

# From combinatorics to ergodic theory and back again

Bryna Kra\*

**Abstract.** Multiple ergodic averages, such as the average of expressions like  $f_1(T^n x)$   $f_2(T^{2n} x) \dots f_k(T^{kn} x)$ , were first studied in the ergodic theoretic proof of Szemerédi's Theorem on arithmetic progressions. It turns out that all constraints on such averages (in a sense that we describe) have an algebraic character, arising from identities in nilpotent groups. We discuss these averages, several generalizations, and combinatorial implications of the results.

**Mathematics Subject Classification (2000).** 37A30, 11B25, 37A45

**Keywords.** Multiple ergodic theorem, multiple recurrence, arithmetic progressions, nil-systems.

## 1. Additive combinatorics and ergodic theory

A classic result of Ramsey Theory was proved by van der Waerden [53] in the 1920's: if the integers are partitioned into finitely many subsets, at least one of the subsets contains arbitrarily long arithmetic progressions. Erdős and Turán [12] conjectured that a weaker assumption suffices: if  $A$  is a set of integers whose *upper density*

$$\bar{d}(A) = \limsup_{N \rightarrow \infty} \frac{1}{N} |A \cap [1, N]|$$

is positive, then  $A$  contains arbitrarily long arithmetic progressions. Clearly the conjecture immediately implies van der Waerden's Theorem.

The first progress on the Erdős-Turán conjecture came in 1952, when Roth [45] used Fourier analysis to establish that a set of integers with positive upper density contains an arithmetic progression of length 3. Further progress was not until 1969, when Szemerédi [48] showed that the conjecture holds for progressions of length 4. Finally in 1975, Szemerédi [49] resolved the general case with an intricate combinatorial proof.

Soon thereafter, Furstenberg [18] used ergodic theory to give a new proof of Szemerédi's Theorem, and this proof marks the birth of the field of ergodic Ramsey Theory. Since then, ergodic theory has been used to prove new results in combinatorics, such as the multidimensional Szemerédi Theorem [22], the density

---

\*The author was partially supported by NSF grant DMS-0244994.

Hales-Jewett Theorem [24], and the polynomial Szemerédi Theorem [4]; many of these results have yet to be obtained by other means. (Some of these results are explained in Section 4.) Furstenberg’s pioneering work laid out the general strategy for these problems: translate the combinatorial statement into a problem on the intersection of sets in a measure preserving system and then study the average associated to this intersection. The convergence of these multiple ergodic averages is the main focus of this article. A key result is the convergence of the averages associated to Szemerédi’s Theorem (see Section 2 for an explanation of the link):

**Theorem 1.1** (Host and Kra [36]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system,<sup>1</sup>  $k \geq 1$  is an integer, and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ . Then the limit*

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

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

It turns out that a subsystem can be substituted for the original system without affecting the convergence or the value of the limit. Furthermore, this subsystem can be completely described algebraically, with a particular role played by nilpotent groups and their homogeneous spaces. We describe the structural analysis of measure preserving systems needed to prove this in Section 3.

This has led us to a greater understanding of other multiple ergodic averages, including averages with polynomial exponents, prime exponents, and certain averages of commuting transformations, and some of these results are discussed in Section 4. In turn, the multiple convergence theorems have led to deeper connections with exciting developments in number theory and combinatorics, and we discuss some of these developments in Sections 4 and 5.

Although the connection between ergodic theory and additive combinatorics is well established, the depth of this connection is only now beginning to be understood. Szemerédi’s original proof is combinatorial and Furstenberg’s proof uses ergodic theory, yet the two proofs have many formal similarities. These features recur in more recent proofs of Szemerédi’s Theorem, such as those of Gowers [26] and of Tao [50]. In the ergodic setup, with our work in [36] we have a complete understanding of the underlying structures in measure preserving systems that arise in the ergodic theoretic proof of Szemerédi’s Theorem. To elucidate the true nature of the link with additive combinatorics, describing corresponding combinatorial constructions remains a deep open question.

---

<sup>1</sup>By an *(invertible) measure preserving (probability) system*, we mean a quadruple  $(X, \mathcal{X}, \mu, T)$  where  $X$  is a compact metrizable set,  $\mathcal{X}$  denotes the Borel  $\sigma$ -algebra on  $X$ ,  $\mu$  is a probability measure on  $(X, \mathcal{X})$ , and  $T: X \rightarrow X$  is an invertible measurable map with  $\mu(A) = \mu(T^{-1}A)$  for all  $A \in \mathcal{X}$ . Even when not explicitly stated, the measure is assumed to be a probability measure and the transformation is assumed to be invertible.

## 2. Multiple ergodic averages

**2.1. Multiple recurrence.** We start with the connection between regularity properties of subsets of integers and recurrence in measure preserving systems:

**Correspondence Principle** (Furstenberg [18], [20]). *Let  $E$  be a set of integers with positive upper density. There exist a measure preserving system  $(X, \mathcal{X}, \mu, T)$  and a subset  $A \subseteq \mathcal{X}$  such that  $\mu(A) = \bar{d}(E)$  and*

$$\bar{d}((E + n_1) \cap (E + n_2) \cap \dots \cap (E + n_k)) \geq \mu(T^{-n_1} A \cap T^{-n_2} A \cap \dots \cap T^{-n_k} A)$$

for any integer  $k \geq 1$  and integers  $n_1, n_2, \dots, n_k \geq 0$ .

Furstenberg then deduced Szemerédi's Theorem by showing that any system  $(X, \mathcal{X}, \mu, T)$  is *multiply recurrent*, meaning that for all  $A \in \mathcal{X}$  with positive measure, there exists  $n \in \mathbb{N}$  such that

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

To produce such  $n \in \mathbb{N}$  using ergodic theoretic methods, it is natural to average the expression in (2) over  $n$ . If one can show that the limit inferior of this average is positive, the existence of some  $n \in \mathbb{N}$  satisfying (2) follows immediately. Thus combined with the Correspondence Principle, Szemerédi's Theorem follows from:

**Multiple Recurrence Theorem** (Furstenberg [18]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system,  $A \in \mathcal{X}$  has positive measure, and  $k \geq 1$  is an integer. Then*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^n A \cap T^{2n} A \cap \dots \cap T^{kn} A) > 0. \quad (3)$$

Poincaré Recurrence is implied by the case  $k = 1$ : for any set  $A$  with positive measure, there exist infinitely many  $n \in \mathbb{N}$  such that  $\mu(A \cap T^n A) > 0$ . Although it is easy to prove Poincaré Recurrence directly, we can also view it as a corollary of the von Neumann Ergodic Theorem, which implies that for a set  $A \in \mathcal{X}$  with positive measure, the limit

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

exists and is positive. For higher order multiple recurrence ( $k \geq 2$ ), this method of studying the corresponding multiple ergodic average is the only known method for producing  $n$  such that (2) holds.

**2.2. Multiple ergodic averages.** A natural question arises: is the “lim inf” in (3) actually a limit? More generally, if  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system,  $k \geq 1$  is an integer, and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ , does the multiple ergodic average

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \quad (4)$$

converge as  $N$  tends to infinity, and in what sense does it converge? Taking each  $f_i$  to be the indicator function  $\mathbf{1}_A$  of a set  $A$ , multiplying by  $\mathbf{1}_A$  and integrating with respect to  $\mu$ , we obtain the average in (3). For  $k = 1$ , the existence of this limit in  $L^2(\mu)$  is the von Neumann Ergodic Theorem.

A measure preserving transformation  $T: X \rightarrow X$  induces an operator  $U_T$ , on functions in  $L^2(\mu)$  defined by  $U_T f(x) = f(Tx)$ . In a standard abuse of notation, we denote the operator  $U_T$  by  $T$  and write  $Tf(x) = f(Tx)$ . In general we assume that the measure preserving system  $(X, \mathcal{X}, \mu, T)$  is *ergodic*, meaning that the only sets  $A \in \mathcal{X}$  satisfying  $T^{-1}A \subseteq A$  have either full or zero measure. Since a general system can be decomposed into its ergodic components, for most of the theorems we consider it suffices to assume that the system is ergodic.

When the system is ergodic, for  $k = 1$  the limit of (4) in  $L^2(\mu)$  is the integral  $\int_X f_1 d\mu$  and in particular is constant. However, without some assumption on the system, for  $k \geq 2$ , the limit in (4) need not be constant. For example, if  $X$  is the circle  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ ,  $T: \mathbb{T} \rightarrow \mathbb{T}$  is the rotation  $Tx = x + \alpha \pmod{1}$  for some  $\alpha \in \mathbb{T}$ ,  $f_1(x) = \exp(4\pi i x)$  and  $f_2(x) = \exp(-2\pi i x)$ , then  $f_1(T^n x)f_2(T^{2n} x) = f_2^{-1}(x)$  for all  $n \in \mathbb{N}$ . In particular, the double average

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

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  $\int f_1(x+t)f_2(x+2t) dt$ , which in general is not constant.)

The limit behavior of the double average depends on rotational behavior in the system. To make this more precise, we introduce some terminology. A *factor* of a measure preserving system  $(X, \mathcal{X}, \mu, T)$  can be defined in one of several equivalent ways: it is a  $T$ -invariant sub- $\sigma$ -algebra  $\mathcal{Y}$  of  $\mathcal{X}$ , it is a measure preserving system  $(Y, \mathcal{Y}, \nu, S)$  and a measurable map  $\pi: X \rightarrow Y$  such that  $\mu \circ \pi^{-1} = \nu$  and  $S \circ \pi(x) = \pi \circ T(x)$  for  $\mu$ -almost all  $x \in X$ , and it is a  $T$ -invariant subspace  $\mathcal{F}$  of  $L^\infty(\mu)$ . The equivalence between the first two definitions follows by identifying  $\pi^{-1}(\mathcal{Y})$  with a  $T$ -invariant sub- $\sigma$ -algebra of  $\mathcal{X}$  and noting that any  $T$ -invariant sub- $\sigma$ -algebra of  $\mathcal{X}$  arises in this way. Setting  $\mathcal{F} = L^\infty(\mathcal{Y})$ , we have that the first definition implies the third and taking  $\mathcal{Y}$  to be the  $\sigma$ -algebra generated by  $\mathcal{F}$ -measurable sets, we have the converse. Depending on the context, we use any of these three characterizations interchangeably. In a slight abuse of notation, we use the same letter to denote the transformation in the whole space and in a factor.

If  $(Y, \mathcal{Y}, \nu, T)$  is a factor of  $(X, \mathcal{X}, \mu, T)$  and  $f \in L^2(\mu)$ , the *conditional expectation*  $\mathbb{E}(f | \mathcal{Y})$  of  $f$  with respect to  $\mathcal{Y}$  is the orthogonal projection of  $f$  onto  $L^2(\nu)$ . Let  $\mathbb{E}(f | Y)$  denote the function on  $Y$  defined by  $\mathbb{E}(f | Y) \circ \pi = \mathbb{E}(f | \mathcal{Y})$ , where  $\pi: X \rightarrow Y$  is the natural projection. This expectation is characterized by

$$\int_Y \mathbb{E}(f | Y)(y)g(y) d\nu(y) = \int_X f(x)g(\pi(x)) d\mu(x)$$

for all  $g \in L^\infty(\mu)$ .

A measure preserving system  $(X, \mathcal{X}, \mu, T)$  is said to be *weakly mixing* if the only measurable eigenfunctions of the operator on  $L^2(\mu)$  induced by the transformation  $T$  are constant. An alternate characterization of weakly mixing can be given in terms of a factor: the measure preserving system  $(X, \mathcal{X}, \mu, T)$  is not weakly mixing if and only if it has a nontrivial factor which is a rotation on a compact abelian group. The maximal such (group rotation) factor is known as the *Kronecker factor*. A rotation on a circle is not weakly mixing.

Taking the rotational behavior into account, the double average  $\frac{1}{N} \sum T^n f_1 \cdot T^{2n} f_2$  can be understood. The obvious phenomenon is that for  $\mu$ -almost every  $x$ , the triple  $(x, T^n x, T^{2n} x)$  projects to an arithmetic progression in the Kronecker factor  $\mathcal{Z}$ . Furstenberg showed that this restriction is the only restriction, meaning that

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

tends to 0 as  $N \rightarrow \infty$ . Thus to prove convergence of the double average, it suffices to replace each  $f_i$ , for  $i = 1, 2$ , by its conditional expectation  $\mathbb{E}(f_i | \mathcal{Z})$  on the Kronecker factor. In particular, this means that one can assume that the system is an ergodic rotation on a compact abelian group. Then one can easily use Fourier analysis to show the existence of the limit. (The Kronecker factor is said to be *characteristic* for the double average. See Section 3.1 for the general definition.) The double average is the simplest example of a “nonconventional ergodic average,” where even if the system is assumed to be ergodic, the limit need not be constant.

Furthermore, if the system is assumed to be weakly mixing, Furstenberg [18] showed the existence of the limit in (4) for all  $k \geq 1$ . Moreover, in this case the limit takes on a particularly simple form: the average converges in  $L^2(\mu)$  to the product of the integrals  $\int f_1 d\mu \int f_2 d\mu \dots \int f_k d\mu$ .

For a general system, the limiting behavior for  $k \geq 3$  is more complicated and group rotations do not suffice for describing the long term behavior. For example, if  $f(Tx) = \lambda f(x)$  for some  $|\lambda| = 1$  and  $F(Tx) = f(x)F(x)$ , then

$$F(T^n x) = f(x) f(Tx) \dots f(T^{n-1} x) F(x) = \lambda^{\frac{n(n-1)}{2}} (f(x))^n F(x) .$$

Therefore

$$F(x) (F(T^n x))^{-3} (F(T^{2n} x))^3 (F(T^{3n} x))^{-1} = 1 .$$

Projection to the Kronecker factor does not capture the behavior of generalized eigenfunctions, meaning that there is some relation among  $x, T^n x, T^{2n} x$  and  $T^{3n} x$  that does not arise from the Kronecker factor. See Furstenberg [21] for a more intricate example, showing that even such generalized eigenfunctions do not suffice in determining the limiting behavior for  $k = 3$ .

Using a new structural analysis for ergodic systems, we describe the algebraic constraints on  $n$ -tuples  $x, T^n x, T^{2n} x, \dots, T^{(k-1)n} x$ , and use this to obtain convergence of the averages in (4). Existence of the limit in  $L^2(\mu)$  for  $k = 1$  is the von Neumann Ergodic Theorem and existence for  $k = 2$  was proven by Furstenberg [18]. Existence of the limit for  $k = 3$  with the hypothesis of total ergodicity,

meaning that  $T$  and all its powers are ergodic, was proven by Conze and Lesigne ([9], [10], and [11]); this is the first place that a natural generalization (playing a major role for higher  $k$ ) of the Kronecker factor, a 2-step nilsystem, appears as a factor. In the general case for  $k = 3$ , existence was shown by Furstenberg and Weiss [25] and by Host and Kra [33] (see also [34]). We proved existence of the limit (1) for all integers  $k \geq 1$  in [36] and this is the statement of Theorem 1.1. More recently, Ziegler [57] has a different approach for showing the existence of the limit in the general case. The existence of the pointwise limit is a much more difficult problem and convergence is only known for  $k = 2$ , due to Bourgain [8].

The key role in the analysis used to prove the existence of the limit in (1) is played by nilpotent groups and their homogeneous spaces. We start with a brief overview of the ingredients in the proof of Theorem 1.1.

### 3. Structural analysis

**3.1. Characteristic factors.** A general strategy for showing the existence of an average, such as that of (1), is to find a factor such that the limiting behavior is unchanged when each function is replaced by its conditional expectation on this factor. More precisely, a factor  $\mathcal{Y} \subseteq \mathcal{X}$  is a *characteristic factor* (or more succinctly, is *characteristic*) for the average

$$\frac{1}{N} \sum_{n=0}^{N-1} T^{a_1(n)} f_1 \cdot T^{a_2(n)} f_2 \cdot \dots \cdot T^{a_k(n)} f_k$$

if the difference between this average and the same average with each function replaced by its conditional expectation on  $\mathcal{Y}$

$$\frac{1}{N} \sum_{n=0}^{N-1} T^{a_1(n)} \mathbb{E}(f_1 | \mathcal{Y}) \cdot T^{a_2(n)} \mathbb{E}(f_2 | \mathcal{Y}) \cdot \dots \cdot T^{a_k(n)} \mathbb{E}(f_k | \mathcal{Y})$$

converges to 0 in  $L^2(\mu)$  as  $N$  tends to infinity. For example, when  $a_1(n) = n$  and  $a_2(n) = 2n$ , the Kronecker factor is characteristic for the double average. Although the term characteristic factor only appeared explicitly in the literature fairly recently [21], the method is implicit in Furstenberg's original proof [18] of Szemerédi's Theorem.

If one can find a characteristic factor for a given average, then it suffices to prove convergence when the characteristic factor is substituted for the original system. Proving convergence for the factor is then easier when the factor has a sufficiently explicit and "simple" description.

We follow this general strategy, but with a different point of view. Rather than manipulating a particular average that we want to understand, we start with an abstract construction of characteristic factors. The construction (following [36]) is based on an inductively defined sequence of measures and of seminorms,<sup>2</sup> which are then used to define the factors. We now outline this construction.

<sup>2</sup>Although the definition and context are on the surface quite different, these seminorms turn

**3.2. Definition of measures and seminorms.** Fix an integer  $k \geq 0$ . We write a point  $\omega \in \{0, 1\}^k$  as  $\omega = \omega_1\omega_2 \dots \omega_k$  with  $\omega_i \in \{0, 1\}$ , omitting commas and parentheses, and let  $|\omega| = \omega_1 + \omega_2 + \dots + \omega_k$ . Fixing an ergodic measure preserving system  $(X, \mathcal{X}, \mu, T)$ , let  $X^{[k]} = X^{2^k}$  and let  $T^{[k]} : X^{[k]} \rightarrow X^{[k]}$  be the map  $T \times T \times \dots \times T$ , taken  $2^k$  times. Elements of  $X^{[k]}$  are written  $\mathbf{x} = (x_\omega : \omega \in \{0, 1\}^k)$ . There is a natural identification of  $X^{[k+1]}$  and  $X^{[k]} \times X^{[k]}$ , with a point  $\mathbf{x} \in X^{[k+1]}$  being identified with  $(\mathbf{x}', \mathbf{x}'') \in X^{[k]} \times X^{[k]}$ , where  $x'_\omega = x_{\omega 0}$  and  $x''_\omega = x_{\omega 1}$  for each  $\omega \in \{0, 1\}^k$ .

By induction, we define a probability measure  $\mu^{[k]}$  on  $X^{[k]}$ , that is invariant under  $T^{[k]}$ . Set  $\mu^{[0]} = \mu$ . Assume that  $\mu^{[k]}$  is defined for some  $k \geq 0$ . Let  $\mathcal{I}^{[k]}$  denote the  $\sigma$ -algebra of  $T^{[k]}$ -invariant subsets of  $X^{[k]}$ .

Under the natural identification of  $X^{[k+1]}$  with  $X^{[k]} \times X^{[k]}$ , define the measure preserving (probability) system  $(X^{[k+1]}, \mu^{[k+1]}, T^{[k+1]})$  to be the *relatively independent joining* of  $(X^{[k]}, \mu^{[k]}, T^{[k]})$  with itself over  $\mathcal{I}^{[k]}$ ; this means that the measure  $\mu^{[k+1]}$  satisfies for all bounded functions  $F'$  and  $F''$  on  $X^{[k]}$ ,

$$\int_{X^{[k+1]}} F'(\mathbf{x}')F''(\mathbf{x}'') d\mu^{[k+1]}(\mathbf{x}) = \int_{X^{[k]}} \mathbb{E}(F' | \mathcal{I}^{[k]}) \mathbb{E}(F'' | \mathcal{I}^{[k]}) d\mu^{[k]}.$$

The measure  $\mu^{[k+1]}$  is invariant under  $T^{[k+1]}$  and the two natural projections on  $X^{[k]}$  are each  $\mu^{[k]}$ . By induction, each of the  $2^k$  natural projections of  $\mu^{[k]}$  on  $X$  is equal to  $\mu$ . Letting  $C : \mathbb{C} \rightarrow \mathbb{C}$  denote the conjugacy map  $z \mapsto \bar{z}$ , we have that for a bounded function  $f$  on  $X$ , the integral

$$\int_{X^{[k]}} \prod_{\omega \in \{0, 1\}^j} C^{|\omega|} f(x_\omega) d\mu^{[k]}(\mathbf{x})$$

is real and nonnegative.

Therefore, for a function  $f \in L^\infty(\mu)$ , we can define

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

One can also view this definition as an average over the cube  $\{0, 1\}^k$ . A convergence theorem for general averages along cubes is also proved in [36], and the connection between averages along cubes and along arithmetic progressions is more fully explained in Host [32].

Using the Ergodic Theorem and the definition of the measures, we have that for any  $f \in L^\infty(\mu)$ ,

$$\|f\|_{k+1} = \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \|f \cdot T^n f\|_k^{2^k} \right)^{1/2^{k+1}}. \quad (5)$$

---

out to be a generalization of the norms introduced by Gowers [26] in his proof of Szemerédi's Theorem. To recover the Gowers norms, consider the space  $\mathbb{Z}/N\mathbb{Z}$ , the transformation  $x \mapsto x + 1 \pmod N$ , and the uniform measure assigning each element of  $\mathbb{Z}/N\mathbb{Z}$  weight  $1/N$ . The Gowers norms were later used by Green and Tao [28] in a spirit closer to ergodic theory and their use in our work [36]. See [32] and [39] for more on this connection.

To show that the map  $f \mapsto \|f\|_k$  is a seminorm on  $L^\infty(\mu)$ , one derives a version of the Cauchy-Schwarz inequality and uses it to show subadditivity. Positivity immediately follows from Equation (5). (See [36] for details.)

We now return to the original averages along arithmetic progressions and show that the long term behavior of the average (1) is controlled by the seminorms we have constructed:

**Theorem 3.1** (Host and Kra [36]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is an ergodic measure preserving probability system. Let  $k \geq 1$  be an integer and assume that  $f_1, f_2, \dots, f_k$  are functions on  $X$  with  $\|f_1\|_\infty, \|f_2\|_\infty, \dots, \|f_k\|_\infty \leq 1$ . Then*

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} T^n f_1 \cdot T^{2n} f_2 \cdot \dots \cdot T^{kn} f_k \right\|_{L^2(\mu)} \leq \min_{1 \leq j \leq k} (j \|f_j\|_k) .$$

The proof relies on a standard method for finding characteristic factors, which is an iterated use of a variation of the van der Corput Lemma on differences (see for example [40] or [1]):

**van der Corput Lemma.** *Assume that  $\mathcal{H}$  is a Hilbert space with inner product  $\langle \cdot, \cdot \rangle$  and norm  $\|\cdot\|$ , and that  $\xi_n, n \geq 0$ , is a sequence in  $\mathcal{H}$  with  $\|\xi_n\| \leq 1$  for all  $n$ . Then*

$$\limsup_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} \xi_n \right\|^2 \leq \limsup_{H \rightarrow \infty} \frac{1}{H} \sum_{h=0}^{H-1} \limsup_{N \rightarrow \infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \langle \xi_{n+h}, \xi_n \rangle \right| .$$

In our context, we apply this to the Hilbert space  $L^2(\mu)$  of unitary operators that is naturally associated with the system  $(X, \mathcal{X}, \mu, T)$ . The seminorms we construct reflect  $k$  successive uses of the van der Corput Lemma, with the number of steps in the iteration increasing with the complexity of the averages. Theorem 3.1 follows using induction, the Cauchy-Schwarz Inequality, and the van der Corput Lemma.

**3.3. The factors.** We then show that for every integer  $k \geq 1$ , the seminorms define factors  $Z_{k-1}$ . One presentation of these factors is obtained by defining their orthogonal complements: for  $k \geq 1$ , it can be shown that there exists a  $T$ -invariant  $\sigma$ -algebra  $\mathcal{Z}_{k-1}$  of  $\mathcal{X}$  such that for  $f \in L^\infty(\mu)$ ,

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

Thus a bounded function  $f$  is measurable with respect to  $\mathcal{Z}_{k-1}$  if and only if  $\int fg d\mu = 0$  for all functions  $g \in L^\infty(\mu)$  with  $\|g\|_{k-1} = 0$ .

Then  $Z_{k-1}$  is defined to be the factor of  $X$  associated to the sub- $\sigma$ -algebra  $\mathcal{Z}_{k-1}$ . Thus defined,  $Z_0$  is the trivial factor,  $Z_1$  is the Kronecker factor and more generally,  $Z_k$  is a compact abelian group extension of  $Z_{k-1}$ . Furthermore, the sequence of factors is increasing

$$Z_0 \leftarrow Z_1 \leftarrow Z_2 \leftarrow \dots \leftarrow X$$



and if  $T$  is weakly mixing, then  $Z_k$  is the trivial factor for all  $k$ . In this terminology, Theorem 3.1 states that the factor  $Z_k$  is characteristic for the average (1).

The bulk of the work, and also the most technical portion, is devoted to the description of these factors. The initial idea is natural: we associate to each of these factors the group of transformations which preserves the natural cubic structure that arises in the construction. This group is nilpotent. We then conclude that for a sufficiently large (for our purposes) class of systems, this group is a Lie group and acts transitively on the space. Therefore, the constructed system is a *translation on a nilmanifold*. More precisely, if  $G$  is a  $k$ -step nilpotent Lie group and  $\Gamma$  is a discrete cocompact subgroup, then the compact space  $X = G/\Gamma$  is said to be a  *$k$ -step nilmanifold*. The group  $G$  acts on  $G/\Gamma$  by left translation and the translation by a fixed element  $a \in G$  is given by  $T_a(g\Gamma) = (ag)\Gamma$ . There exists a unique probability measure  $m_{G/\Gamma}$ , the *Haar measure*, on  $X$  that is invariant under the action of  $G$  by left translations. Fixing an element  $a \in G$ , we call the system  $G/\Gamma$  with its associated Borel  $\sigma$ -algebra, Haar measure  $m_{G/\Gamma}$ , and translation  $T_a$  a  *$k$ -step nilsystem*. The system  $(X, \mathcal{X}, \mu, T)$  is an *inverse limit of a sequence of factors*  $(X_n, \mathcal{X}_n, \mu_n, T)$  if  $\mathcal{X}_{n \in \mathbb{N}}$  is an increasing sequence of  $T$ -invariant  $\sigma$ -algebras such that  $\bigvee_{n \in \mathbb{N}} \mathcal{X}_n = \mathcal{X}$  up to a set of measure 0. If in addition each factor  $(X_n, \mathcal{X}_n, \mu_n, T)$  is isomorphic to a  $k$ -step nilsystem for  $n \in \mathbb{N}$ , the system  $(X, \mathcal{X}, \mu, T)$  is an *inverse limit of  $k$ -step nilsystems*.

The structure theorem states:

**Theorem 3.2** (Host and Kra [36]). *There exists a characteristic factor for the averages in (1) which is isomorphic to an inverse limit of  $k$ -step nilsystems.*

An expository outline of the proof is also given in Host [32]. A posteriori, the role played by the nilpotent structure is not surprising: for a  $k$ -step nilsystem  $(X, \mathcal{X}, \mu, T)$  and  $x \in X$ , the  $(k+1)$ st term  $T^k x$  of an arithmetic progression is constrained by the first  $k$  terms  $x, Tx, T^2x, \dots, T^{k-1}x$ .

Convergence of the linear (meaning the exponents  $n, 2n, \dots, kn$  are linear) multiple ergodic average then follows easily from general properties of nilmanifolds proved by Lesigne [43] for connected groups and proved in the general case by Leibman [41].

## 4. Generalizations of multiple convergence

**4.1. Polynomial averages.** It is natural to ask what configurations, other than arithmetic progressions, must occur in sets of integers with positive upper density. Sárközy [46] showed that if a subset of integers  $E$  has positive upper density and  $p: \mathbb{Z} \rightarrow \mathbb{Z}$  is a polynomial with  $p(0) = 0$ , then there exist  $x, y \in E$  and  $n \in \mathbb{N}$  such that  $x - y = p(n)$ . Furstenberg [19] gave a proof using ergodic theory. Once again, Furstenberg's proof used the correspondence principle and a recurrence result, this time along polynomial times. Bergelson and Leibman generalized the recurrence result for multiple polynomials:

**Theorem 4.1** (Bergelson and Leibman [4]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is an invertible measure preserving system,  $A \in \mathcal{X}$  has positive measure,  $k \geq 1$  is an integer, and  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  are polynomials with  $p_j(0) = 0$  for  $j = 1, 2, \dots, k$ . Then*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(A \cap T^{-p_1(n)} A \cap T^{-p_2(n)} A \cap \dots \cap T^{-p_k(n)} A) > 0 .$$

The result in [4] is actually quite a bit stronger; they prove a multidimensional version of this statement (see Section 4.2), meaning that one replaces the  $j$ -th occurrence of  $T$  by  $T_j$ , for  $k$  commuting measure preserving transformations  $T_1, T_2, \dots, T_k$  of the measure space  $(X, \mathcal{X}, \mu)$ . A polynomial version of Szemerédi's Theorem follows immediately via Furstenberg's Correspondence Principle.

The polynomial recurrence theorem naturally leads to the corresponding convergence question for multiple polynomial averages:

**Theorem 4.2** (Host and Kra [37], Leibman [42]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system,  $k \geq 1$  is an integer,  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  are polynomials, and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ . Then the limit*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{p_1(n)} f_1 \cdot T^{p_2(n)} f_2 \cdot \dots \cdot T^{p_k(n)} f_k \tag{6}$$

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

For a weakly mixing system, convergence of (6) was proved by Bergelson [1]. In an arbitrary measure preserving system, Furstenberg and Weiss [25] proved convergence for  $k = 2$  with  $p_1(n) = n$  and  $p_2(n) = n^2$  and  $p_1(n) = n^2$  and  $p_2(n) = n^2 + n$ . Weak convergence was proven in [37], as well as convergence in  $L^2(\mu)$  in most cases. The remaining case, along with a generalization for multiparameter polynomials, was completed in [42].

As with the linear average corresponding to exponents  $n, 2n, \dots, kn$ , the behavior of a general polynomial average is controlled by the seminorms  $\|\cdot\|_k$ . Using an inductive procedure like that of [1], the averages in (6) can be reduced to an average only with linear exponents and we obtain a result for a polynomial average analogous to Theorem 3.1. Using the structure theorem (Theorem 3.2), we have that a characteristic factor for a polynomial average is once again an inverse limit of nilsystems.

The number of steps needed in the inductive procedure used to reduce the average (6) to linear terms depends on the choice of polynomials. As might be expected, more terms and higher degree increases the number of steps needed and so the complexity of the corresponding nilsystem rises. However, it turns out that the linearly dependent family  $\{n, 2n, \dots, kn\}$  is in some sense the most difficult. For a general polynomial family, the minimal characteristic factor  $Z_k$  (meaning smallest  $k$ ) is unknown. Yet for *rationally independent polynomials*, meaning polynomials  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  with  $\{1, p_1, p_2, \dots, p_k\}$  linearly independent over  $\mathbb{Q}$ , the characteristic factor (and therefore the value of the limit) is particularly simple

and is independent of the choice of polynomials. Answering a question of Bergelson posed in [2], we show:

**Theorem 4.3** (Frantzikinakis and Kra [14]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is a totally ergodic measure preserving system,  $k \geq 1$  is an integer,  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  are rationally independent polynomials, and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ . Then*

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

converges to 0 in  $L^2(\mu)$  as  $N \rightarrow \infty$ .

Our proof uses the machinery of the Structure Theorem, but we ultimately show that the procyclic factor (an inverse limit of cyclic groups), which is contained in the Kronecker factor, is characteristic for this average. It would be interesting to prove the theorem directly, avoiding the use of nilsystems.

**4.2. Averages for commuting transformations.** Furstenberg and Katznelson generalized multiple recurrence for commuting transformations:

**Theorem 4.4** (Furstenberg and Katznelson [22]). *Assume that  $(X, \mathcal{X}, \mu)$  is a probability space,  $k \geq 1$  is an integer,  $T_1, T_2, \dots, T_k: X \rightarrow X$  are commuting measure preserving transformations, and  $A \in \mathcal{X}$  has positive measure. Then*

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

(Other generalizations, including combinations of the commuting and polynomial averages, are contained in [23] and [5].) Furstenberg's correspondence principle immediately implies a combinatorial version, the multidimensional version of Szemerédi's Theorem.

Once again, it is natural to ask about convergence of the corresponding commuting multiple ergodic average:

**Question 4.5.** *If  $k \geq 1$  is an integer,  $T_1, T_2, \dots, T_k: X \rightarrow X$  are commuting measure preserving transformations of a probability space  $(X, \mathcal{X}, \mu)$ ,  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  are polynomials, and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ , does*

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

exist in  $L^2(\mu)$ ?

For two transformations and exponents  $p_1(n) = p_2(n) = n$ , existence of the limit in  $L^2(\mu)$  was shown by Conze and Lesigne [9]. For arbitrary  $k \geq 1$ , under the assumptions that  $T_j$  is ergodic for  $j = 1, 2, \dots, k$  and that  $T_i T_j^{-1}$  is ergodic for  $i \neq j$ ,  $i, j \in \{1, 2, \dots, k\}$ , existence of the limit with exponents  $p_1(n) = p_2(n) = \dots = p_k(n) = n$  in  $L^2(\mu)$  is shown in [15]. However, the general case (even with all exponents equal to  $n$ ) remains open and it is easy to construct systems such that the characteristic factors are not nilsystems.

**4.3. Sequences related to prime numbers.** Recently, a new chapter in ergodic Ramsey Theory was opened, with ergodic theoretic techniques adapted for use outside of the field. A particularly spectacular result in this direction is Green and Tao's proof [28] that the prime numbers contain arbitrarily long arithmetic progressions. The connections between the proof of Green and Tao and ergodic theory are further explained in the expository articles of Host [32], Kra [39], and Tao [51]. In turn, Green and Tao's results make it possible to study convergence for other multiple ergodic averages, leading us to a greater understanding of patterns in a set of integers with positive upper density. Green and Tao ([29], [30], [31]) proved a strong uniformity result on the prime numbers and using this result, we can show that the shifted primes have multiple recurrence properties. (See also the survey articles of Green [27] and of Tao [52].) Letting  $\mathbb{P}$  denote the primes, we show:

**Theorem 4.6** (Frantzikinakis, Host, and Kra [13]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system and  $A \in \mathcal{X}$  has positive measure. Then there exists  $n \in \mathbb{P} - 1$  such that*

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

The same statement holds with  $\mathbb{P} - 1$  replaced by  $\mathbb{P} + 1$ . Thus the shifted primes  $\mathbb{P} - 1$  and  $\mathbb{P} + 1$  are sets of 2-recurrence. For single recurrence, this was proven by Sárközy [47] and reproved using ergodic methods by Wierdl [54]. (Bourgain [7] and Wierdl [55] also proved several stronger results on pointwise convergence along primes.)

An immediate corollary is that a set of integers with positive upper density contains infinitely many arithmetic progressions of length 3 whose common difference is of the form  $p - 1$  for some prime  $p$  (and similarly of the form  $p + 1$ ).

Roughly speaking, we prove this by comparing the associated double average along primes with the usual double average, and show that the difference converges to 0. This relies on the uniformity result on the prime numbers of Green and Tao. It turns out that the Kronecker factor is characteristic for the associated average. The added complication is that one must work with  $\mathbb{Z}/N\mathbb{Z}$  as the underlying space instead of  $\mathbb{Z}$ .

Using the same methods, we also show the existence of the related double ergodic average:

**Theorem 4.7** (Frantzikinakis, Host, and Kra [13]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system and  $f_1, f_2 \in L^\infty(\mu)$ . Then*

$$\lim_{N \rightarrow \infty} \frac{1}{|\mathbb{P} \cap [0, N]|} \sum_{n \in \mathbb{P}, n < N} T^n f_1 \cdot T^{2n} f_2$$

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

The same reduction to a uniformity statement about the prime numbers, for both recurrence and convergence, works for multiple recurrence and convergence of all lengths. However, the needed uniformity result for prime numbers remains open for longer progressions.

## 5. Lower bounds for multiple ergodic averages

**5.1. Khintchine Recurrence.** As described in Section 2, the first step in Furstenberg's Multiple Recurrence Theorem (Poincaré Recurrence) is an immediate corollary of the von Neumann Ergodic Theorem. However, using the full description of the limit, and not only positivity of the limit inferior, one can make a finer statement about the frequency of recurrence. More precisely, a set  $E \subseteq \mathbb{Z}$  is *syndetic*<sup>3</sup> if there exists  $M \in \mathbb{N}$  such that every interval of length  $M$  has nontrivial intersection with the set  $E$ . Khintchine generalized Poincaré Recurrence and showed:

**Theorem 5.1** (Khintchine [38]). *If  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system and  $A \in \mathcal{X}$ , then for all  $\varepsilon > 0$ , the set*

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

*is syndetic.*

As this follows easily from the von Neumann Ergodic Theorem, one can ask for the analogous results corresponding to other multiple recurrence theorems: if  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system,  $A \in \mathcal{X}$ ,  $k \geq 1$  is an integer,  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  are polynomials with  $p_j(0) = 0$  for  $j = 1, 2, \dots, k$ , and  $\varepsilon > 0$ , is the set

$$\{n \in \mathbb{Z}: \mu(A \cap T^{-p_1(n)} A \cap \dots \cap T^{p_k(n)} A) > \mu(A)^{k+1} - \varepsilon\} \quad (7)$$

syndetic?

Surprisingly enough, the answer depends on the number  $k$  of polynomials and on the linear dependencies among the polynomials. For rationally independent polynomials, using the fact that a characteristic factor takes on a simple form, we show that the measure of the intersection in (7) is as large as possible on a syndetic set:

**Theorem 5.2** (Frantzikinakis and Kra [16]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is an invertible measure preserving system,  $A \in \mathcal{X}$ ,  $k \geq 1$  is an integer, and  $p_1, p_2, \dots, p_k: \mathbb{Z} \rightarrow \mathbb{Z}$  are rationally independent polynomials with  $p_j(0) = 0$  for  $j = 1, 2, \dots, k$ . Then for all  $\varepsilon > 0$ , the set*

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

*is syndetic.*

This sharply contrasts the behavior for a family of linearly dependent polynomials, such as the linear polynomials corresponding to Szemerédi's Theorem, where the behavior depends on the number of linear terms. This dependence is illustrated in the following two theorems:

---

<sup>3</sup>A syndetic set is sometimes known in the literature as *relatively dense*. A syndetic set in  $\mathbb{Z}$  is said to have *bounded gaps*.

**Theorem 5.3** (Bergelson, Host, and Kra [3]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is an ergodic measure preserving system,  $A \in \mathcal{X}$ , and  $k \geq 1$  is an integer. Then for all  $\varepsilon > 0$ , the sets*

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

and

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

are syndetic.

While ergodicity is not needed in Khintchine's Theorem, it is a necessary hypothesis in Theorem 5.3. In [3], we construct a counterexample for the nonergodic case.

For arithmetic progressions of length  $\geq 5$ , the analogous result does not hold. Based on a result of Ruzsa contained in the Appendix of [3], we show

**Theorem 5.4** (Bergelson, Host, and Kra [3]). *There exists an ergodic system  $(X, \mathcal{X}, \mu, T)$  such that for all integers  $\ell \geq 1$  and all  $\varepsilon > 0$ , there exists a set  $A = A(\ell, \varepsilon) \in \mathcal{X}$  with positive measure such that*

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

for every integer  $n \neq 0$ .

The proofs of these theorems are based on a decomposition result for the *multicorrelation sequence*

$$\int f(x) f(T^n x) f(T^{2n} x) \dots f(T^{kn} x) d\mu(x), \quad (8)$$

where  $(X, \mathcal{X}, \mu, T)$  is a measure preserving system,  $f \in L^\infty(\mu)$ , and  $k, n \geq 1$  are integers. We decompose such a sequence into two pieces, one of which is small in terms of density and the second of which arises from a nilsystem. We need some terminology to describe this decomposition more precisely. A bounded sequence  $\{a_n\}_{n \in \mathbb{Z}}$  tends to zero in uniform density if

$$\lim_{N \rightarrow \infty} \sup_{M \in \mathbb{Z}} \frac{1}{N} \sum_{n=M}^{M+N-1} |a_n| = 0.$$

If  $k \geq 1$  is an integer, the sequence  $\{x_n\}$  is said to be a *basic  $k$ -step nilsequence* if there exists some  $k$ -step nilmanifold  $X = G/\Gamma$ , a continuous real valued function  $\phi$  on  $X$ ,  $a \in G$  and  $e \in X$  such that  $x_n = \phi(a^n \cdot e)$  for all  $n \in \mathbb{N}$ . A  *$k$ -step nilsequence* is a uniform limit of basic  $k$ -step nilsequences. The general decomposition result is:

**Theorem 5.5** (Bergelson, Host, and Kra [3]). *Assume that  $(X, \mathcal{X}, \mu, T)$  is an ergodic measure preserving system,  $k \geq 1$  is an integer, and  $f \in L^\infty(\mu)$ . The multicorrelation sequence (8) is the sum of a sequence tending to zero in uniform density and a  $k$ -step nilsequence.*

By subtracting a sequence of integers that tends to 0 in uniform density, the sequences in Theorem 5.3 have the same behavior as the associated nilsequences (of lengths 3 and 4), and the problem reduces to studying lower bounds for the associated nilsequences. The factors constructed in [36] are used to understand the structure of these nilsequences and a key ingredient comes from the explicit formula for the average (1) given in Ziegler [56] (an alternate proof is given in [3]).

In [16], we prove a similar multicorrelation result for independent polynomials. In this case, the nilsequence takes on a simple form, as it is induced by a unipotent affine transformation.

**5.2. Combinatorial Implications.** Via a small modification of Furstenberg's Correspondence Principle, each of these results translates to a combinatorial result. The *upper Banach density*  $d^*(E)$  of a set  $E \subseteq \mathbb{Z}$  is defined by

$$d^*(E) = \lim_{N \rightarrow +\infty} \sup_{M \in \mathbb{Z}} \frac{1}{N} |E \cap [M, M + N - 1]| .$$

Let  $\varepsilon > 0$ ,  $E \subseteq \mathbb{Z}$  have positive upper Banach density, and consider the set

$$\{n \in \mathbb{Z} : d^*(E \cap (E + p_1(n)) \cap \dots \cap (E + p_k(n))) \geq d^*(E)^{k+1} - \varepsilon\} . \quad (9)$$

For  $k = 2$  or  $k = 3$  and  $p_j(n) = jn$  for  $j = 1, 2, 3$ , this set is syndetic, while for  $k \geq 4$  and  $p_j(n) = jn$  for  $j = 1, 2, \dots, k$ , there exists a set of integers  $E$  with positive upper Banach density such that the set in (9) is empty. On the other hand, in [16] we show that for all integers  $k \geq 1$ , if  $p_1, p_2, \dots, p_k : \mathbb{Z} \rightarrow \mathbb{Z}$  are rationally independent polynomials with  $p_i(0) = 0$  for  $i = 1, 2, \dots, k$ , then the set in (9) is always syndetic.

**Question 5.6.** *If  $\varepsilon > 0$ ,  $E \subseteq \mathbb{Z}$  has positive upper Banach density, for which polynomials  $p_1, p_2, \dots, p_k : \mathbb{Z} \rightarrow \mathbb{Z}$  with  $p_i(0) = 0$  for  $i = 1, 2, \dots, k$ , is the set*

$$\{n \in \mathbb{Z} : d^*(E \cap (E + p_1(n)) \cap (E + p_2(n)) \cap \dots \cap (E + p_k(n))) \geq d^*(E)^{k+1} - \varepsilon\}$$

*syndetic?*

For the polynomials of Theorems 5.2, 5.3, and 5.4 we know the answer and it is sometimes yes and sometimes no. It is reasonable to conjecture that the answer is yes for  $k = 2$  and 3, as we know it holds for two extreme cases: 2 (or 3) rationally independent polynomials and 2 (or 3) linear polynomials. For higher  $k$ , it may be possible to lift the independence condition of Theorem 5.2 under certain circumstances. The natural approach to the problem is via the corresponding statement in ergodic theory. A first step in answering this question is finding a general formula for the multiple polynomial average (6), generalizing the formula for the linear average (1).

## 6. Future directions

**6.1. Convergence along other sequences.** General conditions on sequences of integers under which one can prove a multiple ergodic theorem are

unknown:

**Question 6.1.** *If  $k \geq 1$  is an integer and  $a_1(n), a_2(n), \dots, a_k(n)$  are sequences of integers with  $a_j(n) \rightarrow \infty$  as  $n \rightarrow \infty$  for  $j = 1, 2, \dots, k$ , when does*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{a_1(n)} f_1 \cdot T^{a_2(n)} f_1 \cdot \dots \cdot T^{a_k(n)} f_k$$

*exist in  $L^2(\mu)$  for all measure preserving systems  $(X, \mathcal{X}, \mu, T)$  and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ ?*

For  $k = 1$ , convenient necessary and sufficient conditions are given by the spectral theorem. However for  $k \geq 2$ , there is no such characterization and the proof of all known sequences for multiple convergence (including arithmetic progressions, polynomials, and the primes) rely in some manner on a use of the van der Corput Lemma. Finding alternate proofs not relying on the van der Corput Lemma is a first step in describing choices for the sequences  $a_j(n)$ ; a full characterization would probably require some sort of higher order spectral theorem.

Another natural question is the convergence of random multiple ergodic averages. Let  $(\Omega, \mathcal{B}, P)$  be a probability space and let  $\{Y_n\}_{n \in \mathbb{N}}$  be a sequence of independent random variables taking on values 0 and 1. Given  $\omega \in \Omega$ , let  $E = E(\omega) = \{n \in \mathbb{N} : Y_n(\omega) = 1\}$ . Ordering  $E$  by size, we have defined a *random sequence*  $\{a(n) = a(n, \omega)\}$  of natural numbers.

**Question 6.2.** *Assume that  $k \geq 1$  is an integer and that  $a(n)$  is a random sequence of natural numbers generated by a sequence of independent random variables on some probability space  $(\Omega, \mathcal{B}, P)$ . When does*

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} T^{a(n)} f_1 \cdot T^{2a(n)} f_2 \cdot \dots \cdot T^{ka(n)} f_k$$

*exist in  $L^2(\mu)$  for all measure preserving systems  $(X, \mathcal{X}, \mu, T)$  and  $f_1, f_2, \dots, f_k \in L^\infty(\mu)$ ?*

For  $k = 1$ , Bourgain [6] showed that for a random nonlacunary sequence, meaning a sequence where  $P(X_n(\omega) = 1) = p_n$  satisfies  $\lim_{n \rightarrow \infty} np_n = \infty$ , which is also decreasing, one has convergence in  $L^2(\mu)$ . For  $k \geq 2$ , convergence for  $k = 1$  is of course a necessary condition, but it is not known if this is sufficient.

**6.2. Connections with additive combinatorics.** Since Furstenberg's proof of Szemerédi's Theorem, there has been a long and fruitful interaction between additive combinatorics and ergodic theory, with results and techniques passing from one field to the other. A major challenge remains: understand the mathematics behind the deep analogies between the two fields. The nilsystems that arise in the structural analysis of measure preserving systems should have some sort of combinatorial analog:



**Question 6.3.** *What is the combinatorial analog for the Structure Theorem (Theorem 3.2)?*

The uniformity norms on  $\mathbb{Z}/N\mathbb{Z}$  (used in Gowers's [26] proof of Szemerédi's Theorem and in Green and Tao's [28] proof that the primes contain arbitrarily long progressions) play a role similar to the role that the seminorms described in Section 3 play in proving convergence of the multiple ergodic average along arithmetic progressions in [36]. A partial answer to this question is given by Green and Tao in [29], in which they show that generalized quadratic functions control the third uniformity norm, analogous to the way that 2-step nilsystems control the third seminorm. These generalized quadratic functions are controlled by 2-step nilsequences, and this gives a partial understanding of the combinatorial objects. It should be interesting and useful to obtain a more complete understanding of the precise nature of the link between these generalized quadratic functions and 2-step nilsequences, with a description in the finite setting of  $\mathbb{Z}/N\mathbb{Z}$  rather than in  $\mathbb{Z}$ . For longer progressions, even partial results are not known.

It is not clear if one can use ergodic theory to prove statements about the primes, as the primes have zero density and Furstenberg's Correspondence Principle only applies for sets of positive upper density. However, even without a version of the Correspondence Principle that applies to zero density subsets, ergodic theory and especially its techniques has and will be further used to understand the finer structure of the primes. In analogy with multiple ergodic averages along polynomial sequences (and the use of seminorms), one may hope to combine techniques of additive combinatorics and ergodic theory to show, for example, that for all integers  $k > 1$ , there exist infinitely many pairs  $(p, n)$  of integers with  $p, n \geq 1$  such that  $p, p + n, p + n^2, \dots, p + n^k$  consists only of prime numbers.

## References

- [1] V. Bergelson. Weakly mixing PET. *Erg. Th. & Dyn. Sys.* **7** (1987), 337–349.
- [2] V. Bergelson. Ergodic Ramsey theory an update. *Ergodic Theory of  $\mathbb{Z}^d$ -actions* (Eds.: M. Pollicott, K. Schmidt). Cambridge University Press, Cambridge (1996), 1–61.
- [3] V. Bergelson, B. Host and B. Kra, with an Appendix by I. Ruzsa. Multiple recurrence and nilsequences. *Inventiones Math.* **160** (2005), 261–303.
- [4] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.* **9** (1996), 725–753.
- [5] V. Bergelson and R. McCutcheon. An Ergodic IP Polynomial Szemerédi Theorem. *Mem. Amer. Math. Soc.* **146**, 2000.
- [6] J. Bourgain. On the maximal ergodic theorem for certain subsets of the positive integers. *Isr. J. Math.* **61** (1988), 39–72.
- [7] J. Bourgain. An approach to pointwise ergodic theorems. *Geometric aspects of functional analysis (1986/87)*, Lecture Notes in Math., 1317, Springer, Berlin (1988), 204–223.

- [8] J. Bourgain. Pointwise ergodic theorems for arithmetic sets. *Publ. I.H.E.S.* **69** (1989), 5–45.
- [9] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France* **112** (1984), 143–175.
- [10] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *Publications de l'Institut de Recherche de Mathématiques de Rennes, Probabilités* 1987.
- [11] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris Série I* **306** (1988), 491–493.
- [12] P. Erdős and P. Turán. On some sequences of integers. *J. London Math. Soc.* **11** (1936), 261–264.
- [13] N. Frantzikinakis, B. Host, and B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. Preprint, 2005.
- [14] N. Frantzikinakis and B. Kra. Polynomial averages converge to the product of the integrals. *Isr. J. Math.* **148** (2005), 267–276.
- [15] N. Frantzikinakis and B. Kra. Convergence of multiple ergodic averages for some commuting transformations. *Erg. Th. & Dyn. Sys.* **25** (2005), 799–809.
- [16] N. Frantzikinakis and B. Kra. Ergodic averages for independent polynomials and applications. Preprint, 2005.
- [17] H. Furstenberg. Strict ergodicity and transformations of the torus. *Amer. J. Math.* **83** (1961), 573–601.
- [18] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* **31** (1977), 204–256.
- [19] H. Furstenberg. Poincaré recurrence and number theory. *Bull. Amer. Math. Soc.* **5** (1981), 211–234.
- [20] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, Princeton, New Jersey, 1981.
- [21] H. Furstenberg. Nonconventional ergodic averages. *Proc. Sympos. Pure Math.* **50** (1990), 43–56.
- [22] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformation. *J. d'Analyse Math.* **34** (1979), 275–291.
- [23] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for IP-systems and combinatorial theory. *J. d'Analyse Math.* **45** (1985), 117–268.
- [24] H. Furstenberg and Y. Katznelson. A density version of the Hales-Jewett theorem. *J. d'Analyse Math.* **57** (1991), 64–119.
- [25] H. Furstenberg and B. Weiss. A mean ergodic theorem for  $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$ . *Convergence in Ergodic Theory and Probability* (Eds.: Bergelson, March, Rosenblatt). Walter de Gruyter & Co, Berlin, New York (1996), 193–227
- [26] T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
- [27] B. Green. Generalising the Hardy-Littlewood method for primes. *Proceedings of ICM 2006, Madrid*.

- [28] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. To appear, *Annals of Math.*
- [29] B. Green and T. Tao. An inverse theorem for the Gowers  $U^3$  norm. Preprint, 2005.
- [30] B. Green and T. Tao. Quadratic uniformity of the Möbius function. Preprint, 2005.
- [31] B. Green and T. Tao. Two linear equations in four prime unknowns. Preprint, 2005.
- [32] B. Host. Convergence of multiple ergodic averages. To appear, Proceedings of school on “Information and Randomness,” Chili.
- [33] B. Host and B. Kra. Convergence of Conze-Lesigne averages. *Erg. Th. & Dyn. Syst.* **21** (2001), 493-509.
- [34] B. Host and B. Kra. An odd Furstenberg-Szemerédi theorem and quasi-affine systems. *J. d'Analyse Math.* **86** (2002), 183–220.
- [35] B. Host and B. Kra. Averaging along cubes. In “Dynamical Systems and Related Topics,” Cambridge University Press, 2004, 123–144.
- [36] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Annals of Math.* **161** (2005), 397–488.
- [37] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.* **149** (2005), 1–19.
- [38] A. Y. Khintchine. Eine Verschärfung des Poincaréschen “Wiederkehrrsatzes.” *Comp. Math.* **1** (1934), 177–179.
- [39] B. Kra. The Green-Tao Theorem on arithmetic progressions in the primes: an ergodic point of view. *Bull. Amer. Math. Soc.* **43** (2006), 3–23.
- [40] L. Kuipers and N. Niederreiter. *Uniform distribution of sequences*. John Wiley & Sons, New York, 1974.
- [41] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Erg. Th. & Dyn. Sys.* (2005), 201-213.
- [42] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.* **146** (2005), 303–316.
- [43] E. Lesigne. Sur une nil-variété, les parties minimales associée à une translation sont uniquement ergodiques. *Erg. Th. & Dyn. Sys.* **11** (1991), 379–391.
- [44] W. Parry. Ergodic properties of affine transformations and flows on nilmanifolds. *Amer. J. Math.* **91** (1969), 757–771.
- [45] K. Roth. Sur quelques ensembles d’entiers. *C. R. Acad. Sci. Paris* **234** (1952), 388–390.
- [46] A. Sárközy. On difference sets of integers I. *Acta Math. Acad. Sci. Hungar.* **31** (1978), 125–149.
- [47] A. Sárközy. On difference sets of integers III. *Acta Math. Acad. Sci. Hungar.* **31** (1978), 355–386.
- [48] E. Szemerédi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89–104.
- [49] E. Szemerédi. On sets of integers containing no  $k$  elements in arithmetic progression. *Acta Arith.* **27** (1975), 199–245.

- [50] T. Tao. A quantitative ergodic theory proof of Szemerédi's theorem. Preprint, 2004.
- [51] T. Tao. Obstructions to uniformity, and arithmetic patterns in the primes. Preprint, 2005.
- [52] T. Tao. The dichotomy between structure and randomness, arithmetic progressions, and the primes. *Proceedings of ICM 2006, Madrid*.
- [53] B. L. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Arch. Wisk.* **15** (1927), 212–216.
- [54] M. Wierdl. Almost everywhere convergence and recurrence along subsequences in ergodic theory. PhD Thesis, Ohio State University, 1989.
- [55] M. Wierdl. Pointwise ergodic theorem along the prime numbers. *Israel J. Math.*, **64** (1988), 315–336.
- [56] T. Ziegler. A non-conventional ergodic theorem for a nilsystem. *Erg. The. Dyn. Sys.* **25** (2005), 1357–1370.
- [57] T. Ziegler. Universal Characteristic Factors and Furstenberg Averages. To appear, *J. Amer. Math. Soc.*

Department of Mathematics, Northwestern University, 2033 Sheridan Road,  
Evanston, IL 60208-2730, USA  
E-mail: kra@math.northwestern.edu