x)+2s)\,dm(s).$$

More generally, the same sort of argument can be used to show that in a weak mixing system, the Kronecker factor is characteristic for the averages (3.1) for all $k \geq 1$, meaning that to prove convergence of these averages in a weak mixing system it suffices to assume that the system is a Kronecker system. Using Fourier analysis, one then gets convergence of the averages (3.1) for weak mixing systems.

5.8. Multiple averages. We want to carry out similar analysis for the multiple averages

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \cdots \cdot T^{kn} f_k$$

and show the existence of the limit in $L^2(\mu)$ as $N \to \infty$. In his proof of Szemerédi's theorem in [16] and subsequent proofs of Szemerédi's theorem via ergodic theory such as [21], the approach of Section 5.7 is not the one used for $k \geq 3$. Namely, they do not show the existence of the limit and then analyze the limit itself to show it is positive. A weaker statement is proved, only giving that the lim inf of (2.1) is positive. We will not discuss the intricate structure theorem and induction needed to prove this.

Already to prove convergence for $k = 3$, one needs to consider more than just rotational behavior.

Example 5.11. Given a system $(X, \mathcal{X}, \mu, T)$, let $F(Tx) = f(x)F(x)$, where

$$f(Tx) = \lambda f(x) \quad \text{and} \quad |\lambda| = 1.$$

Then

$$F(T^n x) = f(x)f(Tx)\cdots f(T^{n-1}x)F(x) = \lambda^{n(n-1)/2}\big(f(x)\big)^n F(x)$$

and so

$$F(x) = \big(F(T^n x)\big)^3 \big(F(T^{2n}x)\big)^{-3} F(T^{3n}x).$$

This means that there is some relation among

$$(x, T^n x, T^{2n} x, T^{3n} x)$$

that does not arise from the Kronecker factor.

One can construct more complicated examples (see Furstenberg [18]) that show that even such generalized eigenfunctions do not suffice for determining the limiting behavior for $k = 3$. More precisely, the factor corresponding to generalized eigenfunctions (the *Abramov factor*) is not characteristic for the average (3.1) with $k = 3$.

To understand the triple average, one needs to take into account systems more complicated than Kronecker and Abramov systems. The simplest such example is a 2-step nilsystem (the use of this terminology will be clarified later):

Example 5.12. Let $X = \mathbb{T}\times\mathbb{T}$, with Borel σ-algebra, and Haar measure. Fix $\alpha \in \mathbb{T}$ and define $T\colon X \to X$ by

$$T(x,y) = (x+\alpha, y+x)$$

The system is ergodic if and only if $\alpha \notin \mathbb{Q}$.

The system is not isomorphic to a group rotation, as can be seen by defining $f(x,y) = e(y) = \exp(2\pi i y)$. Then for all $n \in \mathbb{Z}$,

$$T^n(x,y) = \left(x + n\alpha, y + nx + \frac{n(n-1)}{2}\alpha\right)$$

and so

$$f(T^n(x,y)) = e(y)e(nx)e\left(\frac{n(n-1)}{2}\alpha\right).$$

Quadratic expressions like these do not arise from a rotation on a group.

6. The structure theorem

6.1. Major steps in the proof of Theorem 3.1. In broad terms, there are four major steps in the proof of Theorem 3.1.

For each $k \in \mathbb{N}$, we inductively define a seminorm $|||\cdot|||_k$ that controls the asymptotic behavior of the average. More precisely, we show that if $|f_1| \le 1, \dots, |f_k| \le 1$, then

$$\limsup_{N\to\infty} \left\| \frac{1}{N}\sum_{n=o}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \cdots \cdot T^{kn} f_k \right\|_{L^2(\mu)} \le \min_{1\le j\le k} |||f_j|||_k. \tag{6.1}$$

Using these seminorms, we define factors Z_k of X such that for $f \in L^\infty(\mu)$,

$$\mathbb{E}(f \mid \mathcal{Z}_{k-1}) = 0 \text{ if and only if } |||f|||_k = 0.$$

It follows from (6.1) that the factor Z_{k-1} is characteristic for the average (3.1).

The bulk of the work is then to give a "geometric" description of these factors. This description is in terms of nilpotent groups, and more precisely we show that the dynamics of translations on homogeneous spaces of a nilpotent Lie group determines the limiting behavior of these averages. This is the content of the Structure Theorem, explained in Section 6.2. (A more detailed expository version of this is given in Host [31]; for full details, see [34].)

Finally, we show convergence for these particular types of systems.

Roughly speaking, this same outline applies to other convergence results we consider in the sequel, such as averages along polynomial times, averages along cubes, or averages for commuting transformations. For each average, we find a characteristic factor that can be described in geometric terms, allowing us to prove convergence in the characteristic factor.

6.2. The role of nilsystems. We have already seen that the limit behavior of the double average is controlled by group rotations, meaning the Kronecker factor is characteristic for this average. Furthermore, we have seen that something more is needed to control the limit behavior of the triple average. Our goal here is to explain how the multiple averages of (3.1), and some more general averages, are controlled by nilsystems. We start with some terminology.

Let G be a group. If $g, h \in G$, let $[g, h] = g^{-1}h^{-1}gh$ denote the commutator of g and h. If $A, B \subset G$, we write $[A, B]$ for the subgroup of G spanned by $\{[a, b] : a \in A, b \in B\}$. The lower central series

$$G = G_1 \supset G_2 \supset \cdots \supset G_j \supset G_{j+1} \supset \cdots$$

of G is defined inductively, by setting $G_1 = G$ and $G_{j+1} = [G, G_j]$ for $j \geq 1$. We say that G is *k-step nilpotent* if $G_{k+1} = \{1\}$.

If G is a k-step nilpotent Lie group and Γ is a discrete cocompact subgroup, the compact manifold $X = G/\Gamma$ is a *k-step nilmanifold.*

The group G acts naturally on X by left translation: if $a \in G$ and $x \in X$, the translation T_a by a is given by $T_a(x\Gamma) = (ax)\Gamma$. There is a unique Borel probability measure μ (the *Haar measure*) on X that is invariant under this action. We let $\mathcal{G}/\Gamma$ denote the associated Borel σ-algebra on G/Γ. Fixing an element $a \in G$, the system $(G/\Gamma, \mathcal{G}/\Gamma, T_a, \mu)$ is a *k-step nilsystem* and T_a is a *nilrotation.*

The system $(X, \mathcal{X}, \mu, T)$ is an *inverse limit* of a sequence of factors $\{(X_j, \mathcal{X}_j, \mu_j, T)\}$ if $\{\mathcal{X}_j\}_{j\in\mathbb{N}}$ is an increasing sequence of T-invariant sub-σ-algebras such that $\bigvee_{j\in\mathbb{N}} \mathcal{X}_j = \mathcal{X}$ up to null sets.[4] If each system $(X_j, \mathcal{X}_j, \mu_j, T)$ is isomorphic to a k-step nilsystem, then $(X, \mathcal{X}, \mu, T)$ is an *inverse limit of k-step nilsystems.*

Proving convergence of the averages (3.1) is only possible if one can has a good description of some characteristic factor for these averages. This is the content of the Structure Theorem:

Theorem 6.1 (Host and Kra [34]). *There exists a characteristic factor for the averages* (3.1) *which is isomorphic to an inverse limit of* $(k-1)$*-step nilsystems.*

The advantage of reducing to nilsystems is that convergence of the averages under study is much easier in nilsystems. This is further discussed in Section 8.2.

6.3. Examples of nilsystems. We give two examples of nilsystems that illustrate their general properties.

Example 6.2. Let $G = \mathbb{Z} \times \mathbb{T} \times \mathbb{T}$ with multiplication given by

$$(k, x, y) * (k', x', y') = \big(k + k', x + x' \pmod 1, y + y' + 2kx' \pmod 1\big).$$

The commutator subgroup of G is $\{0\} \times \{0\} \times \mathbb{T}$, and G is 2-step nilpotent. The subgroup $\Gamma = \mathbb{Z} \times \{0\} \times \{0\}$ is discrete and cocompact, and thus $X = G/\Gamma$ is a nilmanifold. Let $\mathcal{X}$ denote the Borel σ-algebra and let μ denote Haar measure on X. Fix some irrational $\alpha \in \mathbb{T}$, let $a = (1, \alpha, \alpha)$, and let $T : X \to X$ be translation by a. Then (X, μ, T) is a 2-step nilsystem.

The Kronecker factor of X is $\mathbb{T}$ with rotation by α. Identifying X with $\mathbb{T}^2$ via the map $(k, x, y) \mapsto (x, y)$, the transformation T takes on the familiar form of a

[4] Recall that if $\mathcal{X}_1$ and $\mathcal{X}_1$ are sub-σ-algebras of $\mathcal{X}$, then $\mathcal{X}_1 \bigvee \mathcal{X}_2$ denotes the smallest sub-σ-algebra of $\mathcal{X}$ containing both $\mathcal{X}_1$ and $\mathcal{X}_2$. Thus $\mathcal{X}_1 \bigvee \mathcal{X}_2$ consists of all sets which are unions of sets of the form the form $A \cap B$ for $A \in \mathcal{X}_1$ and $B \in \mathcal{X}_2$.

skew transformation:

$$T(x,y) = (x+\alpha, y+2x+\alpha).$$

This system is ergodic if and only if $\alpha \notin \mathbb{Q}$: for $x, y \in X$ and $n \in \mathbb{Z}$,

$$T^n(x,y) = (x+n\alpha, y+2nx+n^2\alpha)$$

and equidistribution of the sequence $\{T^n(x,y)\}$ is equivalent to ergodicity.

Example 6.3. Let G be the Heisenberg group $\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ with multiplication given by

$$(x,y,z) * (x',y',z') = (x+x', y+y', z+z'+xy').$$

Then G is a 2-step nilpotent Lie group. The subgroup $\Gamma = \mathbb{Z}\times\mathbb{Z}\times\mathbb{Z}$ is discrete and cocompact and so $X = G/\Gamma$ is a nilmanifold. Letting T be the translation by $a = (a_1, a_2, a_3) \in G$ where a_1, a_2 are independent over $\mathbb{Q}$ and $a_3 \in \mathbb{R}$, and taking $\mathcal{X}$ to be the Borel σ-algebra and μ to be the Haar measure, we have that $(X, \mathcal{X}, \mu, T)$ is a nilsystem. The system is ergodic if and only if a_1, a_2 are independent over $\mathbb{Q}$.

The compact abelian group $G/G_2\Gamma$ is isomorphic to $\mathbb{T}^2$ and the rotation on $\mathbb{T}^2$ by (a_1, a_2) is ergodic (again for a_1, a_2 independent over $\mathbb{Q}$). The Kronecker factor of X is the factor induced by functions on x_1, x_2. The system $(X, \mathcal{X}, \mu, T)$ is (uniquely) ergodic.

The dynamics of the first example gives rise to quadratic sequences, such as $\{n^2\alpha\}$, and the dynamics of the second example gives rise to generalized quadratic sequences such as $\{\lfloor n\alpha\rfloor n\beta\}$.

6.4. Motivation for nilpotent groups. The content of the Structure Theorem is that nilpotent groups, or more precisely the dynamics of a translation on the homogeneous space of a nilpotent Lie group, control the limiting behavior of the averages along arithmetic progressions. We give some motivation as to why nilpotent groups arise.

If G is an abelian group, then

$$\{(g, gz, gz^2, \ldots, gz^n) : g, z \in G\}$$

is a subgroup of G^n. However, this does not hold if G is not abelian. To make these arithmetic progressions into a group, one must take into account the commutators. This is the content of the following theorem, proven in different contexts by Hall [29], Petresco [48], Lazard [42], Leibman [43]. (Recall that $G_i = [G, G_i]$ denotes the ith entry in the lower central series of G.)

Theorem 6.4. *If G is a group, then for any $x, y \in G$, there exist $z \in G$ and $w_i \in G_i$ such that*

$$(x, x^2, x^3, \ldots, x^n) \times (y, y^2, y^3, \ldots, y^n) \\ = \Big(z, z^2 w_1, z^3 w_1^3 w_2, \ldots, z^{\binom{n}{1}} w_1^{\binom{n}{2}} w_2^{\binom{n}{3}} \ldots w_{n-1}^{\binom{n}{n}}\Big).$$

Furthermore, these expressions form a group.

If G is a group, a *geometric progression* is a sequence of the form

$$g, gz, gz^2 w_1, gz^3 w_1^3 w_2, \ldots, gz^{\binom{n}{1}} w_1^{\binom{n}{2}} \ldots w_{n-1}^{\binom{n}{n}}, \ldots$$

where $g, z \in G$ and $w_i \in G_i$.

Thus if G is abelian, g and z determine the whole sequence. On the other hand, if G is k-step nilpotent with $k < n$, the first k terms determine the whole sequence. (This holds because each w_i appears first in the ith term of the sequence and with exponent 1, and so it is completely determined, and for $i > k$, each w_i is trivial.)

Similarly, if $(G/\Gamma, \mathcal{G}/\Gamma, \mu, T_a)$ is a k-step nilsystem and

$$x_1 = g_1\Gamma, \quad x_2 = g_2\Gamma, \ldots, \quad x_k = g_k\Gamma, \ldots, \quad x_n = g_n\Gamma$$

is a geometric progression in G/Γ, then the first k terms determine the rest. Thus in a k-step nilsystem, $a^{k+1}x\Gamma$ is a function of the first k terms $ax\Gamma, a^2x\Gamma, \ldots, a^kx\Gamma$.

This means that the $(k+1)$st term $T^{(k+1)n}x$ in an arithmetic progression $T^nx, \ldots, T^{kn}x$ is constrained by the first k terms. More interestingly, the converse also holds: in an arbitrary system $(X, \mathcal{X}, \mu, T)$, any k-step nilpotent factor places a constraint on $(x, T^nx, T^{2n}x, \ldots, T^{kn}x)$.

7. Building characteristic factors

The material in this and the next section is based on [34] and the reader is referred to [34] for full proofs. To describe characteristic factors for the averages (3.1), for each $k \in \mathbb{N}$ we define a seminorm and use it to define these factors. We start by defining certain measures that are then used to define the seminorms. Throughout this section, we assume that $(X, \mathcal{X}, \mu, T)$ is an ergodic system.

7.1. Definition of the measures. Let $X^{[k]} = X^{2^k}$ and define $T^{[k]} \colon X^{[k]} \to X^{[k]}$ by $T^{[k]} = T \times \cdots \times T$ (taken 2^k times).

We write a point $\mathbf{x} \in X^{[k]}$ as $\mathbf{x} = \big(x_\epsilon : \epsilon \in \{0,1\}^k\big)$ and make the natural identification of $X^{[k+1]}$ with $X^{[k]} \times X^{[k]}$, writing $\mathbf{x} = (\mathbf{x}', \mathbf{x}'')$ for a point of $X^{[k+1]}$, with $\mathbf{x}', \mathbf{x}'' \in X^{[k]}$.

By induction, we define a measure $\mu^{[k]}$ on $X^{[k]}$ invariant under $T^{[k]}$. Set $\mu^{[0]} := \mu$. Let $\mathcal{I}^{[k]}$ be the invariant σ-algebra of $(X^{[k]}, \mathcal{X}^{[k]}, \mu^{[k]}, T^{[k]})$. (Note that this system is not necessarily ergodic.) Then $\mu^{[k+1]}$ is defined to be the *relatively independent joining* of $\mu^{[k]}$ with itself over $\mathcal{I}^{[k]}$, meaning that if F and G are bounded functions on $X^{[k]}$,

$$\text{(7.1)} \quad \int_{X^{[k+1]}} F(\mathbf{x}') \cdot G(\mathbf{x}'') \,\mathrm{d}\mu^{[k+1]}(\mathbf{x}) = \int_{X^{[k]}} \mathbb{E}(F \mid \mathcal{I}^{[k]})(\mathbf{y}) \cdot \mathbb{E}(G \mid \mathcal{I}^{[k]})(\mathbf{y}) \,\mathrm{d}\mu^{[k]}(\mathbf{y}).$$

Since $(X, \mathcal{X}, \mu, T)$ is assumed to be ergodic, $\mathcal{I}^{[0]}$ is trivial and $\mu^{[1]} = \mu \times \mu$. If the system is weak mixing, then for all $k \geq 1$, $\mu^{[k]}$ is the product measure $\mu \times \mu \times \cdots \times \mu$, taken 2^k times.

7.2. Symmetries of the measures. Writing a point $\mathbf{x} \in X^{[k]}$ as

$$\mathbf{x} = (x_\epsilon \colon \epsilon \in \{0,1\}^k),$$

we identify the indexing set $\{0,1\}^k$ of this point with the vertices of the Euclidean cube.

An isometry σ of $\{0,1\}^k$ induces a map $\sigma_* \colon X^{[k]} \to X^{[k]}$ by *permuting the coordinates*:

$$(\sigma_*(\mathbf{x}))_\epsilon = x_{\sigma(\epsilon)}.$$

For example, from the diagonal symmetries for $k=2$, we have the permutations

$$(x_{00},x_{01},x_{10},x_{11}) \mapsto (x_{00},x_{10},x_{01},x_{11})$$
$$(x_{00},x_{01},x_{10},x_{11}) \mapsto (x_{11},x_{01},x_{10},x_{00}).$$

By induction, the measures are invariant under permutations:

Lemma 7.1. *For each $k \in \mathbb{N}$, the measure $\mu^{[k]}$ is invariant under all permutations of coordinates arising from isometries of the unit Euclidean cube.*

7.3. Defining seminorms. For each $k \in \mathbb{N}$, we define a seminorm on $L^\infty(\mu)$ by setting

$$|||f|||_k^{2^k} = \int_{X^{[k]}} \prod_{\epsilon\in\{0,1\}^k} f(x_\epsilon)\,\mathrm{d}\mu^{[k]}(\mathbf{x}).$$

By definition of the measure $\mu^{[k]}$, this integral is equal to

$$\int_{X^{[k-1]}} \mathbb{E}\Big(\prod_{\epsilon\in\{0,1\}^{k-1}} f(x_\epsilon)\mid \mathcal{I}^{[k-1]}\Big)^2 \,\mathrm{d}\mu^{[k-1]}$$

and so in particular it is nonnegative.

From the symmetries of the measure $\mu^{[k]}$ (Lemma 7.1), we have a version of the Cauchy – Schwarz inequality for the seminorms, referred to as a Cauchy – Schwarz – Gowers inequality:

Lemma 7.2. *For $\epsilon \in \{0,1\}^k$, let $f_\epsilon \in L^\infty(\mu)$. Then*

$$\Big|\int \prod_{\epsilon\in\{0,1\}^k} f_\epsilon(x_\epsilon)\,\mathrm{d}\mu^{[k]}(\mathbf{x})\Big| \le \prod_{\epsilon\in\{0,1\}^k} |||f_\epsilon|||_k.$$

As a corollary, the map $f \mapsto |||f|||_k$ is subadditive (meaning that $|||f+g|||_k \le |||f|||_k + |||g|||_k$ for all $f,g\in L^\infty(\mu)$) and so:

Corollary 7.3. *For every $k\in\mathbb{N}$, $|||\cdot|||_k$ is a seminorm on $L^\infty(\mu)$.*

Since the system $(X,\mathcal{X},\mu,T)$ is ergodic, the σ-algebra $\mathcal{I}^{[0]}$ is trivial, $\mu^{[1]}=\mu\times\mu$ and $|||f|||_1 = \left|\int f\,\mathrm{d}\mu\right|$. By induction,

$$|||f|||_1 \le |||f|||_2 \le \cdots \le |||f|||_k \le \cdots \le \|f\|_\infty.$$

If the system is weak mixing, then $|||f|||_k = |||f|||_1$ for all $k\in\mathbb{N}$.

By induction and the ergodic theorem, we have a second presentation of these seminorms:

Lemma 7.4. *For every $k \ge 1$,*

$$|||f|||_{k+1}^{2^{k+1}} = \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} |||f\cdot T^n f|||_k^{2^k}.$$

7.4. Seminorms control the averages (3.1). The seminorms $|||\cdot|||_k$ control the averages along arithmetic progressions:

Lemma 7.5. *Assume that $(X,\mathcal{X},\mu,T)$ is ergodic and let $k\in\mathbb{N}$. If $\|f_1\|_\infty,\dots,$ $\|f_k\|_\infty \le 1$, then*

$$\limsup_{N\to\infty}\Big\|\frac{1}{N}\sum_{n=0}^{N-1} T^n f_1\cdot T^{2n}f_2\cdot\,\cdots\,\cdot T^{kn}f_k\Big\|_{L^2(\mu)} \le \min_{\ell=1,\dots,k} \ell|||f_\ell|||_k.$$

PROOF. We proceed by induction on k. For $k = 1$, by the ergodic theorem

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=0}^{N-1} T^n f_1\right\|_{L^2(\mu)} = \left|\int f_1 \,\mathrm{d}\mu\right| = |||f_1|||_1.$$

Assume it holds for $k \geq 1$. Let $f_1, f_2, \ldots, f_{k+1} \in L^\infty(\mu)$ with $\|f_j\|_\infty \leq 1$ for $j = 1, 2, \ldots, k+1$ and define $u_n = T^n f_1 \cdot T^{2n} f_2 \cdots T^{(k+1)n} f_{k+1}$. Assume that $\ell \in \{2, 3, \ldots, k+1\}$ (the case $\ell = 1$ is similar). Then

$$\left|\frac{1}{N}\sum_{n=0}^{N-1}\langle u_{n+h}, u_n\rangle\right| = \left|\int (f_1 \cdot T^h f_1)\frac{1}{N}\sum_{n=0}^{N-1}\prod_{j=2}^{k+1} T^{(j-1)n}(f_j \cdot T^{jh} f_j)\,\mathrm{d}\mu\right|$$
$$\leq \|f_1 \cdot T^h f_1\|_{L^2(\mu)}\left\|\frac{1}{N}\sum_{n=0}^{N-1}\prod_{j=2}^{k+1} T^{(j-1)n}(f_j \cdot T^{jh} f_j)\right\|_{L^2(\mu)}.$$

Set

$$\gamma_h = \limsup_{N\to\infty}\left|\frac{1}{N}\sum_{n=0}^{N-1}\langle u_{n+h}, u_n\rangle\right|.$$

Then by the inductive hypothesis, with f_{j-1} replaced by $f_j \cdot T^{jh} f_j$ for $j = 2, 3, \ldots, k+1$, we have that

$$\gamma_h \leq \ell \cdot |||f_\ell \cdot T^{\ell h} f_\ell|||_k.$$

Thus

$$\frac{1}{H}\sum_{h=0}^{H-1}\gamma_h \leq \ell^2\,\frac{1}{\ell H}\sum_{n=0}^{\ell H-1}|||f_\ell \cdot T^n f_\ell|||_k$$

and the statement follows from the van der Corput lemma (Lemma 5.9) and the definition of the seminorm $|||\cdot|||_{k+1}$. □

7.5. The Kronecker factor, revisited ($k = 2$). We have seen two presentations of the Kronecker factor $(Z_1, \mathcal{Z}_1, m, T)$: it is the largest abelian group rotation factor and it is the sub-σ-algebra of $\mathcal{X}$ that gives rise to all eigenfunctions. Another equivalent formulation is that it is the smallest sub-σ-algebra of $\mathcal{X}$ such that all invariant functions of $(X \times X, \mathcal{X} \times \mathcal{X}, \mu \times \mu, T \times T)$ are measurable with respect to $\mathcal{Z}_1 \times \mathcal{Z}_1$. Recall that $\pi_1 \colon X \to \mathcal{Z}_1$ denotes the factor map.

We give an explicit description of the measure $\mu^{[2]}$, and thus give yet another description of the Kronecker factor. For $f \in L^\infty(\mu)$, write $\tilde{f} = \mathbb{E}(f \mid \mathcal{Z}_1)$.

For $s \in Z_1$ and $f_0, f_1 \in L^\infty(\mu)$, we define a probability measure μ_s on $X \times X$ by

$$\int_{X\times X} f_0(x_0) f_1(x_1)\,\mathrm{d}\mu_s(x_0, x_1) := \int_{Z_1}\tilde{f}_0(z)\tilde{f}_1(z+s)\,\mathrm{d}m(z).$$

This measure is $T\times T$-invariant and the ergodic decomposition of $\mu\times\mu$ under $T\times T$ is given by

$$\mu\times\mu = \int_{Z_1}\mu_s\,\mathrm{d}m(s).$$

Thus for m-almost every $s \in Z_1$, the system $(X \times X, \mathcal{X} \times \mathcal{X}, \mu_s, T \times T)$ is ergodic and

$$\mu^{[2]} = \int_{Z_1}\mu_s \times \mu_s\,\mathrm{d}m(s).$$

More generally, if f_ϵ, $\epsilon \in \{0,1\}^2$, are measurable functions on X, then

$$\int_{X^{[2]}} f_{00} \otimes f_{01} \otimes f_{10} \otimes f_{11} \, \mathrm{d}\mu^{[2]} = \int_{Z_1^3} \tilde{f}_{00}(z) \cdot \tilde{f}_{01}(z+s) \cdot \tilde{f}_{10}(z+t) \cdot \tilde{f}_{11}(z+s+t) \, \mathrm{d}m(z) \, \mathrm{d}m(s) \, \mathrm{d}m(t).$$

It follows immediately that:

$$|||f|||_2^4 := \int f \otimes f \otimes f \otimes f \, \mathrm{d}\mu^{[2]} = \int_{Z_1^3} \tilde{f}(z) \cdot \tilde{f}(z+s) \cdot \tilde{f}(z+t) \cdot \tilde{f}(z+s+t) \, \mathrm{d}m(z) \, \mathrm{d}m(s) \, \mathrm{d}m(t).$$

As a corollary, $|||f|||_2$ is the ℓ^4-norm of the Fourier Transform of $\tilde{f}$ and the factor Z_1, defined by $|||f|||_2 = 0$ if and only if $\mathbb{E}(f \mid \mathcal{Z}_1) = 0$ for $f \in L^\infty(\mu)$, is the Kronecker factor of $(X, \mathcal{X}, \mu, T)$.

7.6. Factors for all $k \geq 1$. Using these seminorms, we define factors $Z_k = Z_k(X)$ for $k \geq 1$ of X that generalize the relation between the Kronecker factor Z_1 and the second seminorm $|||\cdot|||_2$. We define $\mathcal{Z}_k$ as follows: for $f \in L^\infty(\mu)$, $\mathbb{E}(f \mid \mathcal{Z}_k) = 0$ if and only if $|||f|||_{k+1} = 0$. We let Z_k denote the associated factor. That this does define a factor needs proof and to further explain this and the definition, we start by describing some geometric properties of the measures $\mu^{[k]}$.

Indexing $X^{[k]}$ by the coordinates $\{0,1\}^k$ of the Euclidean cube, it is natural to use geometric terms like *side, edge, vertex* for subsets of $\{0,1\}^k$. For example, Figure 1 illustrates the point $\mathbf{x} \in X^{[3]}$ with the side $\alpha = \{010, 011, 110, 111\}$:

Let $\alpha \subset \{0,1\}^k$ be a side. The *side transformation* $T_\alpha^{[k]}$ of $X^{[k]}$ is defined by:

$$\left(T_\alpha^{[k]}\mathbf{x}\right)_{\boldsymbol{\epsilon}} = \begin{cases} Tx_{\boldsymbol{\epsilon}} & \text{if } \boldsymbol{\epsilon} \in \alpha; \\ x_{\boldsymbol{\epsilon}} & \text{otherwise.} \end{cases}$$

We can represent the transformation T_α associated to the side $\{010, 011, 110, 111\}$ by Figure 2.

Since permutations of coordinates leave the measure $\mu^{[k]}$ invariant and act transitively on the sides, we have:

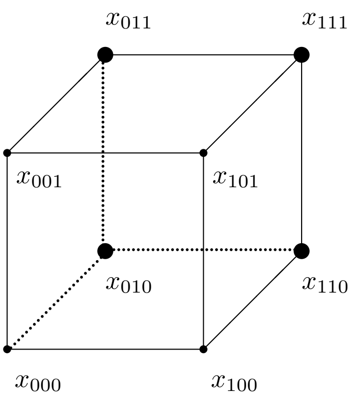


FIGURE 1.

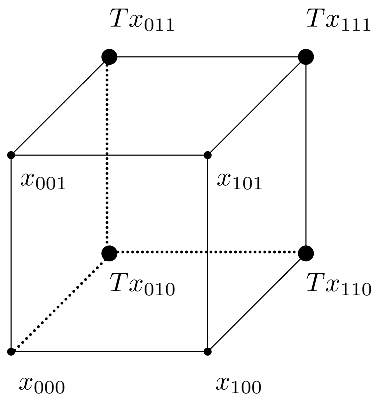


FIGURE 2.

Lemma 7.6. *For all $k \in \mathbb{N}$, the measure $\mu^{[k]}$ is invariant under the side transformations.*

We now view $X^{[k]}$ in a different way, identifying $X^{[k]} = X \times X^{2^k-1}$. A point $\mathbf{x} \in X^{[k]}$ is now written as

$$\mathbf{x} = (x_{\mathbf{0}}, \tilde{x}) \quad \text{where } \tilde{x} \in X^{2^k-1},\ x_{\mathbf{0}} \in X, \text{ and } \mathbf{0} = (00\dots0) \in \{0,1\}^k.$$

Although the 0 coordinate has been singled out and seems to play a particular role, it follows from the symmetries of the measure $\mu^{[k]}$ (Lemma 7.1) that any other coordinate could have been used instead.

If $\alpha \subset \{0,1\}^k$ is a side that does not contain $\mathbf{0}$ (there are k such sides), the transformation $T_\alpha^{[k]}$ leaves the coordinate $\mathbf{0}$ invariant. It follows from induction and the definition of the measure $\mu^{[k]}$ that:

Proposition 7.7. *Let $k \in \mathbb{N}$. If $B \subset X^{2^k-1}$, there exists $A \subset X$ with*

$$\mathbf{1}_A(x_{\mathbf{0}}) = \mathbf{1}_B(\tilde{x}) \quad \text{for almost all } \mathbf{x} = (x_{\mathbf{0}}, \tilde{x}) \in X^{[k]} \tag{7.2}$$

if and only if $X \times B$ is invariant under the k transformations $T_\alpha^{[k]}$ arising from the k sides α not containing $\mathbf{0}$.

This means that the subsets $A \subset X$ such that there exists $B \subset X^{2^k-1}$ satisfying (7.2) form an invariant sub-σ-algebra $\mathcal{Z}_{k-1} = \mathcal{Z}_{k-1}(X)$ of $\mathcal{X}$. We define $Z_{k-1} = Z_{k-1}(X)$ to be the associated factor. Thus $\mathcal{Z}_{k-1}(X)$ is defined to be the sub-σ-algebra of sets $A \subset X$ such that (7.2) holds for some set $B \subset X^{2^k-1}$.

We give some properties of the factors:

Proposition 7.8. (1) *For every bounded function f on X,*

$$|||f|||_k = 0 \text{ if and only if } \mathbb{E}(f \mid \mathcal{Z}_{k-1}) = 0.$$

(2) *For bounded functions f_ϵ, $\epsilon \in \{0,1\}^k$, on X,*

$$\int \prod_{\epsilon \in \{0,1\}^k} f_\epsilon(x_\epsilon)\, d\mu^{[k]}(\mathbf{x}) = \int \prod_{\epsilon \in \{0,1\}^k} \mathbb{E}(f_\epsilon \mid \mathcal{Z}_{k-1})(x_\epsilon)\, d\mu^{[k]}(\mathbf{x}).$$

Furthermore, $\mathcal{Z}_{k-1}$ is the smallest sub-σ-algebra of $\mathcal{X}$ with this property.

(3) *The invariant sets of* $(X^{[k]}, \mathcal{X}^{[k]}, \mu^{[k]}, T^{[k]})$ *are measurable with respect to* $\mathcal{Z}_k^{[k]}$. *Furthermore,* $\mathcal{Z}_k$ *is the smallest sub-*σ*-algebra of* $\mathcal{X}$ *with this property.*

The proof of this proposition relies on showing a similar formula to that used (in (7.1)) to define the measures $\mu^{[k]}$, but with respect to the new identification separating the 0 coordinate from the $2^k - 1$ others. Namely, for bounded functions f on X and F on X^{2^k-1},

$$\int_{X^{[k]}} f(x_{\mathbf{0}}) \cdot F(\tilde{x}) \, \mathrm{d}\mu^{[k]}(\mathbf{x}) = \int_{X^{[k-1]}} \mathbb{E}(f \mid \mathcal{Z}_{k-1}) \cdot \mathbb{E}(F \mid \mathcal{Z}_{k-1}) \, \mathrm{d}\mu^{[k-1]}.$$

The given properties then follow using induction and the symmetries of the measures.

We have already seen that Z_0 is the trivial factor and Z_1 is the Kronecker factor. More generally, the sequence of factors is increasing:

$$Z_0 \leftarrow Z_1 \leftarrow \cdots \leftarrow Z_k \leftarrow Z_{k+1} \leftarrow \cdots \leftarrow X.$$

If X is weak mixing, then $Z_k(X)$ is the trivial factor for every k.

An immediate consequence of Lemma 7.5 and the definition of the factors is that the factor Z_{k-1} is characteristic for the average along arithmetic progressions:

Proposition 7.9. *For all* $k \geq 1$, *the factor* Z_{k-1} *is characteristic for the convergence of the averages*

$$\frac{1}{N} \sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \cdots \cdot T^{kn} f_k.$$

This means that in order to understand the long term behavior of the multiple average along a k-term arithmetic progression, it suffices to assume that the space itself is Z_k. In particular, once we show that the factor Z_k has some useful structure (and this is the content of the Structure Theorem of [34], Theorem 8.1, discussed in Section 8), we are able to prove the existence of the limit of the average along arithmetic progressions. Proposition 7.9 would be meaningless if we were not able to explicitly describe the structure of Z_k in some way other than the abstract definition already given, and then use that description to prove convergence.

8. Structure theorem

8.1. Systems of order $\boldsymbol{k}$**.** For $k \geq 0$, an ergodic system X is said to be of *order* k if $Z_k(X) = X$. This means that $|||\cdot|||_{k+1}$ is a norm on $L^\infty(\mu)$.

Given an ergodic system $(X, \mathcal{X}, \mu, T)$, $Z_k(X)$ is a system of order k, since $Z_k(Z_k(X)) = Z_k(X)$. The unique system of order zero is the trivial system, and a system of order 1 is an ergodic rotation. By definition, if a system is of order k, then it is also of order k' for any $k' > k$.

By Proposition 7.9, to show convergence of

$$\frac{1}{N} \sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \cdots \cdot T^{kn} f_k$$

in an arbitrary system, it suffices to assume that each function is defined on the factor Z_{k-1}. But since $Z_{k-1}(X)$ is a system of order k, it suffices to prove convergence of this average for systems of order $k - 1$.

In this language, the Structure Theorem becomes:

Theorem 8.1 (Host and Kra [34]**).** *A system of order k is the inverse limit of a sequence of k-step nilsystems.*

Before turning to the proof of the Structure Theorem, we show convergence for the average along arithmetic progressions in a nilsystem. Combining this convergence with Theorem 8.1 completes the proof of Theorem 3.1.

8.2. Convergence on a nilmanifold. Using general properties of nilmanifolds (see Furstenberg [15] and Parry [47]), Lesigne [46] showed for connected group G and Leibman [44] showed in the general case, convergence in a nilsystem:

Theorem 8.2. *If $(X = G/\Gamma, \mathcal{G}/\Gamma, \mu, T)$ is a nilsystem and f is a continuous function on X, then*

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$$

converges for every $x \in X$.

(See also Ratner [50] and Shah [53] for related convergence results.)

As a corollary, we have convergence in $L^2(\mu)$ for the average along arithmetic progressions in a nilmanifold:

Corollary 8.3. *If $(X = G/\Gamma, \mathcal{G}/\Gamma, \mu, T)$ is a nilsystem, $k \in \mathbb{N}$, and $f_1, f_2, \dots, f_k \in L^\infty(\mu)$, then*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \dots \cdot T^{kn} f_k$$

exists in $L^2(\mu)$.

Proof. By density, we can assume that the functions are continuous. By assumption, G^k is a nilpotent Lie group, Γ^k is a discrete cocompact subgroup and $X^k = G^k/\Gamma^k$ is a nilmanifold. Let

$$s = (t, t^2, \dots, t^k) \in G^k$$

and let $S\colon X^k \to X^k$ be the translation by s, meaning that

$$S = T \times T^2 \times \dots \times T^k.$$

We apply Theorem 8.2 to (X^k, S) with the continuous function

$$F(x_1, x_2, \dots, x_k) = f_1(x_1) f_2(x_2) \dots f_k(x_k)$$

at the point $y = (x, x, \dots, x)$ and so the averages converge everywhere. □

Thus Theorem 3.1 holds in a nilsystem, and we are left with proving the Structure Theorem.

8.3. A group of transformations. To each ergodic system, we associate a group of measure preserving transformations. The general approach is to show that for sufficiently many systems of order k, this group is a nilpotent Lie group. The bulk of the work is to then show that this group acts transitively on the system. Thus the system can be given the structure of a nilmanifold and the Structure Theorem (Theorem 8.1) follows.

Most proofs are sketched or omitted completely, and the reader is referred to [34] for the details.

Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system. If $S\colon X \to X$ and $\alpha \subset \{0,1\}^k$, define $S_\alpha^{[k]}\colon X^{[k]} \to X^{[k]}$ by:

$$\big(S_\alpha^{[k]}\mathbf{x}\big)_{\boldsymbol{\epsilon}} = \begin{cases} Sx_{\boldsymbol{\epsilon}} & \text{if } \boldsymbol{\epsilon} \in \alpha; \\ x_{\boldsymbol{\epsilon}} & \text{otherwise.} \end{cases}$$

Let $\mathcal{G} = \mathcal{G}(X)$ be the group of transformations $S\colon X \to X$ such that for all $k \in \mathbb{N}$ and all sides $\alpha \subset \{0,1\}^k$, the measure $\mu^{[k]}$ is invariant under $S_\alpha^{[k]}$.

Some properties of this group are immediate. By symmetry, it suffices to consider one side. By definition, $T \in \mathcal{G}$, and if $ST = TS$ then we also have that $S \in \mathcal{G}$. If $S \in \mathcal{G}$ and $k \in N$, then $\mu^{[k]}$ is invariant under $S^{[k]}\colon X^{[k]} \to X^{[k]}$. Furthermore, $S^{[k]}E = E$ for every $E \in \mathcal{I}^{[k]}$.

By induction, the invariance of the measure $\mu^{[k]}$ under the side transformations, and commutator relations, we have:

Proposition 8.4. *If X is a system of order k, then $\mathcal{G}(X)$ is a k-step nilpotent group.*

8.4. Proof of the Structure Theorem. We proceed by induction. By the inductive assumption, we can assume that we are given a system $(X, \mathcal{X}, \mu, T)$ of order k. We have a factor $(Y, \mathcal{Y}, \nu, T)$, where $Y = Z_{k-1}(X)$ and $\pi\colon X \to Y$ is the factor map. Furthermore, Y is an inverse limit of a sequence of $(k-1)$-step nilsystems

$$Y = \varprojlim Y_i; \quad Y_i = G_i/\Gamma_i.$$

We want to show that X is an inverse limit of k-step nilsystems.

We have already shown that if $f_{\boldsymbol{\epsilon}}$, $\boldsymbol{\epsilon} \in \{0,1\}^k$, are bounded functions on X, then

$$\int \prod_{\boldsymbol{\epsilon}\in\{0,1\}^k} f_{\boldsymbol{\epsilon}}(x_{\boldsymbol{\epsilon}})\, \mathrm{d}\mu^{[k]}(\mathbf{x}) = \int \prod_{\boldsymbol{\epsilon}\in\{0,1\}^k} \mathbb{E}\big(f_{\boldsymbol{\epsilon}} \mid \mathcal{Y}\big)(x_{\boldsymbol{\epsilon}})\, \mathrm{d}\mu^{[k]}(\mathbf{x}).$$

In particular, for $f \in L^\infty(\mu)$,

$$|||f|||_k = 0 \text{ if and only if } \mathbb{E}(f \mid \mathcal{Y}) = 0.$$

Furthermore, X does not admit a strict sub-σ-algebra $\mathcal{Z}$ such that all invariant sets of $(X^{[k]}, \mu^{[k]}, T^{[k]})$ are measurable with respect to $\mathcal{Z}^{[k]}$. Recall also that the system $(X^{[k]}, \mu^{[k]}, T^{[k]})$ is defined as a relatively independent joining.

In [16], Furstenberg described the invariant σ-algebra for an arbitrary relatively independent joining. It follows that X is an *isometric extension* of Y, meaning that $X = Y \times H/K$ where H is a compact group and K is a closed subgroup, $\mu = \nu \times m$, where m is the Haar measure of H/K, and the transformation T is given by

$$T(y, u) = (Ty, \rho(y) \cdot u)$$

for some map $\rho\colon Y \to H$. (Note that we are making a slight, but standard, abuse of notation in using the same letter T to denote both the transformation in X and Y.)

Lemma 8.5. *For every $h \in H$, the transformation $(y, u) \mapsto (y, h \cdot u)$ of X belongs to the center of $\mathcal{G}(X)$.*

Thus H is abelian. We can substitute H/K for H, and we use additive notation for H.

We therefore have more information: X is an *abelian extension* of Y, meaning that $X = Y \times H$ for some compact abelian group H, $\mu = \nu \times m$, where m is the

Haar measure of H, and the transformation T is given by $T(y,u) = (Ty, u + \rho(y))$ for some map $\rho\colon Y \to H$. We call ρ the *cocycle* defining the extension.

Furthermore, we show that the cocycle defining this extension has a particular form, given by a particular functional equation:

Proposition 8.6. *If $(X, \mathcal{X}, \mu, T)$ is a system of order k and $(Y, \mathcal{Y}, \nu, T) = Z_{k-1}(X)$, then X is an abelian extension of Y via a compact group H and for the cocycle ρ defining this extension, there exists a map $\Phi\colon Y^{[k]} \to H$ such that*

$$\sum_{\boldsymbol{\epsilon}\in\{0,1\}^k} (-1)^{\epsilon_1+\cdots+\epsilon_k} \rho(y_{\boldsymbol{\epsilon}}) = \Phi(T^{[k]}\mathbf{y}) - \Phi(\mathbf{y}) \tag{8.1}$$

for $\nu^{[k]}$-a.e. $\mathbf{y} \in Y^{[k]}$.

We can make a few more assumptions on our system. Namely, by induction we can deduce that H is connected. Since every connected compact abelian group H is an inverse limit of a sequence of tori, we can further reduce to the case that $H = \mathbb{T}^d$.

8.5. The case $k = 2$ (The Conze – Lesigne equation). We maintain notation of the preceding section and review what this means for the case $k = 2$. By assumption, we have that $(Y, \mathcal{Y}, \nu, T)$ is a system of order 1, meaning it is a group rotation. The measure $\nu^{[2]}$ is the Haar measure of the subgroup

$$\big\{(y, y+s, y+t, y+s+t) : y, s, t \in Y\big\}$$

of Y^4. The functional equation of Proposition 8.6 is: there exists $\Phi\colon Y^3 \to \mathbb{T}^d$ with

$$\rho(y) - \rho(y+s) - \rho(y+t) + \rho(y+s+t) = \Phi(y+1, s, t) - \Phi(y, s, t)$$

It follows that for every $s \in Y$, there exists $\phi_s\colon Y \to \mathbb{T}^d$ and $c_s \in \mathbb{T}^d$ satisfying the *Conze – Lesigne equation* (see [9]):

$$\rho(y) - \rho(y+s) = \phi_s(y+1) - \phi_s(y) + c_s. \tag{CL}$$

The group $\mathcal{G}(X)$ associated to the system is the group of transformations of $X = Y \times \mathbb{T}^d$ of the form

$$(y, h) \mapsto (y+s, h+\phi_s(y))$$

where s and ϕ_s satisfy (CL).

8.6. Structure theorem in general. We give a short outline of the steps needed to complete the proof of the Structure Theorem for $k \geq 3$. We have that $Y = Z_{k-1}(X)$ is a system of order $k-1$, $X = Y \times \mathbb{T}^d$, $T(y,h) = (Ty, h+\rho(y))$, and $\rho\colon Y \to \mathbb{T}^d$ satisfies the functional equation (8.1). By the induction hypothesis $Y = \varprojlim Y_i$ where each $Y_i = G_i/\Gamma_i$ is a $(k-1)$-step nilsystem.

We first show that the cocycle ρ is *cohomologous* to a cocycle measurable with respect to $\mathcal{Y}_i$ for some i, meaning that the difference between the two cocycles is a coboundary. This reduces us to the case that ρ is measurable with respect to some $\mathcal{Y}_i$, and so we can assume that $Y = \mathcal{Y}_i$ for some i. Thus Y is a $(k-1)$-step nilsystem and we can assume that $Y = G/\Gamma$ with $G = \mathcal{G}(Y)$.

We then use the functional equation (8.1) to lift every transformation $S \in G$ to a transformation of X belonging to $\mathcal{G}(X)$. Starting with the case $S \in G_{k-1}$, we move up the lower central series of G. Lastly we show that we obtain sufficiently many elements of the group $\mathcal{G}(X)$ in this way.

8.7. Relations to the finite case. The seminorms $|||\cdot|||_k$ play the same role that the Gowers norms play in Gowers's proof [23] of Szemerédi's theorem and in Green and Tao's proof [25] that the primes contain arbitrarily long arithmetic progressions. We let U_k denote the k-th Gowers norm. For the finite system $\mathbb{Z}/N\mathbb{Z}$, $|||f|||_k = \|f\|_{U_k}$. Furthermore, $\|\cdot\|_{U_k}$ is a norm, not only a seminorm. The analog of Lemma 7.5 is that if $\|f_0\|_\infty, \|f_1\|_\infty, \ldots, \|f_k\|_\infty \leq 1$, then there exists some constant $C_k > 0$ such that

$$\left|\mathbb{E}\big(f_0(x)f_1(x+y)\ldots f_k(x+ky) \mid x, y \in \mathbb{Z}/p\mathbb{Z}\big)\right| \leq C_k \min_{0\leq j\leq k} \|f_j\|_{U_k}.$$

Other parts of the program are not as easy to translate to the finite setting. Consider defining a factor of the system using the seminorms. If p is prime, then $\mathbb{Z}/p\mathbb{Z}$ has no nontrivial factor and so there is no factor of $\mathbb{Z}/p\mathbb{Z}$ playing the role of the factor Z_k, meaning there is no factor with

$$\mathbb{E}(f \mid \mathcal{Z}_k) = 0 \text{ if and only if } \|f\|_{U_k} = 0.$$

Instead, the corresponding results have a different flavor: if $\|f\|_{U_k}$ is large in some sense, then f has large conditional expectation on some (noninvariant) σ-algebra or it has large correlation with a function of some particular class. Although we have a complete characterization of the seminorms $|||\cdot|||_k$ (and so also of the factors Z_k) in terms of nilmanifolds, there are only partial combinatorial characterizations in this direction (see [26–28]).

9. Other patterns

9.1. Commuting transformations. Ergodic theory has been used to detect other patterns that occur in sets of positive upper density, using Furstenberg's correspondence principle and an appropriately chosen strengthening of Furstenberg multiple recurrence. A first example is for commuting transformations:

Theorem 9.1 (Furstenberg and Katznelson [19]**).** *Let $(X, \mathcal{X}, \mu)$ be a probability measure space, let $k \geq 1$ be an integer, and assume that $T_j \colon X \to X$ are commuting measure preserving transformations for $j = 1, 2, \ldots, k$. Then for all $A \in \mathcal{X}$ with $\mu(A) > 0$, there exist infinitely many $n \in \mathbb{N}$ such that*

$$\mu(A \cap T_1^{-n}A \cap T_2^{-n}A \cap \cdots \cap T_k^{-n}A) > 0. \tag{9.1}$$

(In [20], Furstenberg and Katznelson proved a strengthening of this result, showing that one can place some restrictions on the choice of n; we do not discuss these "IP" versions of this theorem or the theorems given in the sequel.) Via correspondence, a multidimensional version of Szemerédi's theorem follows: if $E \subset \mathbb{Z}^r$ has positive upper density and $F \subset \mathbb{Z}^r$ is a finite subset, then there exist $z \in \mathbb{Z}^r$ and $n \in \mathbb{N}$ such that $z + nF \subset E$.

Again, this theorem is proven by showing that the associated lim inf of the average of the quantity in Equation (9.1) is positive. And again, it is natural to ask whether the limit

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T_1^{-n}A \cap \cdots \cap T_k^{-n}A)$$

exists in $L^2(\mu)$ for commuting maps $T_1, \ldots, T_k$. Only partial results are known. For $k = 2$, Conze and Lesigne ([8, 9]) proved convergence. For $k \geq 3$, the only known results rely on strong hypotheses of ergodicity:

Theorem 9.2 (Frantzikinakis and Kra [13]**).** *Let $k \in \mathbb{N}$ and assume that $T_1, T_2, \ldots, T_k$ are commuting invertible ergodic measure preserving transformations of a measure space $(X, \mathcal{X}, \mu)$ such that $T_iT_j^{-1}$ is ergodic for all $i, j \in \{1, 2, \ldots, k\}$ with $i \neq j$. If $f_1, f_2, \ldots, f_k \in L^\infty(\mu)$ the averages,*

$$\frac{1}{N}\sum_{n=0}^{N-1} T_1^n f_1 \cdot T_2^n f_2 \cdot \cdots \cdot T_k^n f_k$$

converge in $L^2(\mu)$ as $N \to \infty$.

The idea is to prove an analog of Lemma 7.5 for commuting transformations, thus reducing the problem to working in a nilsystem. The factors Z_k that are characteristic for averages along arithmetic progressions are also characteristic for these particular averages of commuting transformations. Without the strong hypotheses of ergodicity, this no longer holds and the general case remains open.

9.2. Averages along cubes. Another type of average is along k-dimensional cubes, the natural objects that arise in the definition of the seminorms. For example, a 2-dimensional cube is an expression of the form:

$$f(x)f(T^m x)f(T^n x)f(T^{m+n}x).$$

In [4], Bergelson showed the existence in $L^2(\mu)$ of

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{n,m=0}^{N-1} T^n f_1 \cdot T^m f_2 \cdot T^{n+m} f_3,$$

where $f_1, f_2, f_3 \in L^\infty(\mu)$. Similarly, one can define a 3-dimensional cube:

$$f_1(T^m x)f_2(T^n x)f_3(T^{m+n}x)f_4(T^p x)f_5(T^{m+p}x)f_6(T^{n+p}x)f_7(T^{m+n+p}x)$$

and existence of the limit of the average of this expression $L^2(\mu)$ for bounded functions $f_1, f_2, \ldots, f_7$ was shown in [33].

More generally, this theorem holds for cubes of $2^k - 1$ functions. Recalling the notation of Section 7, we have for $\epsilon = \epsilon_1 \ldots \epsilon_k \in \{0,1\}^k$ and $\mathbf{n} = (n_1, \ldots, n_k) \in \mathbb{Z}^k$,

$$\epsilon \cdot \mathbf{n} = \epsilon_1 n_1 + \epsilon_2 n_2 + \cdots + \epsilon_k n_k,$$

and $\mathbf{0}$ denotes the element $00\ldots0$ of $\{0,1\}^k$. We have:

Theorem 9.3 (Host and Kra [34]**).** *Let $(X, \mathcal{X}, \mu, T)$ be a system, let $k \geq 1$ be an integer, and let f_ϵ, $\epsilon \in \{0,1\}^k \setminus \{\mathbf{0}\}$, be $2^k - 1$ bounded functions on X. Then the averages*

$$\frac{1}{N^k} \cdot \sum_{\mathbf{n}\in[0,N-1]^k} \prod_{\substack{\epsilon\in\{0,1\}^k\\ \epsilon\neq\mathbf{0}}} T^{\epsilon\cdot\mathbf{n}} f_\epsilon$$

converge in $L^2(\mu)$ as $N \to \infty$.

The same result holds for translated averages, meaning the average for $\mathbf{n} \in [M_1, N_1] \times \cdots \times [M_k, N_k]$, as $N_1 - M_1, \ldots, N_k - M_k \to \infty$.

By Furstenberg's correspondence principle, this translates to a combinatorial statement. A subset $E \subset \mathbb{Z}$ is *syndetic* if $\mathbb{Z}$ can be covered by finitely many translates of E. In other words, there exists $N > 0$ such that every interval of size N contains at least one element of E. (Thus it is natural to refer to a syndetic set in the integers as a set with *bounded gaps.*) More generally, $E \subset \mathbb{Z}^k$ is *syndetic* if there exists an integer $N > 0$ such that

$$E \cap \big([M_1, M_1 + N] \times \cdots \times [M_k, M_k + N]\big) \neq \varnothing$$

for all $M_1, \dots, M_k \in \mathbb{Z}$.

Restricting Theorem 9.3 to indicator f unctions, the limit of the averages

$$\prod_{i=1}^{k} \frac{1}{N_i - M_i} \cdot \sum_{n_1 \in [M_1, N_1], \dots, n_k \in [M_k, N_k]} \mu\Big(\bigcap_{\epsilon \in \{0,1\}^k} T^{\epsilon \cdot \mathbf{n}} A\Big)$$

exists and is greater than or equal to $\mu(A)^{2^k}$ when $N_1 - M_1, \dots, N_k - M_k \to \infty$. Thus for every $\varepsilon > 0$,

$$\Big\{\mathbf{n} \in \mathbb{Z}^k : \mu\Big(\bigcap_{\epsilon \in \{0,1\}^k} T^{\epsilon \cdot \mathbf{n}} A\Big) > \mu(A)^{2^k} - \varepsilon\Big\}$$

of $\mathbb{Z}^k$ is syndetic.

By the correspondence principle, we have that if $E \subset \mathbb{Z}^k$ has upper density $d^*(E) > \delta > 0$ and $k \in \mathbb{N}$, then

$$\Big\{\mathbf{n} \in \mathbb{Z}^k \colon d^*\Big(\bigcap_{\epsilon \in \{0,1\}^k} (E + \epsilon \cdot \mathbf{n})\Big) \geq \delta^{2^k}\Big\}$$

is syndetic.

9.3. Polynomial patterns. In a different direction, one can restrict the iterates arising in Furstenberg's multiple recurrence. A natural choice is polynomial iterates, and the corresponding combinatorial statement is that a set of integers with positive upper density contains elements who differ by a polynomial:

Theorem 9.4 (Sárközy [51], Furstenberg [17]). *If $E \subset \mathbb{N}$ has positive upper density and $p\colon \mathbb{Z} \to \mathbb{Z}$ is a polynomial with $p(0) = 0$, then there exist $x, y \in E$ and $n \in \mathbb{N}$ such that $x - y = p(n)$.*

As for arithmetic progressions, Furstenberg's proof relies on the correspondence principle and an averaging theorem:

Theorem 9.5 (Furstenberg [17]). *Let $(X, \mathcal{X}, \mu, T)$ be a system, let $A \in \mathcal{X}$ with $\mu(A) > 0$ and let $p\colon \mathbb{Z} \to \mathbb{Z}$ be a polynomial with $p(0) = 0$. Then*

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-p(n)} A) > 0.$$

The multiple polynomial recurrence theorem, simultaneously generalizing this single polynomial result and Furstenberg's multiple recurrence, was proven by Bergelson and Leibman:

Theorem 9.6 (Bergelson and Leibman [6]). *Let $(X, \mathcal{X}, \mu, T)$ be a system, let $A \in \mathcal{X}$ with $\mu(A) > 0$, and let $k \in \mathbb{N}$. If $p_1, p_2, \dots, p_k\colon \mathbb{Z} \to \mathbb{Z}$ are polynomials with $p_j(0) = 0$ for $j = 1, \dots, k$, then*

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\big(A \cap T^{-p_1(n)} A \cap \cdots \cap T^{-p_k(n)} A\big) > 0. \tag{9.2}$$

By the correspondence principle, one immediately deduces a polynomial Szemerédi theorem: if $E \subset \mathbb{Z}$ has positive upper density, then it contains arbitrary polynomial patterns, meaning there exists $n \in \mathbb{N}$ such that

$$x, x+p_1(n), x+p_2(n), \dots, x+p_k(n) \in E.$$

(More generally, Bergelson and Leibman proved a version of Theorem 9.6 for commuting transformations, with a multidimensional polynomial Szemerédi theorem as a corollary.)

Again, it is natural to ask whether the lim inf in (9.2) is actually a limit. A first result in this direction was given by Furstenberg and Weiss [22], who proved convergence in $L^2(\mu)$ of

$$\frac{1}{N}\sum_{n=0}^{N-1} T^{n^2} f_1 \cdot T^n f_2$$

and

$$\frac{1}{N}\sum_{n=0}^{N-1} T^{n^2} f_1 \cdot T^{n^2+n} f_2$$

for bounded functions f_1, f_2.

The proof of convergence for general polynomial averages uses the technology of the seminorms, reducing to the same characteristic factors Z_k that can be described using nilsystems, as for averages along arithmetic progressions:

Theorem 9.7 (Host and Kra [35]**, Leibman** [45]**).** *Let $(X, \mathcal{X}, \mu, T)$ be a system, $k \in \mathbb{N}$, and $f_1, f_2, \dots, f_k \in L^\infty(\mu)$. Then for any polynomials $p_1, p_2, \dots, p_k \colon \mathbb{Z} \to \mathbb{Z}$, the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} T^{p_1(n)} f_1 \cdot T^{p_2(n)} f_2 \cdot \cdots \cdot T^{p_k(n)} f_k$$

converge in $L^2(\mu)$.

Recently, Johnson [36] has shown that under strong ergodicity conditions, similar to those in Theorem 9.2, one can generalize this and prove $L^2(\mu)$-convergence of the polynomial averages for commuting transformations:

$$\frac{1}{N}\sum_{n=0}^{N-1} T_1^{p_1(n)} f_1 \cdot T_2^{p_2(n)} f_2 \cdot \cdots \cdot T_k^{p_k(n)} f_k$$

for $f_1, f_2, \dots, f_k \in L^\infty(\mu)$.

For a *totally ergodic system* (meaning that T^n is ergodic for all $n \in \mathbb{N}$), Furstenberg and Weiss showed a stronger result, giving an explicit and simple formula for the limit:

$$\frac{1}{N}\sum_{n=0}^{N-1} T^n f_1 \cdot T^{n^2} f_2 \to \int f_1 \, \mathrm{d}\mu \cdot \int f_2 \, \mathrm{d}\mu$$

in $L^2(\mu)$.

Bergelson [3] asked whether the same result holds for k polynomials of different degrees, meaning that the limit of the polynomial average for a totally ergodic system is the product integrals. In [12], we show that the answer is yes under a more general condition. A family of polynomials $p_1, p_2, \dots, p_k \colon \mathbb{Z} \to \mathbb{Z}$ is *rationally independent* if the polynomials $\{1, p_1, p_2, \dots, p_k\}$ are linearly independent over the rationals. We show:

Theorem 9.8 (Frantzikinakis and Kra [12]**).** *Let $(X, \mathcal{X}, \mu, T)$ be a totally ergodic system, let $k \geq 1$ be an integer, and assume that $p_1, p_2 \ldots, p_k \colon \mathbb{Z} \to \mathbb{Z}$ are rationally independent polynomials. If $f_1, f_2, \ldots, f_k \in L^\infty(\mu)$,*

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^{p_1(n)} f_1 \cdot T^{p_2(n)} f_2 \cdot \cdots \cdot T^{p_k(n)} f_k - \prod_{i=1}^{k} \int f_i \, \mathrm{d}\mu \right\|_{L^2(\mu)} = 0.$$

As a corollary, if $(X, \mathcal{X}, \mu, T)$ is totally ergodic, $\{1, p_1, \ldots, p_k\}$ are rationally independent polynomials taking on integer values on the integers, and $A_0, A_1, \ldots, A_k \in \mathcal{X}$ with $\mu(A_i) > 0$, $i = 0, \ldots, k$, then

$$\mu(A_0 \cap T^{-p_1(n)} A_1 \cap \cdots \cap T^{-p_k(n)} A_k) > 0$$

for some $n \in \mathbb{N}$. Thus in a totally ergodic system, one can strengthen Bergelson and Leibman's multiple polynomial recurrence theorem, allowing the sets A_i to be distinct, and allowing the polynomials p_i to have nonzero constant term. It is not clear whether or not this has a combinatorial interpretation.

10. Strengthening Poincaré recurrence

10.1. Khintchine recurrence. Poincaré recurrence states that a set of positive measure returns to intersect itself infinitely often. One way to strengthen this is to ask that the set return to itself often with 'large' intersection. Khintchine made this notion precise, showing that large self intersection occurs on a syndetic set:

Theorem 10.1 (Khintchine [37]**).** *Let $(X, \mathcal{X}, \mu, T)$ be a system, let $A \in \mathcal{X}$ have $\mu(A) > 0$, and let $\varepsilon > 0$. Then*

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A) > \mu(A)^2 - \varepsilon\}$$

is syndetic.

It is natural to ask for a simultaneous generalization of Furstenberg multiple recurrence and Khintchine recurrence. More precisely, if $(X, \mathcal{X}, \mu, T)$ is a system, $A \in \mathcal{X}$ has positive measure, $k \in \mathbb{N}$, and $\varepsilon > 0$, is the set

$$\left\{n \in \mathbb{Z} \colon \mu\big(A \cap T^n A \cap \cdots \cap T^{kn} A\big) > \mu(A)^{k+1} - \varepsilon\right\}$$

syndetic?

Furstenberg multiple recurrence implies that there exists some constant $c = c(\mu(A)) > 0$ such that

$$\{n \in \mathbb{Z} \colon \mu(A \cap T^n A \cap \cdots \cap T^{kn} A) > c\}$$

is syndetic. But to generalize Khintchine recurrence, one needs $c = \mu(A)^{k+1}$. It turns out that the answer depends on the length k of the arithmetic progression.

Theorem 10.2 (Bergelson, Host and Kra [5]**).** *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system and let $A \in \mathcal{X}$. Then for every $\varepsilon > 0$, the sets*

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A \cap T^{2n} A) > \mu(A)^3 - \varepsilon\}$$

and

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A) > \mu(A)^4 - \varepsilon\}$$

are syndetic.

Furthermore, this result fails on average, meaning that the average of the left hand side expressions is not necessarily greater than $\mu(A)^3 - \varepsilon$ or $\mu(A)^4 - \varepsilon$, respectively.

On the other hand, based on an example of Ruzsa contained in the appendix of [5], we have:

Theorem 10.3 (Bergelson, Host and Kra [5]**).** *There exists an ergodic system $(X, \mathcal{X}, \mu, T)$ and for all $\ell \in \mathbb{N}$ there exists a set $A = A(\ell) \in \mathcal{X}$ with $\mu(A) > 0$ such that*

$$\mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A \cap T^{4n} A) \leq \mu(A)^\ell/2$$

for every integer $n \neq 0$.

We now briefly outline the major ingredients in the proofs of these theorems.

10.2. Positive ergodic results. We start with the ergodic results needed to prove Theorem 10.2. Fix an integer $k \geq 1$, an ergodic system $(X, \mathcal{X}, \mu, T)$, and $A \in \mathcal{X}$ with $\mu(A) > 0$. The key ingredient is the study of the *multicorrelation sequence*

$$\mu\big(A \cap T^n A \cap T^{2n} A \cap \cdots \cap T^{kn} A\big).$$

More generally, for a real valued function $f \in L^\infty(\mu)$, we consider the *multicorrelation sequence*

$$I_f(k, n) := \int f \cdot T^n f \cdot T^{2n} f \cdot \cdots \cdot T^{kn} f \, \mathrm{d}\mu(x).$$

When $k = 1$, Herglotz's theorem implies that the correlation sequence $I_f(1, n)$ is the Fourier transform of some positive measure $\sigma = \sigma_f$ on the torus $\mathbb{T}$:

$$I_f(1, n) = \widehat{\sigma}(n) := \int_{\mathbb{T}} \mathrm{e}^{2\pi i n t} \, \mathrm{d}\sigma(t).$$

Decomposing the measure σ into its continuous part σ^c and its discrete part σ^d, can write the multicorrelation sequence $I_f(1, n)$ as the sum of two sequences

$$I_f(1, n) = \widehat{\sigma^c}(n) + \widehat{\sigma^d}(n).$$

The sequence $\{\widehat{\sigma^c}(n)\}$ *tends to* 0 *in density*, meaning that

$$\lim_{N \to \infty} \sup_{M \in \mathbb{Z}} \frac{1}{M} \sum_{n=M}^{M+N-1} |\widehat{\sigma^c(n)}| = 0. \tag{10.1}$$

Equivalently, for any $\varepsilon > 0$, the upper Banach density[5] of the set $\{n \in \mathbb{Z} : |\widehat{\sigma^c(n)}| > \varepsilon\}$ is zero. The sequence $\{\widehat{\sigma^d}(n)\}$ is *almost periodic*, meaning that there exists a compact abelian group G, a continuous real valued function ϕ on G, and $a \in G$ such that $\widehat{\sigma^d}(n) = \phi(a^n)$ for all n.

A compact abelian group can be approximated by a compact abelian Lie group. Thus any almost periodic sequence can be uniformly approximated by an almost periodic sequence arising from a compact abelian Lie group.

In general, however, for higher k the answer is more complicated. We find a similar decomposition for the multicorrelation sequences $I_f(k, n)$ for $k \geq 2$. The notion of an almost periodic sequence is replaced by that of a *nilsequence*: for an

[5]The *upper Banach density* $\bar{d}(E)$ of a set $E \subset \mathbb{Z}$ is defined by $\bar{d}(e) = \lim_{N \to \infty} \sum_{M \in \mathbb{Z}} (1/N) |E \cap [M, M+N-1]|$.

integer $k \geq 2$, a k-step nilmanifold $X = G/\Gamma$, a continuous real (or complex) valued function ϕ on G, $a \in G$, and $e \in X$, the sequence $\{\phi(a^n \cdot e)\}$ is called a *basic k-step nilsequence.* A *k-step nilsequence* is a uniform limit of basic k-step nilsequences.

It follows that a 1-step nilsequence is the same as an almost periodic sequence. An inverse limit of compact abelian Lie groups is a compact group. However an inverse limit of k-step nilmanifolds is not, in general, the homogeneous space of some locally compact group, and so for higher k, the decomposition result must take into account the uniform limits of basic nilsequences. We have:

Theorem 10.4 (Bergelson, Host and Kra [5]**).** *Let $(X, \mathcal{X}, \mu, T)$ be an ergodic system, $f \in L^\infty(\mu)$ and $k \geq 1$ an integer. The sequence $\{I_f(k,n)\}$ is the sum of a sequence tending to zero in density and a k-step nilsequence.*

Due to the connections between the use of the seminorms in ergodic theory and the Gowers uniformity norms in additive combinatorics, it is natural that nilsequences also have a role to play on the combinatorial side. Recently, Green and Tao (see [26–28]) have adapted the idea of a nilsequence to combinatorics, and this plays a role in the asymptotics for the number of arithmetic progressions of length 4 in the primes. Ben Green's notes in this volume have more on this connection.

Finally, we explain how Theorem 10.4 can be used to prove Theorem 10.2. Let $\{a_n\}_{n\in\mathbb{Z}}$ be a bounded sequence of real numbers. The *syndetic supremum* of this sequence is defined to be

$$\sup\{c \in \mathbb{R} : \{n \in \mathbb{Z} : a_n > c\} \text{ is syndetic}\}.$$

Every nilsequence $\{a_n\}$ is uniformly recurrent.[6] In particular, if $S = \sup(a_n)$ and $\varepsilon > 0$, then $\{n \in \mathbb{Z}\colon a_n \geq S - \varepsilon\}$ is syndetic.

If $\{a_n\}$ and $\{b_n\}$ are two sequences of real numbers such that $a_n - b_n$ tends to 0 in density (in the sense of definition (10.1)), then the two sequences have the same syndetic supremum. Therefore the syndetic supremums of the sequences

$$\{\mu(A \cap T^n A \cap T^{2n} A)\}$$

and

$$\{\mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A)\}$$

are equal to the supremum of the associated nilsequences, and we are reduced to showing that they are greater than or equal to $\mu(A)^3$ and $\mu(A)^4$, respectively.

10.3. Nonergodic counterexample. Ergodicity is not needed for Khintchine's theorem, but is essential for Theorem 10.2:

Theorem 10.5 (Bergelson, Host, and Kra [5]**).** *There exists a (nonergodic) system $(X, \mathcal{X}, \mu, T)$, and for every $\ell \in \mathbb{N}$ there exists $A \in \mathcal{X}$ with $\mu(A) > 0$ such that*

$$\mu\big(A \cap T^n A \cap T^{2n} A\big) \leq \frac{1}{2}\mu(A)^\ell.$$

for integer $n \neq 0$.

Actually there exists a set A of arbitrarily small positive measure with

$$\mu\big(A \cap T^n A \cap T^{2n} A\big) \leq \mu(A)^{-c\log(\mu(A))}$$

for every integer $n \neq 0$ and for some positive universal constant c.

[6] A sequence $\{a_n\}$ of real numbers is said to be *uniformly recurrence* if for all $\varepsilon > 0$ and all $h \in \mathbb{N}$, the set $\{n\colon |a_{n+h} - a_n| < \varepsilon\}$ is syndetic

The proof is based on Behrend's construction of a set containing no arithmetic progression of length 3:

Theorem 10.6 (Behrend [1]**).** *For all $L \in \mathbb{N}$, there exists a subset $E \subset \{0, 1, \ldots, L-1\}$ having more than $L \exp(-c\sqrt{\log L})$ elements that does not contain any nontrivial arithmetic progression of length* 3.

Proof (of Theorem 10.5). Let $X = \mathbb{T} \times \mathbb{T}$, with Haar measure $\mu = m \times m$ and transformation $T \colon X \to X$ given by $T(x, y) = (x, y + x)$.

Let $E \subset \{0, 1, \ldots, L-1\}$, not containing any nontrivial arithmetic progression of length 3. Define

$$B = \bigcup_{j \in E} \Big[\frac{j}{2L}, \frac{j}{2L} + \frac{1}{4L}\Big),$$

which we consider as a subset of the torus and $A = \mathbb{T} \times B$.

For every integer $n \neq 0$, we have $T^n(x, y) = (x, y + nx)$ and

$$\begin{aligned}\mu\big(A \cap T^n A \cap T^{2n} A\big) &= \iint_{\mathbb{T}\times\mathbb{T}} \mathbf{1}_B(y)\mathbf{1}_B(y+nx)\mathbf{1}_B(y+2nx)\,\mathrm{d}m(y)\,\mathrm{d}m(x) \\ &= \iint_{\mathbb{T}\times\mathbb{T}} \mathbf{1}_B(y)\mathbf{1}_B(y+x)\mathbf{1}_B(y+2x)\,\mathrm{d}m(y)\,\mathrm{d}m(x).\end{aligned}$$

Bounding this integral, we have that:

$$\begin{aligned}\mu\big(A \cap T^n A \cap T^{2n} A\big) &= \iint_{\mathbb{T}\times\mathbb{T}} \mathbf{1}_B(y)\mathbf{1}_B(y+x)\mathbf{1}_B(y+2x)\,\mathrm{d}m(x)\,\mathrm{d}m(y) \\ &\leq \frac{m(B)}{4L}.\end{aligned}$$

By Behrend's theorem, we can choose the set E with cardinality on the order of $L \exp(-c\sqrt{\log L})$. Choosing L sufficiently large, a simple computation gives the statement. □

For longer arithmetic progressions, the counterexample of Theorem 10.3 is based on a construction of Ruzsa. When P is a nonconstant integer polynomial of degree ≤ 2, the subset

$$\big\{P(0), P(1), P(2), P(3), P(4)\big\}$$

of $\mathbb{Z}$ is called a *quadratic configuration of* 5 *terms*, written QC5 for short.

Any QC5 contains at least 3 distinct elements. An arithmetic progression of length 5 is a QC5, corresponding to a polynomial of degree 1.

Theorem 10.7 (Ruzsa [5]**).** *For all $L \in \mathbb{N}$, there exists a subset $E \subset \{0, 1, \ldots, L-1\}$ having more than $L \exp(-c\sqrt{\log L})$ elements that does not contain any* QC5.

Based on this, we show:

Theorem 10.8 (Bergelson, Host and Kra [5]**).** *There exists an ergodic system $(X, \mathcal{X}, \mu, T)$ and, for every $\ell \in \mathbb{N}$, there exists $A \in \mathcal{X}$ with $\mu(A) > 0$ such that*

$$\mu(A \cap T^n A \cap T^{2n} A \cap T^{3n} A \cap T^{4n} A) \leq \frac{1}{2}\mu(A)^\ell$$

for every integer $n \neq 0$.

Once again, the proof gives the estimate $\mu(A)^{-c\log(\mu(A))}$, for some constant $c > 0$.

The construction again involves a simple example: $\mathbb{T}$ is the torus with Haar measure m, $X = \mathbb{T}\times\mathbb{T}$, and $\mu = m\times m$. Let $\alpha \in \mathbb{T}$ be irrational and let $T\colon X \to X$ be

$$T(x, y) = (x + \alpha, y + 2x + \alpha).$$

Combinatorially this example becomes: for all $k \in \mathbb{N}$, there exists $\delta > 0$ such that for infinitely many integers N, there is a subset $A \subset \{1, \ldots, N\}$ with $|A| \geq \delta N$ that contains no more than $\frac{1}{2}\delta^k N$ arithmetic progressions of length ≥ 5 with the same difference.

10.4. Combinatorial consequences. Via a slight modification of the correspondence principle, each of these results translates to a combinatorial statement. For $\varepsilon > 0$ and $E \subset \mathbb{Z}$ with positive upper Banach density (see the definition in Footnote 5), consider the set

$$\{n \in \mathbb{Z}\colon \bar{d}(E \cap (E + n) \cap (E + 2n) \cap \cdots \cap (E + kn)) \geq \bar{d}(E^{k+1}) - \varepsilon\}. \tag{10.2}$$

From Theorems 10.2 and 10.3, for $k = 2$ and for $k = 3$, this set is syndetic, while for $k \geq 4$ there exists a set of integers E with positive upper Banach density such that the set in (10.2) is empty.

We can refine this a bit further. Recall the notation from Szemerédi's theorem: for every $\delta > 0$ and $k \in \mathbb{N}$, there exists $N(\delta, k)$ such that for all $N > N(\delta, k)$, every subset of $\{1, \ldots, N\}$ with at least δN elements contains an arithmetic progression of length k.

For an arithmetic progression $\{a, a + s, \ldots, a + (k - 1)s\}$, s is the *difference* of the progression. Write $\lfloor x\rfloor$ for integer part of x. ¿From Szemerédi's theorem, we can deduce that every subset E of $\{1, \ldots, N\}$ with at least δN elements contains at least $\lfloor cN^2\rfloor$ arithmetic progressions of length k, where $c = c(k, \delta) > 0$ is a constant. Therefore the set E contains at least $\lfloor c(k, \delta)N\rfloor$ progressions of length k with the same difference.

The ergodic results of Theorem 10.2 give some improvement for $k = 3$ and $k = 4$ (see [5] for the precise statement). For $k = 3$, this was strengthened by Green:

Theorem 10.9 (Green [24]**).** *For all $\delta, \varepsilon > 0$, there exists $N_0(\delta, \varepsilon)$ such that for all $N > N_0(\delta, \varepsilon)$ and any $E \subset \{1, \ldots, N\}$ with $|E| \geq \delta N$, E contains at least $(1 - \varepsilon)\delta^3 N$ arithmetic progressions of length* 3 *with the same difference.*

On the other hand, the similar bound for longer progressions with length $k \geq 5$ does not hold. The proof in [5], based on an example of Rusza, does not use ergodic theory. We show that for all $k \in \mathbb{N}$, there exists $\delta > 0$ such that for infinitely many N, there exists a subset E of $\{1, \ldots, N\}$ with $|E| \geq \delta N$ that contains no more than $\frac{1}{2}\delta^k N$ arithmetic progressions of length ≥ 5 with the same step.

10.5. Polynomial averages. One can ask whether similar lower bounds hold for the polynomial averages. For independent polynomials, using the fact that the characteristic factor is the Kronecker factor, we can show:

Theorem 10.10 (Frantzikinakis and Kra [14]**).** *Let $k \in \mathbb{N}$, $(X, \mathcal{X}, \mu, T)$ be a system, $A \in \mathcal{X}$, and let $p_1, p_2, \ldots, p_k\colon \mathbb{Z} \to \mathbb{Z}$ be rationally independent polynomials* *with* $p_i(0) = 0$ *for* $i = 1, 2, \ldots, k$. *Then for every* $\varepsilon > 0$, *the set*

$$\{n \in \mathbb{Z} : \mu(A \cap T^{p_1(n)}A \cap T^{p_2(n)} \cap \cdots \cap T^{p_k(n)}A) > \mu(A)^{k+1} - \varepsilon\}$$

is syndetic.

Once again, this result fails on average.

Via correspondence, analogous to the results of Section 10.4, we have that for $E \subset \mathbb{Z}$ and rationally independent polynomials $p_1, p_2, \ldots, p_k : \mathbb{Z} \to \mathbb{Z}$ with $p_i(0) = 0$ for $i = 1, 2, \ldots, k$, then for all $\varepsilon > 0$, the set

$$\{n \in \mathbb{Z}\colon \bar{d}\big(E \cap \big(E + p_1(n)\big) \cap \cdots \cap \big(E + p_k(n)\big)\big) \geq \bar{d}(E)^{k+1} - \varepsilon\}$$

is syndetic.

Moreover, in [14] we strengthen this and show that there are many configurations with the same n giving the differences: if $p_1, p_2, \ldots, p_k : \mathbb{Z} \to \mathbb{Z}$ are rationally independent polynomials with $p_i(0) = 0$ for $i = 1, 2, \ldots, k$, then for all $\delta, \varepsilon > 0$, there exists $N(\delta, \varepsilon)$ such that for all $N > N(\delta, \varepsilon)$ and any subset $E \subset \{1, \ldots, N\}$ with $|E| \geq \delta N$ contains at least $(1 - \varepsilon)\delta^{k+1}N$ configurations of the form

$$\{x, x + p_1(n), x + p_2(n), \ldots, x + p_k(n)\}$$

for a fixed $n \in \mathbb{N}$.

References

1. F. A. Behrend, *On sets of integers which contain no three terms in arithmetic progression*, Proc. Nat. Acad. Sci. U.S.A. **32** (1946), 331 – 332.
2. V. Bergelson, *Weakly mixing PET*, Ergodic Theory Dynam. Systems **7** (1987), no. 3, 337 – 349.
3. ———, *Ergodic Ramsey theory — an update*, Ergodic Theory of $\mathbb{Z}^d$-Actions (Warwick, 1993 – 1994) (M. Pollicott and K. Schmidt, eds.), London Math. Soc. Lecture Note Ser., vol. 228, Cambridge Univ. Press, Cambridge, 1996.
4. ———, *The multifarious Poincaré recurrence theorem*, Descriptive Set Theory and Dynamical Systems (Marseille-Luminy, 1996) (M. Foreman, A. S. Kechris, A. Louveau, and B. Weiss, eds.), London Math. Soc. Lecture Note Ser., vol. 277, Cambridge Univ. Press, Cambridge, 2000, pp. 31 – 57.
5. V. Bergelson, Host B., and B. Kra, *Multiple recurrence and nilsequences*, Invent. Math. **160** (2005), no. 2, 261 – 303.
6. V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725 – 753.
7. J. Bourgain, *On the maximal ergodic theorem for certain subsets of the positive integers*, Israel J. Math. **61** (1988), no. 1, 39 – 72.
8. J.-P. Conze and E. Lesigne, *Sur un théorème ergodique pour des mesures diagonales*, Probabilités, Publ. Inst. Rech. Math. Rennes, vol. 1987-1, Univ. Rennes I, Rennes, 1987, pp. 1 – 31.
9. ———, *Sur un théorème ergodique pour des mesures diagonales*, C. R. Acad. Sci. Paris Sér. I Math. **306** (1988), no. 12, 491 – 493.
10. I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinaĭ, *Ergodic theory*, Grundlehren Math. Wiss., vol. 245, Springer, New York, 1982.
11. P. Erdős and P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261 – 264.
12. N. Frantzikinakis and B. Kra, *Polynomial averages converge to the product of the integrals*, Israel J. Math. **148** (2005), 267 – 276.
13. ———, *Convergence of multiple ergodic averages for some commuting transformations*, Ergodic Theory Dynam. Systems **25** (2005), no. 3, 799 – 809.
14. ———, *Ergodic averages for independent polynomials and applications*, J. London Math. Soc. (2) **74** (2006), no. 1, 131 – 142.

15. H. Furstenberg, *Strict ergodicity and transformations of the torus*, Amer. J. Math. **83** (1961), 573 – 601.
16. ______, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204 – 256.
17. ______, *Recurrence in ergodic theory and combinatorial number theory*, M. B. Porter Lectures, Princeton Univ. Press, Princeton, NJ, 1981.
18. ______, *Nonconventional ergodic averages*, The legacy of John von Neumann (Hempstead, NY, 1988), Proc. Sympos. Pure Math., vol. 50, Amer. Math. Soc., Providence, RI, 1990, pp. 43 – 56.
19. H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformation*, J. Analyse Math. **34** (1978), 275 – 291.
20. ______, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Analyse Math. **45** (1985), 117 – 268.
21. H. Furstenberg, Y. Katznelson, and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 3, 527 – 552.
22. H. Furstenberg and B. Weiss, *A mean ergodic theorem for* $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$, Convergence in Ergodic Theory and Probability (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ., vol. 5, de Gruyter, Berlin, 1996, pp. 193 – 227.
23. W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465 – 588.
24. B. J. Green, *A Szemerédi-type regularity lemma in abelian groups, with applications*, Geom. Funct. Anal. **15** (2005), no. 2, 340 – 376.
25. B. J. Green and T. C. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2), to appear.
26. ______, *An inverse theorem for the Gowers* U^3*-norm, with applications*, Proc. Edinb. Math. Soc. (2), to appear.
27. ______, *Quadratic uniformity of the Möbius function*, preprint.
28. ______, *Linear equations in primes*, preprint.
29. P. Hall, *A contribution to the theory of groups of prime-power order*, Proc. London Math. Soc. **36** (1933), 29 – 95.
30. P. R. Halmos and J. von Neumann, *Operator methods in classical mechanics.* II, Ann. of Math. (2) **43** (1942), 332 – 350.
31. B. Host, *Convergence of multiple ergodic averages*, School on Information and Randomness (Santiago, 2004), to appear.
32. B. Host and B. Kra, *Convergence of Conze – Lesigne averages*, Ergodic Theory Dynam. Systems **21** (2001), no. 2, 493 – 509.
33. ______, *Averaging along cubes*, Modern Dynamical Systems and Applications, Cambridge Univ. Press, Cambridge, 2004, pp. 123 – 144.
34. ______, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397 – 488.
35. ______, *Convergence of polynomial ergodic averages*, Israel J. Math. **149** (2005), 1 – 19.
36. M. Johnson, *Convergence of polynomial ergodic averages of several variables for some commuting transformations*, preprint.
37. A. Y. Khintchine, *Eine Verschärfung des Poincaréschen "Wiederkehrsatzes"*, Compositio Math. **1** (1934), 177 – 179.
38. B. O. Koopman and J. von Neumann, *Dynamical systems of continuous spectra*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), 255 – 263.
39. B. Kra, *The Green – Tao theorem on arithmetic progressions in the primes: an ergodic of view*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 1, 3 – 23.
40. ______, *From combinatorics to ergodic theory and back again*, Proceddings of the International Congress of Mathematicians, Vol. 3 (Madrid, 2006).
41. L. Kuipers and N. Niederreiter, *Uniform distribution of sequences*, Pure Appl. Math., John Wiley & Sons, New York, 1974.
42. M. Lazard, *Sur certaines suites d'éléments dans les groupes libres et leurs extensions*, C. R. Acad. Sci. Paris **236** (1953), 36 – 38.
43. A. Leibman, *Polynomial sequences in groups*, J. Algebra **201** (1998), no. 1, 189 – 206.
44. ______, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 201 – 213.

45. ______, *Convergence of multiple ergodic averages along polynomials of several variables*, Israel J. Math. **146** (2005), 303 – 316.
46. E. Lesigne, *Sur une nil-variété, les parties minimales associeée à une translation sont uniquement ergodiques*, Ergodic Theory Dynam. Systems **11** (1991), no. 2, 379 – 391.
47. W. Parry, *Ergodic properties of affine transformations and flows on nilmanifolds*, Amer. J. Math. **91** (1969), 757 – 771.
48. J. Petresco, *Sur les commutateurs*, Math. Z. **61** (1954), 348 – 356.
49. H. Poincaré, *Les méthodes nouvelles de la mécanique céleste.* I, Gathiers-Villars, Paris, 1892; II, 1893; III, 1899.
50. M. Ratner, *On Raghunathan's measure conjecture*, Ann. of Math. (2) **134** (1991), no. 3, 545 – 607.
51. A. Sárközy, *On difference sets of sequences of integers.* I, Acta Math. Acad. Sci. Hungar. **31** (1978), no. 1-2, 125 – 149.
52. ______, *On difference sets of sequences of integers.* III, Acta Math. Acad. Sci. Hungar. **31** (1978), no. 3-4, 355 – 386.
53. N. Shah, *Invariant measures and orbit closures on homogeneous spaces for actions of subgroups*, Lie Groups and Ergodic Theory (Mumbai, 996), Tata Inst. Fund. Res., Bombay, 1998.
54. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199 – 245.
55. J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. U.S.A. **18** (1932), 70 – 82.
56. T. Ziegler, *A non-conventional ergodic theorem for a nilsystem*, Ergodic Theory Dynam. Systems **25** (2005), no. 4, 1357 – 1370.
57. ______, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), no. 1, 53 – 97.

Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730, USA

E-mail address: `kra@math.northwestern.edu`