BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 58, Number 2, April 2021, Pages 295–303
https://doi.org/10.1090/bull/1731
Article electronically published on February 18, 2021

Nilpotent structures in ergodic theory, by Bernard Host and Bryna Kra, Mathematical Surveys and Monographs, Vol. 236, American Mathematical Society, 2018, x+427 pp., ISBN 978-1-4704-4780-9

1. Some history

Ergodic theory is a broad subject that studies properties of measure preserving systems given by quadruples $(X, \mathcal{X}, \mu, T)$, where $(X, \mathcal{X}, \mu)$ is a measure space with $\mu(X) = 1$ and $T\colon X \to X$ is a measurable and measure preserving transformation. At the very early stages of the theory, a central problem was to understand statistical properties of the orbits $(T^n x)_{n\in\mathbb{N}}$ for typical points $x \in X$, where T^n denotes the composition $T \circ \cdots \circ T$. A satisfactory qualitative answer came in the early 1930s with the results of Birkhoff and von Neumann, who showed that for ergodic systems these orbits are equidistributed in the space X for almost every $x \in X$. A functional version of this result states that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x) = \int f \, d\mu$$

for every $f \in L^\infty(\mu)$, with the equality holding almost everywhere (Birkhoff) and in $L^2(\mu)$ (von Neumann). The ergodic theorem had an enormous number of applications in a variety of fields, including statistical mechanics, functional and harmonic analysis, probability theory, geometry, combinatorics, and number theory.

A related problem, the importance of which surfaced four decades later, is the study of distributional properties of sequences of the form

$$(T^n x, T^{2n} x, \ldots, T^{\ell n} x), \quad n \in \mathbb{N}.$$

It is a deep and unexpected result that this problem, and similar ones involving other choices of iterates, are intimately linked with various nilpotent structures that lurk in the background of abstract measure preserving systems, and that these structures control the distributional properties of such sequences in a manner that will be made precise shortly. The main objective of the remarkable book of Host and Kra is to unearth these nilpotent structures and in the process introduce a variety of concepts and tools that have transformed parts of ergodic theory and have led to fruitful interactions with combinatorics and number theory.

These interactions started in the mid-1970s, when Furstenberg observed that structural properties of positive density sets of integers can be reformulated as multiple recurrence properties of measure preserving systems; this is the context of the so-called *Furstenberg correspondence principle*. Consider for example the celebrated theorem of Szemerédi [22], which asserts that any subset of the integers with positive upper density contains arbitrarily long arithmetic progressions. Furstenberg noticed that this result is equivalent to the following multiple recurrence property: if $(X, \mathcal{X}, \mu, T)$ is a measure preserving system and A has positive measure, then for every $\ell \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that

$$\mu(A \cap T^{-n} A \cap \cdots \cap T^{-\ell n} A) > 0. \tag{1}$$

For $\ell = 1$ this is the classical theorem of Poincaré that can be proved elementarily in a few lines, and more refined versions follow from the von Neumann ergodic theorem. But already for $\ell = 2$, the result is far less trivial and the complexity of the proof increases dramatically when $\ell \geq 3$. Furstenberg proved this result in his seminal article [8] ([9, 11] contain useful variants of this approach) using a novel structural result for measure preserving systems that allowed him to infer the asserted multiple recurrence property for general systems from two mutually exclusive classes of systems: the weakly mixing systems, which model randomness; and the distal systems, which model structure. A key ingredient in his approach is the study of the limiting behavior in $L^2(\mu)$ of the following multiple (sometimes also called nonconventional) ergodic averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T^n x)\cdot \ldots \cdot f_\ell(T^{\ell n}x), \tag{2}$$

where $f_1, \ldots, f_\ell \in L^\infty(\mu)$. Although Furstenberg was not able to prove mean convergence for these averages, he managed to show a positivity property that suffices to verify the multiple recurrence result.

Furstenberg's pioneering work set out the general strategy for several other deep combinatorial results that were proved subsequently using ergodic theory. This laid out the foundations for a new area, referred to as ergodic Ramsey theory, that has flourished in subsequent years and remains vibrant. Two major results that were first proved using ergodic theory are the following:

- The multidimensional Szemerédi theorem, proved by Furstenberg and Katznelson in [10], states that for every $d, \ell \in \mathbb{N}$ and vectors $v_1, \ldots, v_\ell \in \mathbb{Z}^d$, every subset of $\mathbb{Z}^d$ with positive upper density contains patterns of the form $m, m + nv_1, \ldots, m + nv_\ell$, for some $m \in \mathbb{Z}^d$ and $n \in \mathbb{N}$.
- The polynomial Szemerédi theorem, proved by Bergelson and Leibman in [4], states that for every $\ell \in \mathbb{N}$ and polynomials $p_1, \ldots, p_\ell \in \mathbb{Z}[t]$ with zero constant terms, every subset of the integers with positive upper density contains patterns of the form $m, m+p_1(n), \ldots, m+p_\ell(n)$ for some $m, n \in \mathbb{N}$.

To deal with such results it was necessary to study more general multiple ergodic averages of the form

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T_1^{p_1(n)}x)\cdot \ldots \cdot f_\ell(T_\ell^{p_\ell(n)}x), \tag{3}$$

where $T_1, \ldots, T_\ell \colon X \to X$ are commuting measure preserving transformations. As in the case of the ergodic theoretic proof of Szemerédi's theorem, the relevant multiple recurrence results were proved without either identifying the limit of these averages or establishing mean convergence.

We come back to the question of mean convergence of the averages (2). The main approach is to find a suitable factor of the given system, that is often identified with a closed subspace of $L^2(\mu)$ and referred to as the *characteristic factor* (a notion coined by Furstenberg and Weiss in [12]), such that the limit behavior of the relevant averages remains unchanged when each function is replaced by its conditional expectation on this factor. If this factor is simple enough, one could then hope to prove mean convergence by explicit computation. For $\ell = 2$ such a factor was identified by Furstenberg in [8], and for ergodic systems it is the closed subspace spanned by all the eigenfunctions of the operator T (often refered to as the Kronecker factor). Moreover, using an old result of Halmos and von Neumann from the 1940s, it follows that the action of T on this factor is isomorphic to a rotation on a compact Abelian group. This can be used to establish mean convergence and to explicitly evaluate the limit of the averages (2) when $\ell = 2$.

But the problem of mean convergence of the averages (2) turned out to be much more challenging when $\ell \geq 3$. In [8] Furstenberg proved that the $(\ell-1)$-step distal factor is a characteristic factor for these averages. One-step distal systems are isomorphic to compact Abelian group rotations with the Haar measure, and an example of a 2-step distal system is given by the transformation

$$T(x,y) = (x+\alpha, y+\rho(x)), \quad x,y \in G, \tag{4}$$

where $(G,+)$ is a compact Abelian group, $\alpha \in G$, $\rho\colon G \to G$ is Borel measurable, and we consider the Haar measure on $G \times G$. While at first glance this seems like a simple class of measure preserving systems, it turns out to be too complicated for explicit computations that could be used to prove mean convergence of the averages (2).

The next significant step came in the mid-1980s, when Conze and Lesigne in [5, 6] proved mean convergence of the averages (2) for $\ell = 3$ for all totally ergodic systems. The key step was to give a convenient description of the corresponding characteristic factors. It is simple to verify that for $\ell = 3$ affine systems, such as those arising from (4) when $\rho(y) = y$, lead to cancellation in the averages (2), thus, such affine subsystems should be accounted for when describing potential characteristic factors. But there are also other, somewhat more exotic, systems that must also be taken into account. One such example is given by the transformation T that acts on $\mathbb{T}^3$ with the Haar measure as

$$T(x,y,z) = (x+\alpha, y+\beta, z+(x+\alpha)[y+\beta] - x[y] - \alpha y), \quad x,y,z \in \mathbb{T}, \tag{5}$$

where $\alpha, \beta \in [0,1)$. A crucial observation, made by Furstenberg and Weiss (and recorded in print several years later in [12]), is that such transformations arise from rotations on the Heisenberg nilmanifold. To see this, define multiplication on $G = \mathbb{R}^3$ by

$$g_1 \cdot g_2 = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + x_1 y_2),$$

where $g_1 = (x_1, y_1, z_1)$ and $g_2 = (x_2, y_2, z_2)$, in which case $(G, \cdot)$ is a 2-step nilpotent group and the discrete subgroup $\Gamma = \mathbb{Z}^3$ is cocompact. If we let $b = (\alpha, \beta, 0)$, where $\alpha, \beta \in \mathbb{R}$, then the system induced by the rotation $x \mapsto bx$ on $X = G/\Gamma$ with the Haar measure $m_{\mathbb{T}^3}$, is measure theoretically isomorphic to the system $(\mathbb{T}^3, m_{\mathbb{T}^3}, T)$, where T is given by (5). More generally, if $(G, \cdot)$ is an ℓ-step nilpotent Lie group and Γ is a discrete cocompact subgroup of G, then the homogeneous space $X = G/\Gamma$ is called an *ℓ-step nilmanifold* and for $b \in G$ the system induced by the rotation $x \mapsto bx$ on X with the Haar measure is called an *ℓ-step nilsystem.* A key result proved in [5, 6] is that if $\ell = 3$, then for totally ergodic systems the averages (2) have characteristic factors that can be approximated by 2-step nilsystems.

The previous advances triggered hopes that characteristic factors for the averages (2) admit similar explicit structural descriptions for arbitrary $\ell \in \mathbb{N}$; namely they can be well approximated by $(\ell-1)$-step nilsystems. It took more than 15 years before this hope was realized in the seminal paper of Host and Kra [18] (with a different proof given subsequently by Ziegler [27]), thus bringing to a close the problem of mean convergence of the averages (2). But more important than the convergence result itself, were the conceptual advances and powerful tools developed in [18] that transformed the field of ergodic Ramsey theory and had far-reaching impact well beyond the field of ergodic theory. In the next few paragraphs we briefly describe some of these key ideas; the reader will find a much more detailed account in the book under review.

A posteriori, for $\ell = 3$ it is not very surprising that 2-step nilsystems play an important role in the description of the characteristic factors of the averages (2). Roughly speaking, for such systems simple algebraic manipulations show that the values of $T^{3n}x$ are constrained by the values $x, T^n x, T^{2n}x$, and such pointwise constraints induce nontrivial cancellation in the averages (2) for appropriate choices of functions. A similar phenomenon occurs when one studies the *cubic configurations*,

$$(x, T^m x, T^n x, T^r x, T^{m+n}x, T^{m+r}x, T^{n+r}x, T^{m+n+r}x); \tag{6}$$

namely, it can be shown that for 2-step nilsystems the last coordinate is constrained by the first seven coordinates. Host and Kra went a step further by showing that a measure theoretic variant of this observation can be phrased rigorously so that it becomes a characterization of systems that can be well approximated by 2-step nilsystems. Moreover, similar characterizations hold for ℓ-step nilsystems and involve cubic configurations with $2^{\ell+1}$ terms. This deep and surprising fact has opened a new window in our understanding of the limiting behavior of the averages (2). The link comes from certain seminorms that these cubic patterns define and control the limiting behavior of these averages. We introduce these seminorms next.

For a given ergodic system Host and Kra inductively defined a family of seminorms $\|\!|\cdot|\!\|_\ell$ on $L^\infty(\mu)$ as follows: for $\ell = 1$, we let $\|\!|f|\!\|_1 = |\int f\, d\mu|$, and for $\ell \in \mathbb{N}$ we let $\|\!|f|\!\|_{\ell+1}^{2^{\ell+1}} = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \|\!|T^n f \cdot \overline{f}|\!\|_\ell^{2^\ell}$. For instance, we have

$$\|\!|f|\!\|_2^4 = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \lim_{M\to\infty} \frac{1}{M}\sum_{m=1}^{M} \int f \cdot T^m\overline{f} \cdot T^n\overline{f} \cdot T^{m+n} f\, d\mu. \tag{7}$$

Similarly, $\|\!|f|\!\|_3^8$ is defined via a similar formula that involves the cubic configurations in (6), and so on for $\|\!|f|\!\|_\ell^{2^\ell}$. The motivation behind these seminorms is that, via successive uses of the van der Corput lemma, they can be used to bound the $L^2(\mu)$ norm of several multiple ergodic averages; in particular the averages (2). The seminorm $\|\!|\cdot|\!\|_\ell$ naturally defines a *cubic structure* of order ℓ, meaning a measure $\mu^{[[\ell]]}$ on the space X^{2^ℓ} that enjoys a variety of symmetries, and from this measure, structured factors $\mathcal{Z}_{\ell-1}$ are defined such that for every $f \in L^\infty(\mu)$ one has

$$\mathbb{E}(f|\mathcal{Z}_{\ell-1}) = 0 \Longleftrightarrow \|\!|f|\!\|_\ell = 0.$$

Since the seminorms $\|\!|\cdot|\!\|_\ell$ control the averages (2), the factor $\mathcal{Z}_{\ell-1}$ is characteristic for these averages.

The bulk of the paper [18] is devoted to an in-depth structural analysis of the factors $\mathcal{Z}_\ell$. It is not hard to verify that for $\ell = 1$ the factor $\mathcal{Z}_1$ coincides with the Kronecker factor of the system, but matters become more complicated when $\ell \geq 2$. Using the cubic structures alluded to above, a group of measure preserving transformations of $(X, \mathcal{X}, \mu)$ is defined and it is shown that for systems of order ℓ (that is, those systems that satisfy $\mathcal{X} = \mathcal{Z}_\ell$), this group is ℓ-step nilpotent. The hardest and most technical step is to show that for a sufficiently large class of systems this group is a Lie group and acts transitively on the space; the argument depends on a complicated analysis of properties of cocycles of type ℓ that naturally arise when one studies systems of order ℓ. From this fact one deduces that the factor $\mathcal{Z}_\ell$ can be well approximated by ℓ-step nilsystems.

The Host–Kra seminorms and the related structural result immediately became essential tools in the area and were used to control and understand a variety of other multiple ergodic averages. This initiated a stream of applications, and the first one was obtained by Host and Kra in [19] and Leibman in [21], who proved mean convergence for multiple ergodic averages with polynomial iterates. Using the polynomial exhaustion technique of Bergelson [2], they showed that if the polynomials $p_1, \dots, p_\ell \in \mathbb{Z}[t]$ are nonconstant and have pairwise nonconstant differences, then there exist $k \in \mathbb{N}$ and $C > 0$ such that for all systems (X, μ, T) and functions $f_1, \dots, f_\ell \in L^\infty(\mu)$ we have

$$\limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=1}^{N} T^{p_1(n)} f_1 \cdot \ldots \cdot T^{p_\ell(n)} f_\ell \right\|_{L^2(\mu)} \leq C \min_{i=1,\dots,\ell} \| f_i \|_k. \tag{8}$$

An immediate consequence of this and the Host–Kra structure theory, is that characteristic factors for these averages can be well approximated by nilsystems. With this in mind, mean convergence of the averages in (8) can be deduced from known equidistribution results of polynomial sequences on nilmanifolds due to Leibman [20]. Using a similar approach, it has also been possible to identify the limit for special classes of polynomial iterates, and this has led to a variety of more refined multiple recurrence and combinatorial results than those covered by the theorem of Szemerédi and its polynomial version. We mention one such application next.

It is natural to inquire about specific lower bounds in Furstenberg's multiple recurrence result (1). Examples of weakly mixing systems show that such lower bounds cannot be greater than $(\mu(A))^{\ell+1}$, and the question is whether for all ergodic systems $(X, \mathcal{X}, \mu, T)$, measurable sets A, and $\varepsilon > 0$, we have for some $n \in \mathbb{N}$ that

$$\mu(A \cap T^{-n} A \cap \cdots \cap T^{-\ell n} A) \geq (\mu(A))^{\ell+1} - \varepsilon.$$

Bergelson, Host, and Kra in [3] showed that this holds when $\ell = 1, 2, 3$. But surprisingly enough, when $\ell \geq 4$ for some systems and sets, the estimate fails for all $n \in \mathbb{N}$, and in fact no fixed power of $\mu(A)$ can be used as a lower bound. The proof of the lower bounds makes essential use of the theory of characteristic factors of Host and Kra and proceeds by carefully analyzing the limiting behavior of certain weighted variants of the averages (2) using equidistribution results on nilmanifolds. Moreover, in the same article the authors introduced a new class of sequences, called *nilsequences*, and used it to prove the following decomposition result for multiple correlation sequences: for every ergodic system $(X, \mathcal{X}, \mu, T)$ and functions $f_0, f_1, \dots, f_\ell \in L^\infty(\mu)$, we have

$$\int f_0 \cdot T^n f_1 \cdot \ldots \cdot T^{\ell n} f_\ell \, d\mu = \psi(n) + e(n), \tag{9}$$

where $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |e(n)| = 0$ and the sequence ψ is a uniform limit of ℓ-step nilsequences, meaning, sequences of the form $(\Psi(b^n x))_{n\in\mathbb{N}}$, where Ψ is a continuous function on a nilmanifold $X = G/\Gamma$ and $b \in G$. In subsequent years, nilsequences were used as a substitute for linear exponential sequences in the rapidly evolving area of *higher order Fourier analysis*, in order to analyze the asymptotic behavior of a variety of multilinear expressions that naturally occur in additive combinatorics and number theory; these are problems where traditional Fourier analytic tools have failed to give us a proper understanding.

After successfully handling mean convergence problems for the averages (3) when all the transformations are equal, the next goal was to deal with analogous problems for arbitrary commuting measure preserving transformations $T_1, \ldots, T_\ell \colon X \to X$, starting with the averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(T_1^n x) \cdot \ldots \cdot f_\ell(T_\ell^n x). \tag{10}$$

A major block in handling these averages is that the Host–Kra seminorms with respect to the individual transformations do not control their $L^2(\mu)$ norm, and as a consequence, the Host–Kra theory of characteristic factors is not applicable. When $\ell = 2$, mean convergence of the averages (10) was established by Conze and Lesigne [5], but their methods did not generalize to $\ell = 3$ and progress stalled for more than two decades. The next breakthrough came when Tao established in [23] mean convergence for general $\ell \in \mathbb{N}$ by recasting the problem in finitary terms and using a delicate induction that allowed him to infer convergence properties for ℓ transformations from related convergence properties for $\ell - 1$ transformations. In another breakthrough, this approach was extended by Walsh in [26] to cover general polynomial iterates, thus finally establishing mean convergence for the averages (3). A drawback of the methods of Tao and Walsh is that they do not give explicit information for the limit or the related characteristic factors, and an in-depth understanding of the limiting behavior of the averages (3) is still lacking. In the case of the averages (10), this has been partially rectified by works of Austin [1] and Host [17], in which alternative ergodic proofs of mean convergence were given that provide some information for the limit.

The Host–Kra theory of characteristic factors gave us access to several other problems that were previously considered intractable. A variety of deep multiple recurrence and convergence results were obtained for iterates arising from smooth functions, random sequences, and the sequence of prime numbers, setting the stage for a thriving area of research. In a different direction, the conceptual advances made and the ergodic tools developed through the years had amazing and often unexpected impact in some notoriously difficult problems in combinatorics and number theory. Some examples outside the area of ergodic Ramsey theory include-the following.

- The proof of Green and Tao [14] that the primes contain arbitrarily long arithmetic progressions drew crucial insight from the notion of characteristic factors and the ergodic theoretic proof of Szemerédi's theorem. The subsequent article of Tao and Ziegler [25], which established arbitrarily long polynomial progressions in the primes, used the polynomial Szemerédi theorem as a black box whose only known proof to this day uses ergodic theory.
- The Gowers norms in $\mathbb{Z}_N$ were introduced in [13] and play a central role in additive combinatorics. An inverse theorem for these norms was established by Green, Tao, and Ziegler [16] in which finitary analogues of nilsequences were used to classify the sequences that obstruct Gowers uniformity. This result, in addition to others, was used by Green and Tao [15] in order to settle several cases of the Hardy–Littlewood prime-tuples conjecture and prove Gowers uniformity of a modification of the von Mangoldt function. The proofs of all these results crucially used the Host–Kra cubic structures from [18] as well as quantitative variants of ergodic equidistribution results on nilmanifolds.

- In a further twist, very recently, ergodic theory played an important role in the study of statistical properties of bounded multiplicative functions, the most prominent examples being the Möbius and the Liouville functions. Tao and Teräväinen [24] verified a logarithmically averaged variant of the Chowla conjecture regarding correlations of odd order of the Liouville function, and their argument crucially used a variant of the decomposition result for multiple correlation sequences given in (9). Frantzikinakis and Host [7] verified the logarithmically averaged Möbius disjointness conjecture of Sarnak for ergodic weights, and their argument used a variety of limit formulas for multiple ergodic averages that were previously proved using the Host–Kra theory of characteristic factors.

2. The book

The results presented above give only a taste of the topics covered in this wonderful book. Anyone interested in a deeper understanding of the Host–Kra theory of characteristic factors and its applications in ergodic theory, combinatorics, and number theory, will find this book an invaluable asset.

The first part of the book covers background material in ergodic theory and topological dynamics that is tailored for the needs of this book and is in some cases hard to find in print. This is a valuable source for the nonexpert.

The second part introduces some of the key elements of the Host–Kra structure theorem. The cubic structures are explained in progressively more complex settings, namely, algebraic, topological, and finally ergodic settings. The ergodic cubic structures are then used to naturally introduce the ergodic seminorms $|\!|\!| \cdot |\!|\!|_\ell$ and the structured factors $\mathcal{Z}_\ell$. In the process, dual functions and their properties are studied.

The third part covers basic facts about nilmanifolds, nilsystems and their factors, polynomial sequences on nilmanifolds, and related equidistribution results. Several of the concepts and tools presented here are spread out in various places in the literature, and the book masterfully organizes these topics and presents them in a coherent way with meticulous attention to details that are in some cases missing from the original sources.

The first half of the fourth part contains a functional form of the main structure theorem and states variants of the structure theorem in topological dynamics and in the finitary setting of $\mathbb{Z}_N$. The second half contains the heart of the proof of the Host–Kra structure theorem. This is the hardest and most technical part of the book. The authors follow the proof in [18] but with simplifications at various parts. The argument is first presented in a case that offers significant technical simplifications, allowing the reader to get a broad overview of the proof strategy in a friendlier setup before going through the details of the more convoluted argument needed for the general case.

The fifth and last part of the book contains applications of the material presented in the previous chapters, most of them stemming from research papers that were written in the last fifteen years.

- The method of characteristic factors is used to prove mean convergence of multiple ergodic averages with linear iterates, and then, after carefully explaining the polynomial exhaustion technique, the case of polynomial iterates is handled. Furthermore, for linearly independent iterates, using qualitative equidistribution results on nilmanifolds, the limit function is identified.

- Uniformity seminorms for bounded sequences are defined and their relation with the Gowers norms in $\mathbb{Z}_N$ and the ergodic seminorms $\|\cdot\|_\ell$ is studied. As an application, a pointwise convergence result for cubic averages is given.
- An inverse theorem for uniformity seminorms of bounded sequences and a related decomposition result is proved. As a consequence, convergence criteria for multiple ergodic averages are given.
- A variant of the decomposition result (9) that covers ℓ commuting transformations is proved, and applications to mean convergence results of weighted multiple ergodic averages involving commuting transformations are given.
- Finally, multiple recurrence and convergence results for sequences related to the prime numbers are deduced from known results using the Gowers uniformity of a modification of the von Mangoldt function.

Each chapter is followed by a short section that is filled with bibliographical and historical notes. This essential input should be appreciated by anyone who wants to learn more on the subject and study the original sources.

This book was a long time coming! Prior to its publication, basic results in the field were scattered in many research articles and books, making it difficult for newcomers to get up to speed with current research. The book does a wonderful job in rectifying this situation by skillfully providing a synthesis of background material along with foundational and more recent contributions in the field. It will be appreciated by beginners and experts alike and will undoubtedly become a key resource for years to come for anyone that wants to explore or be actively involved with this beautiful area of mathematics.

References

[1] T. Austin, *On the norm convergence of non-conventional ergodic averages*, Ergodic Theory Dynam. Systems **30** (2010), no. 2, 321–338, DOI 10.1017/S014338570900011X. MR2599882

[2] V. Bergelson, *Weakly mixing PET*, Ergodic Theory Dynam. Systems **7** (1987), no. 3, 337–349, DOI 10.1017/S0143385700004090. MR912373

[3] V. Bergelson, B. Host, and B. Kra, *Multiple recurrence and nilsequences*, Invent. Math. **160** (2005), no. 2, 261–303, DOI 10.1007/s00222-004-0428-6. MR2138068

[4] 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, DOI 10.1090/S0894-0347-96-00194-4. MR1325795

[5] J.-P. Conze and E. Lesigne, *Théorèmes ergodiques pour des mesures diagonales* (French, with English summary), Bull. Soc. Math. France **112** (1984), no. 2, 143–175. MR788966

[6] J.-P. Conze and E. Lesigne, *Sur un théorème ergodique pour des mesures diagonales* (French), Probabilités, Publ. Inst. Rech. Math. Rennes, vol. 1987, Univ. Rennes I, Rennes, 1988, pp. 1–31. MR989141

[7] N. Frantzikinakis and B. Host, *The logarithmic Sarnak conjecture for ergodic weights*, Ann. of Math. (2) **187** (2018), no. 3, 869–931, DOI 10.4007/annals.2018.187.3.6. MR3779960

[8] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256, DOI 10.1007/BF02813304. MR498471

[9] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981. MR603625

[10] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979), DOI 10.1007/BF02790016. MR531279

[11] 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, DOI 10.1090/S0273-0979-1982-15052-2. MR670131

[12] 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. MR1412607

[13] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588, DOI 10.1007/s00039-001-0332-9. MR1844079

[14] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547, DOI 10.4007/annals.2008.167.481. MR2415379

[15] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850, DOI 10.4007/annals.2010.171.1753. MR2680398

[16] B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers* $U^{s+1}[N]$*-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372, DOI 10.4007/annals.2012.176.2.11. MR2950773

[17] B. Host, *Ergodic seminorms for commuting transformations and applications*, Studia Math. **195** (2009), no. 1, 31–49, DOI 10.4064/sm195-1-3. MR2539560

[18] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488, DOI 10.4007/annals.2005.161.397. MR2150389

[19] B. Host and B. Kra, *Convergence of polynomial ergodic averages*, Israel J. Math. **149** (2005), 1–19, DOI 10.1007/BF02772534. MR2191208

[20] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 201–213, DOI 10.1017/S0143385704000215. MR2122919

[21] A. Leibman, *Convergence of multiple ergodic averages along polynomials of several variables*, Israel J. Math. **146** (2005), 303–315, DOI 10.1007/BF02773538. MR2151605

[22] E. Szemerédi, *On sets of integers containing no* k *elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, DOI 10.4064/aa-27-1-199-245. MR369312

[23] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems **28** (2008), no. 2, 657–688, DOI 10.1017/S0143385708000011. MR2408398

[24] T. Tao and J. Teräväinen, *The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures*, Duke Math. J. **168** (2019), no. 11, 1977–2027, DOI 10.1215/00127094-2019-0002. MR3992031

[25] T. Tao and T. Ziegler, *The primes contain arbitrarily long polynomial progressions*, Acta Math. **201** (2008), no. 2, 213–305, DOI 10.1007/s11511-008-0032-5. MR2461509

[26] M. N. Walsh, *Norm convergence of nilpotent ergodic averages*, Ann. of Math. (2) **175** (2012), no. 3, 1667–1688, DOI 10.4007/annals.2012.175.3.15. MR2912715

[27] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), no. 1, 53–97, DOI 10.1090/S0894-0347-06-00532-7. MR2257397

Nikos Frantzikinakis
Department of Mathematics and Applied Mathematics
University of Crete
Voutes University Campus
Heraklion 71003, Greece
Email address: `frantzikinakis@gmail.com`