Proceedings of the International Congress of Mathematicians
Hyderabad, India, 2010

Ergodic Structures and Non-conventional Ergodic Theorems

Hillel Furstenberg*

Mathematics Subject Classification (2000). Primary 28D99; Secondary 11B75.

Keywords. arithmetic progression, characteristic factors, distal, dynamical system, ergodic, Gowers seminorm, measure, nilpotent, nilsystem, recurrence, Szemerédi

1. Introduction

Ergodic theory treats measure preserving dynamical systems. We recall: a quadruple $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ is a *measure preserving system* (m.p.s.) if $(X, \mathcal{B}, \mu)$ is a measure space with $\mu(X) < \infty$, and $T : X \to X$ is a measurable, measure preserving map. That is to say, for $B \in \mathcal{B}$, $T^{-1}(B) \in \mathcal{B}$ and $\mu(T^{-1}B) = \mu(B)$. The dynamical character of such a system appears when the transformation T is iterated so that $T^n x$ describes the state at the time n when the initial state is x. There are two theorems at the foundation of classical ergodic theory:

Poincaré's Recurrence Theorem. *If $(X, \mathcal{B}, \mu, T)$ is an m.p.s. and $A \in \mathcal{B}$ with $\mu(A) > 0$, then for some $n = 1, 2, 3, \ldots,$ $\mu(A \cap T^{-n}A) > 0$.*

It is not hard to deduce from this that in fact almost every point of A returns to A infinitely often.

Birkhoff's Pointwise Ergodic Theorem. *If $(X, \mathcal{B}, \mu, T)$ is an m.p.s. and $f \in L^1(X, \mathcal{B}, \mu)$, then for almost every $x \in X$, the limit as $N \to \infty$ of ergodic averages*

$$A_N(f, x) = \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) \tag{1.1}$$

exists, and the limit function $\bar{f}(x) = \lim A_N(f, x)$ satisfies $\bar{f}(Tx) = \bar{f}(x)$ and $\int \bar{f} d\mu = \int f d\mu$.

*Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Safra Campus Giv'at Ram, Manchester House 202, Jerusalem, 91904, Israel. E-mail: harry@math.huji.ac.il

Our focus here will be on "non-conventional ergodic averages",

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^{p_1(n)}x)f_2(T^{p_2(n)}x)\cdots f_k(T^{p_k(n)}x) \tag{1.2}$$

and their limits, in which several functions are involved simultaneously, and these are evaluated on the orbit of a point x at polynomial times $p_1(n), p_2(n), \ldots, p_k(n)$ respectively. The polynomial character of the times has no special dynamical significance, but is meaningful for diophantine applications.

The diophantine significance of expressions of the form (1.2) showed up first in the ergodic theoretic proof of Szemerédi's theorem. This theorem states that if a set E of integers has positive upper density, then it contains arbitrarily long arithmetic progressions. It can be shown — via a correspondence principle ([EW10], [TT09, p. 163]) — that this is equivalent to the following extension of the Poincaré recurrence theorem:

The Multiple Recurrence Theorem. *For any m.p.s.* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$*, if* $A \in \mathcal{B}$ *with* $\mu(A) > 0$ *and* $k \in \mathbb{N}$*, then for some* n

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{1.3}$$

This recurrence result was first proved by a consideration of averages. Namely, ([FU77]) one showed that for any k,

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > 0. \tag{1.4}$$

This raises the question as to whether the limit in question exists, and this will be the case, if, setting $f(x) = 1_A(x)$ the "non-conventional average"

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)f(T^{2n}x)\cdots f(T^{kn}x)$$

exists in $L^2(X, \mathcal{B}, \mu)$. This in fact is true but considerably more effort was required to obtain this "mean ergodic theorem" than was needed for (1.4). (See [EW10], [BL96] and [FU81] for a more detailed exposition.)

One is now able to extend the two types of phenomena further to polynomial times, as we'll see. We can talk of a "polynomial mean ergodic theorem" as well as a "polynomial multiple recurrence theorem". The former is of interest in its own right as a legitimate topic in ergodic theory; the latter is of interest also for its diophantine and combinatorial implications. The polynomial mean ergodic theorem is the statement that for any bounded measurable functions $f_1, f_2, \ldots, f_k$ and integer valued polynomials $p_1(n), p_2(n), \ldots, p_k(n)$, the averages in (1.2) converge, as $N \to \infty$, in $L^2(X, \mathcal{B}, \mu)$. It is believed that one also has almost everywhere convergence but this has been proved so far only for some special cases. A polynomial multiple recurrence theorem is the analogue of (1.3) or (1.4) with $n, 2n, \ldots, kn$ replaced by a suitable set of polynomials. Some restrictions on the polynomials $p_j(n)$ have to be made since, e.g., it is easy to construct systems with $\mu(A \cap T^{-(2n+1)}A) = 0$ for certain A for all n.

2. Ergodicity, Factors and the Basic Structure Theorems

A system $(X, \mathcal{B}, \mu, T)$ is *ergodic* if for $A, B \in \mathcal{B}$ with $\mu(A), \mu(B) > 0$ there exists n with $\mu(A \cap T^{-n}B) > 0$. This is equivalent to the condition that if f is measurable and $f \circ T = f$ then f is almost everywhere constant. The ergodic theorem implies in this case that the limit $\bar{f}$ of (1.1) is constant, and the condition $\int \bar{f} d\mu = \int f d\mu$ implies that $\bar{f}(x) = \frac{1}{\mu(X)} \int f d\mu$. We will be assuming throughout that $\mu(X) = 1$, so that for ergodic systems we obtain $A_N(f, x) \to \int f d\mu$ a.e.

For two measure spaces $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{D}, \nu)$, a map $\pi : X \to Y$ is *measurable* if the σ-algebra $\pi^{-1}(\mathcal{D}) \subset \mathcal{B}$ and π is *measure preserving* if for $D \in \mathcal{D}$, $\mu(\pi^{-1}(D)) = \nu(D)$. For $(X, \mathcal{B}, \mu)$ a "Lebesgue space" we have the notion of "decomposition of μ relative to $(Y, \mathcal{D}, \nu)$" and *conditional expectation*. (See [GL03], [FU81] for details). Namely there is an almost everywhere defined map from Y to probability measures on $X, y \to \mu_y$, so that $\mu = \int \mu_y d\nu(y)$, meaning that $\int f d\mu = \int \{\int f d\mu_y\} d\nu(y)$ for $f \in L^1(X, \mathcal{B}, \mu)$. The function $\phi(y) = \int f d\mu_y$ is denoted $E(f|\pi^{-1}(\mathcal{D}))$ (See [DO53]). The lift of the latter function to X, $E(f|\pi^{-1}(\mathcal{D})) \circ \pi$ belongs to $L^1(X, \mathcal{B}, \mu)$ and for $f \in L^2(X, \mathcal{B}, \mu)$, the linear map $f \to E(f|\pi^{-1}\mathcal{D}) \circ \pi$ is the orthogonal projection of $L^2(X, \mathcal{B}, \mu)$ to the subspace $L^2(Y, \mathcal{D}, \nu) \circ \pi$. We will use the notation $E(f|Y)$ interchangeably for the function $E(f|\pi^{-1}(\mathcal{D}))$ on Y and its lift to X.

For two measure preserving systems $(X, \mathcal{B}, \mu, T), (Y, \mathcal{D}, \nu, S)$ we will speak of a measurable, measure preserving map $\pi : X \to Y$ as a *homomorphism* if for a.e. $x \in X$, $S\pi(x) = \pi(Tx)$. It will follow that for almost every y, $T(\mu_y) = \mu_{Sy}$ and that $E(f \circ T|Y) = E(f|Y) \circ S$ as functions on Y. When we have a homomorphism of a system $\mathbf{X}$ to a system $\mathbf{Y}$ we speak of the latter as a *factor* of the former and of the former as an *extension* of the latter.

Suppose $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ is a *degenerate* system meaning that $Sy = y$ for each $y \in Y$, and suppose $\pi : \mathbf{X} \to \mathbf{Y}$ is a homomorphism. Then the measures μ_y are T-invariant for a.e. y. We can then form systems $(X, \mathcal{B}, \mu_y, T)$. One now has the *ergodic decomposition theorem*:

Theorem. *For any m.p.s.* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *there is a degenerate factor* $\mathbf{Y} = (Y, \mathcal{S}, \nu, S)$ *for which the systems* $(X, \mathcal{B}, \mu_y, T)$ *are almost all ergodic.*

A consequence of this ergodic decomposition theorem, together with the representation $\mu = \int \mu_y d\nu(y)$, is that the issues we are dealing with, recurrence and convergence of ergodic averages, can be confined to the case of an ergodic system. We proceed to present a structure theorem for ergodic systems. We will describe two types of extensions for ergodic systems and the basic structure theorem for ergodic systems will be the assertion that combining these two forms of extensions one can arrive at any ergodic system starting from the trivial 1-point system.

For a compact metric space M we denote by $Isom(M)$ the compact group of isometries of M. We will say that $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ is an *isometric* extension of $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ if the former can be represented as $X = Y \times M$ for compact metric M with $\mu = \nu \times m_M$ where $m_M \in \mathcal{P}(M)$ is invariant under isometries, and $T(y, u) = (Sy, \rho(y)u)$ where $\rho : Y \to Isom(M)$ is measurable. When $\mathbf{Y}$ is a trivial system and $\mathbf{X}$ an ergodic isometric extension, it can be seen that $X \approx M$ is a compact abelian group and $Tx = ax$ where $a \in M$ generates a dense subgroup of M. We call such a system a *Kronecker system* and denote the action of S additively: $z \to z + \alpha$, and denote the system (M, Borel sets, Haar measure, translation by α) briefly by (M, α).

Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be an ergodic m.p.s. with Kronecker factor (Z, α) and let $\varphi : X \to Z$ define the corresponding homomorphism. Any character $\chi : Z \to S^1$ satisfies $\chi(z + \alpha) = \chi(\alpha)\chi(z)$ and so lifting to X, if $f = \chi \circ \varphi$, $f(Tx) = \chi \circ \varphi(Tx) = \chi(\varphi(x) + \alpha) = \chi(\alpha)\chi \circ \varphi(x) = \chi(\alpha)f(x)$, we obtain an eigenfunction f of the operator $f \to f \circ T$. For ergodic systems all eigenfunctions come about in this way, and indeed, using the group of eigenfunctions of the induced operator $Tf = f \circ T$, we can construct a "universal" Kronecker factor $(\tilde{Z}, \tilde{\alpha})$ of $\mathbf{X}$ such that all Kronecker factors of $\mathbf{X}$ are factors of $(\tilde{Z}, \tilde{\alpha})$. We refer to $(\tilde{Z}, \tilde{\alpha})$ as *the* Kronecker factor of $\mathbf{X}$.

A broader family of systems is obtained by taking successive isometric extensions of previously defined systems. This leads to the notion of a *distal* system: $\mathbf{X}$ is *distal* if it is a member of a (possibly) transfinite tower of systems $\{\mathbf{X}_\eta, \eta \text{ ordinal}\}$ having at its base $\mathbf{X}_0$ the trivial 1-point system, and with $\mathbf{X}_{\eta+1}$ an isometric extension of $\mathbf{X}_\eta$, and for a limit ordinal η, $\mathbf{X}_\eta = \lim_{\xi < \eta} \mathbf{X}_\xi$.

The other type of extension which will appear in our general structure theorem is that of a (relatively) weakly mixing extension, which we abbreviate to WM extension. Recall that a system is (absolutely) weakly mixing if $\mathbf{X} \times \mathbf{X}$ is ergodic. In the relative notion we introduce the "relative product". If $(X_i, \mathcal{B}_i, \mu_i)$, $i = 1, 2$, are two measure spaces, $f_i, i = 1, 2$, two measurable function on these spaces respectively, we denote by $f_1 \otimes f_2$ the function on $X_1 \times X_2$ with $f_1 \otimes f_2(x_1, x_2) = f_1(x_1)f_2(x_2)$. Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are both extensions of a system $\mathbf{Y}$ there will be a unique measure $\tilde{\mu}$ or $X_1 \times X_2$ with $\int f_1 \otimes f_2 d\tilde{\mu} = \int E(f_1|Y)E(f_2|Y)d\nu(y)$. If $\mathbf{X}_i = (X_i, \mathcal{B}_i, \mu_i, T_i)$ then $T_1 \times T_2$ will preserve the measure $\tilde{\mu}$. We can speak of the m.p.s. $(X_1 \times X_2, \mathcal{B}_1 \times \mathcal{B}_2, \tilde{\mu}, T_1 \times T_2)$ which we denote $\mathbf{X}_1 \underset{\mathbf{Y}}{\times} \mathbf{X}_2$. We now make the definition:

Definition. A system $\mathbf{X}$ is a *WM extension* of a factor $\mathbf{Y}$ if $\mathbf{X} \underset{\mathbf{Y}}{\times} \mathbf{X}$ is ergodic.

Our main structure theorem is:

Theorem. *Every ergodic system is a WM extension of its maximal distal factor.*

It follows from this that every ergodic system arises by taking successively isometric and WM extensions beginning with the trivial system.

The ergodic decomposition theorem together with the foregoing structure theorem were made use of in the original proof of (linear) multiple recurrence in the form (1.4) which implies Szemerédi's theorem. (See [FKO82]). A variant of that argument in the spirit of the proof of Szemerédi's theorem for commuting transformations ([FK78]) is the following. Call a system an MR system when (1.4) holds for all sets A of positive measure and for all k. It is relatively straightforward to show that a WM extension of an MR system is MR. Using van der Waerden's theorem on arithmetic progressions, one can show that the MR property is also preserved under isometric extensions. Finally one argues that every system has a maximal MR factor and this proves that every ergodic system is MR. Ultimately by ergodic decomposition the phenomenon of (linear) multiple recurrence is established.

A similar strategy was adopted by V. Bergelson and A. Leibman in [BL96] to obtain a polynomial multiple recurrence theorem:

Theorem. *Let $p_1(n), p_2(n), \ldots, p_k(n)$ be polynomials with integer coefficients and with vanishing constant term ($p_i(0) = 0$), then for any m.p.s. $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ and a set $A \in \mathcal{B}$ with $\mu(A) > 0$, there exists $n \neq 0$ with*

$$\mu(A \cap T^{-p_1(n)}A \cap T^{-p_2(n)}A \cap \cdots \cap T^{-p_k(n)}A) > 0$$

The proof in [FK78] of multiple recurrence which is needed for Szemerédi's theorem on arithmetic progressions and its higher dimensional analogues makes use of the related, classical van der Waerden theorem. For the Bergelson-Leibman polynomial version, a polynomial version of van der Waerden's theorem is needed and this too is established in their paper [BL96].

We remark that the formulation in the foregoing theorem is not the final word on multiple recurrence. The result can be refined to include certain sets of polynomials which do not vanish at 0, but this will require additional machinery which will be discussed.

We may make use of the same correspondence principle alluded to earlier to derive the following result regarding "polynomial progressions":

Theorem. *Let $E \subset \mathbb{Z}$ be a subset of positive upper density, and let $p_1(n)$, $p_2(n), \ldots, p_k(n)$ be k polynomials vanishing for $n = 0$. Then E contains a progression $\{a, a + p_1(n), a + p_2(n), \ldots, a + p_k(n)\}$ with $n \neq 0$.*

3. Characteristic Factors and the van der Corput Lemma

We shall refer to families $\{p_1(n), p_2(n), \ldots, p_k(n)\}$ of integer valued polynomials as *schemes*.

Definition. If $\mathbf{X}$ is a m.p.s. and $\mathbf{Y}$ is a factor of $\mathbf{X}$, we shall say that $\mathbf{Y}$ is a *characteristic factor* for the scheme $\{p_1(n), \ldots, p_k(n)\}$ if for every choice of $f_1, f_2, \ldots, f_k \in L^\infty(X, \mathcal{B}, \mu)$,

$$\frac{1}{N}\sum_0^{N-1} T^{p_1(n)}f_1 T^{p_2(n)}f_2 \cdots T^{p_k(n)}f_k - \frac{1}{N}\sum_0^{N-1} T^{p_1(n)}E(f_1|Y)T^{p_2(n)}E(f_2|Y)\cdots T^{p_k(n)}E(f_k|Y) \to 0$$

in $L^2(X, \mathcal{B}, \mu)$.

Here we have abbreviated $f \circ T$ to Tf. Finding a characteristic factor for a scheme often gives a reduction of the problem of evaluating limit behavior of non-conventional averages to special systems. This will be the case in the proof of the polynomial mean ergodic theorem, which is carried out by first showing the convergence for nilsystems and showing that the latter serve as characteristic factors for all polynomial schemes.

Perhaps the principal tool in identifying characteristic factors is the following lemma which we will refer to as the "Hilbert space van der Corput lemma":

Lemma. *Let $\mathcal{H}$ be a Hilbert space with inner product $\langle\ ,\ \rangle$. Let $\{u_n\}$ be a bounded sequence of vectors in $\mathcal{H}$ and assume that for each m, the limit*

$$\gamma_m = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+m}\rangle$$

exists. If $\frac{1}{M}\sum_1^M \gamma_m \to 0$ as $M \to \infty$, then $\|\frac{1}{N}\sum_1^N u_n\| \to 0$.

We will illustrate the use of this lemma in showing that for any ergodic system $\mathbf{X}$, its Kronecker factor (Z, α) is a characteristic factor for the scheme $\{n, 2n\}$. It suffices to show that if $E(f|Z) = 0$ or $E(g|Z) = 0$ then

$$\frac{1}{N}\sum_{n=1}^{N} T^n f T^{2n} g \to 0$$

in $L^2(X, \mathcal{B}, \mu)$. Regarding the products $T^n f T^{2n} g$ as elements in $L^2(X, \mathcal{B}, \mu)$, we set $u_n = T^n f T^{2n} g$. Then

$$\langle u_n, u_{n+m}\rangle = \int T^n(fT^m\bar{f})T^{2n}(gT^{2m}\bar{g})d\mu = \int fT^m\bar{f}\cdot T^n(gT^{2m}\bar{g})d\mu$$

By the ergodic theorem the average of these expressions over n exists and by ergodicity

$$\gamma_m = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N}\langle u_n, u_{n+m}\rangle = \int fT^m \bar{f} d\mu \int gT^{2m}\bar{g}d\mu$$
$$= \int f\otimes g(T\times T^2)^m \bar{f}\otimes \bar{g}d(\mu\times\mu).$$

The average over m exists: $\frac{1}{M}\sum_1^M \gamma_m \to \int f\otimes gHd(\mu\times\mu)$ where $T\times T^2 H = H$. Now invariant functions on a product system are formed from products of eigenfunctions for the individual systems, from which it follows that if either $E(f|Z) = 0$ or $E(g|Z) = 0$, then $\int f \otimes g\ H\ d(\mu\times\mu) = 0$. This proves that the Kronecker factor is a characteristic factor for $\{n, 2n\}$ as claimed. We remark that following T. Ziegler [ZI07], for any scheme and any system there exists a "minimal" characteristic factor. If we take into account expressions $\frac{1}{N}\Sigma_{n=1}^N T^n\varphi^2 T^{2n}\bar{\varphi}$ where φ is an eigenfunction we see that all eigenfunctions of T appear in any characteristic factor for $\{n, 2n\}$. Thus we have identified the minimal characteristic factor for $\{n, 2n\}$ as the Kronecker factor.

One conclusion that can be drawn is the existence of $\lim \frac{1}{N}\Sigma_{n=1}^N T^n f T^{2n} g$ in $L^2(X,\mathcal{B},\mu)$ for any system. From the foregoing this is reduced to the special case of a Kronecker system and L^2-convergence is readily established in this case. Namely, for convergence in L^2 it suffices to consider f, g in an L^4-dense subset of L^2, and particularly for f, g continuous. For this case we can use the equidistribution of $\{n\alpha\}$ in Z:

$$\frac{1}{N}\sum_{n=1}^{N} f(z+n\alpha)g(z+2n\alpha) \to \int f(z+\theta)g(z+2\theta)d\theta$$

which is true pointwise and consequently also in L^2.

Since strong convergence in $L^2(X,\mathcal{B},\mu)$ implies weak convergence, we can formulate a consequence of the foregoing:

For $f, g, h \in L^\infty(X,\mathcal{B},\mu)$,

$$\frac{1}{N}\sum \int f(x)g(T^n x)h(T^{2n}x)d\,\mu \to$$
$$\int E(f|Z)(z)E(g|Z)(z+\theta)E(h|Z)(z+2\theta)dzd\theta.$$

An instructive interpretation of this is that as x ranges over X and n ranges over non-negative integers, the triple $(x, T^n x, T^{2n}x)$ ranges "freely" over $X\times X\times X$ subject to the condition that $\varphi(x), \varphi(T^n x), \varphi(T^{2n}x)$ form an arithmetic progression in Z, where $\varphi : X \to Z$ is the projection of X to its Kronecker factor. Thus the role played by the characteristic factor here is that of determining the constraints on $(x, T^n x, T^{2n}x)$. It is remarkable that the constraints are purely algebraic.

There is a situation when no constraints exist on $(x, T^{p_1(n)}x, T^{p_2(n)}x, \ldots, T^{p_k(n)}x)$. Another way of saying this is to say that the characteristic factor of $\mathbf{X}$ for $\{p_1(n), p_2(n), \ldots, p_k(n)\}$ is trivial so that

$$\frac{1}{N}\sum T^{p_1(n)}f_1 T^{p_2(n)}f_2 \ldots T^{p_n(n)}f_k \to \int f_1 d\mu \cdot \int f_2 d\mu \ldots \int f_k d\mu$$

in $L^2(X, \mathcal{B}, \mu)$. This will be the case when $\mathbf{X}$ is weakly mixing - or a WM extension of the trivial system - provided the polynomials $p_i - p_j$ for $i \neq j$ differ not only in their constant term. This result was proved by Bergelson [BE87] and the proof makes repeated use of the Hilbert space van der Corput lemma.

4. Geometric Progressions in Nilpotent Groups and on Nilmanifolds

Turning to the general case, one finds that for $k > 2$ the $(k+1)$-tuples $(x, T^n x, T^{2n}x, \ldots, T^{kn}x)$ are subject to further restrictions not implicit in the projection to a $(k+1)$-term arithmetic progression in the Kronecker factor of the system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$. These come from "nil-factors", i.e., factors $(Y, \mathcal{D}, \nu, S)$ where $Y = \mathcal{N}/\Gamma$, $\mathcal{N}$ a nilpotent Lie group, Γ a cocompact discrete subgroup. ν is an $\mathcal{N}$-invariant measure, and $S(u\Gamma) = a_\circ u\Gamma$ for $a_\circ$ fixed in $\mathcal{N}$. The existence of a nil-factor $\pi : X \to \mathcal{N}/\Gamma$ for a nilpotent group $\mathcal{N}$ of level $k-1$ imposes an algebraic condition on $(k+1)$-tuples $(x, T^n x, T^{2n}x, \ldots, T^{kn}x)$. This condition can be stated as the requirement that $(\pi(x), \pi(T^n x), \ldots, \pi(T^{kn}x))$ belong to a submanifold of $(\mathcal{N}/\Gamma)^{k+1}$ which we designate $HP_{k+1}(\mathcal{N}/\Gamma)$. H-P stands for Hall and Petresco who studied the term by term products of geometric progressions for non-commutative groups, these no longer having to be geometric progressions.

Definition. Let $G \supset G^{(1)} \supset G^{(2)} \supset \cdots$ be the lower central series of a group G, $G^{(i+1)} = [G, G^{(i)}]$. A $(k+1)$-term sequence $\{u_0, u_1, u_2, \ldots, u_k\}$ is an HP$_{k+1}$-sequence if there exist $x_1 \in G, x_2 \in G^{(1)}, \ldots, x_k \in G^{(k-1)}$ so that

$$u_1 = x_1 u_0,\ u_2 = x_2 x_1^2 u_0,\ u_3 = x_3 x_2^3 x_1^3 u_0,\ u_4 = x_4 x_3^4 x_2^6 x_1^4 u_0\,,\ \ldots,\ u_k = x_k x_{k-1}^k \cdots x_1^k u_0.$$

The significance of this notion is that the HP_{k+1} sequences form a group in G^{k+1} ([LE98], [TT09, p. 217]). We can define the projection of such sequences on a homogeneous space G/Γ as HP_{k+1}-progressions which form a subvariety of $(G/\Gamma)^{k+1}$. The role played by nilpotence comes from the following:

Lemma. *If $\mathcal{N}$ is k-step nilpotent, i.e., $\mathcal{N}^{(k)} = \{1\}$, then the first $k+1$ terms of an HP_ℓ-progression, $\ell > k+1$, determine all successive terms.*

Now let $\mathbf{X}$ be an arbitrary m.p.s. possessing a k-step nilfactor, then the projections $\pi(x), \pi(T^n x), \ldots, \pi(T^{\ell n} x)$ form a $HP_{\ell+1}$ sequence on the factor and this imposes new constraints on $(x, T^n x, T^{2n} x, \ldots, T^{\ell n} x)$. As in the case of $\{n, 2n\}$ these constraints turn out to be the only ones on $(x, T^n x, T^{2n} x, \ldots, T^{kn} x)$. To make this precise we formulate the notion of a k-step pro-nilsystem: a k-step pro-nilsystem is an inverse limit of nilsystems $\underleftarrow{\lim} \mathcal{N}_j/\Gamma_j$ where $\mathcal{N}_j$ is a nilpotent Lie group with $\mathcal{N}_j^{(k)} = \{1\}$, and on each of these the measure preserving action is translation by an element of $\mathcal{N}_j$, so that the inverse system is consistently defined. Every ergodic system $\mathbf{X}$ will have a maximal k-step pro-nilflow factor $\mathbf{Z}_k$ and Ziegler's theorem asserts that $\mathbf{Z}_k$ is characteristic for $\{n, 2n, \ldots, kn\}$, and, more generally, for any linear family $\{a_1 n, a_2 n, \ldots, a_k n\}$ with distinct a_i [ZI07].

As remarked earlier, a consequence of this identification of the characteristic factor for any ergodic system enables us to prove convergence in $L^2(X, \mathcal{B}, \mu)$ of

$$\frac{1}{N} \sum_{n=1}^{N} f_1(T^{a_1 n} x) f_2(T^{a_2 n} x) \cdots f_k(T^{a_k n} x)$$

as $N \to \infty$, since this will now follow for any system once it is known for translations on nilmanifolds. For nilmanifolds this was established in 1969 by W. Parry ([PA69]), and is also a special case of theorems of N. Shah ([SH96]) and Leibman ([LE05]). An explicit description of the limit appears in [ZI05] and for the special case $k = 3$ was given by E. Lesigne ([Le89]). The entire theory was developed for $k = 3$ by J.P. Conze and E. Lesigne in [CL84] and [CL87], who first recognized the role of nilmanifolds for 3-term non-conventional averages.

In fact pro-nilsystems serve as characteristic factors for any scheme, and both the polynomial mean ergodic theorem and polynomial multiple recurrence can be deduced from this. With the identification of the characteristic factor for any scheme, the polynomial mean ergodic theorem as well as a polynomial multiple recurrence theorem will follow for arbitrary measure preserving systems, once they are known for nilsystems. As regards the polynomial mean ergodic theorem one has available for nilsystems a pointwise ergodic theorem which is valid for all points for continuous functions by results of Leibman ([LE05]). In addition, the analysis of distribution of polynomial orbits on a nilmanifold leads to the following refinement of our earlier multiple recurrence theorem:

Call a set of integer valued polynomial $\{q_1(n), \ldots, q_r(n)\}$ *intersective* if for any $m \in \mathbb{N}$, there exists n such that m divides each $q_i(n)$. Then if $\{q_1(n), \ldots, q_r(n)\}$ is an intersective family, for any m.p.s. $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$ there exists an integer n with $\mu(A \cap T^{-q_1(n)} A \cap T^{-q_2(n)} A \cap \cdots \cap T^{-q_r(n)} A) > 0$.

Note that the sufficient condition for multiple recurrence given in §2 is a special case of the above. But there are intersective families of polynomials which don't have a common 0 and so the present theorem is strictly stronger than the earlier one. This is noteworthy since the theorem in [BL96] makes use implicitly of the distal factor of a given ergodic system, whereas the present refinement in [BLL08] makes use of a special distal factor - namely, the pro-nilfactor.

5. Conze-Lesigne Factors

The main result of these investigations is identifying the nilfactor of an ergodic system as the characteristic factor for all schemes. Two approaches have been taken up and these show up in considering the schemes $\{a_1n, a_2n, \ldots, a_kn\}$. The approach of Conze and Lesigne has been mentioned and this was generalized by T. Ziegler from the case $k = 3$ to arbitrary k ([ZI07]). In line with this approach is the treatment in [FW96] of characteristic factors for $\{a_1n, a_2n, a_3n\}$ which, as is shown there, is also characteristic for $\{n, n^2\}$. This was the first instance of a non-linear scheme to be treated, and for which a mean ergodic theorem was proved. B. Host and B. Kra have an entirely different approach leading ultimately to the same conclusion ([HK05]).

We begin with what can be called the Conze-Lesigne approach. With Ziegler ([ZI07]) we denote by $\mathbf{Y}_k = \mathbf{Y}_k(\mathbf{X})$ the "universal" characteristic factor for schemes $\{a_1n, a_2n, \ldots, a_kn\}$ which, first of all, is shown to exist. It is manifest that $\mathbf{Y}_{k+1}$ is an extension of $\mathbf{Y}_k$. It can be shown to be an isometric extension and moreover an extension $Y_{k+1} = Y_k \times W_k$, where W_k is a compact abelian group. The action on Y_{k+1} is given by $T(y, w) = (Ty, \rho(y)w)$ and further analysis shows that the "cocycle" ρ is not arbitrary but satisfies a functional equation. This has led to the important notion of a "Conze-Lesigne cocycle" which appears in contemporary treatments of more general convergence questions. In the simplest situation $k = 2$ where $\mathbf{Y}_2$ has already been shown to coincide with the Kronecker factor $(\tilde{Z}, \tilde{\alpha})$ of $\mathbf{X}$, the Conze-Lesigne condition takes the form: there exist measurable functions K and L with

$$\frac{\rho(z+u)}{\rho(z)} = K(u)\frac{L(u, z+\alpha)}{L(u, z)}.$$

Conze and Lesigne arrived at this equation in their direct treatment of convergence of ergodic averages, but Ziegler makes use of it and its analogs for higher k to show that the k-universal factor is a $(k-1)$-step pro-nilsystem which can be denoted $\mathbf{Z}_{k-1}(\mathbf{X})$.

6. Gowers Seminorms and Host-Kra Factors

In his proof of Szemerédi's theorem ([GO01]), T. Gowers introduced a notion of mixing (he calls it "uniformity") which is useful in studying non-conventional averages. With B. Host and B. Kra one defines an ergodic theoretic analog of an expression studied by Gowers: the k-seminorm $|||f|||_k$ of a bounded measurable function which can be defined inductively by

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

and $|||f|||_0 = \int f d\mu$. These are non-decreasing with k so that the condition $|||f|||_k = 0$ becomes more and more restrictive. It can be shown that if f is orthogonal to the distal component of a system $\mathbf{X}$, then $|||f|||_k = 0$ for all k. On the other hand $|||f|||_k = 0$ if $f \perp g$ for functions g on X coming from the $(k-1)$-step pro-nilfactor, and indeed this nilfactor can be characterized by this quantitative condition on its orthogonal complement. A direct definition of $|||\ \ |||_k$ is given in [HK05] where the seminorm appears as an autocorrelation of values of a function on "cubes", these being special 2^k-tuples of points in X. For our purposes, the main result is the theorem of Leibman [LE05,1].

Theorem. *For any $r, b \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that for any system of non-constant essentially distinct integer valued polynomials $p_1, \ldots, p_r$ of degree $\leq b$ and any $f_1, f_2, \ldots, f_r \in L^\infty(X, \mathcal{B}, \mu)$ for a m.p.s. $\mathbf{X}$ for which $|||f_1|||_k = 0$, one has*

$$\frac{1}{N} \sum_{0}^{N-1} T^{p_1(n)} f_1 \, T^{p_2(n)} f_2 \cdots T^{p_r(n)} f_r \to 0$$

in $L^2(X, \mathcal{B}, \mu)$ as $N \to \infty$.

It follows from this theorem that for any scheme $\{p_1(n), p_2(n), \ldots, p_r(n)\}$, the $(k-1)$-step pro-nilfactor of $\mathbf{X}$ serves as a characteristic factor provided k is sufficiently large.

Pro-nilfactors appear as characteristic in a related but different context. Namely one can form multi-parameter averages:

$$\lim_{N_1, N_2 \cdots N_k \to \infty} \frac{1}{N_1 N_2 \cdots N_k} \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \cdots \sum_{n_k=1}^{N_k} \prod_{\varepsilon_1, \ldots, \varepsilon_k \in \{0,1\}} T^{\varepsilon_1 n_1 + \cdots + \varepsilon_k n_k} f_{\varepsilon_1 \varepsilon_2 \cdots \varepsilon_k} d\mu.$$

These were first considered by Bergelson for $k = 2$, who showed that

$$\lim_{N_1, N_2 \to \infty} \frac{1}{N_1 N_2} \sum_{n_1=1}^{N_1} f(T^{n_1} x) f(T^{n_1} x) g(T^{n_2} x) h(T^{n_1+n_2} x)$$

exists in $L^2(X, \mathcal{B}, \mu)$. It turns out that $\mathbf{Z}_{k-1}(\mathbf{X})$ is a characteristic factor (in this extended sense) for this expression as well. ([BE00])

We have only skimmed the surface of a large area, which is still growing. Much work has already been done when powers of a single transformation are replaced by more general commuting transformations. Another notion of interest is that of IP-limit (see [BE06]) replacing the usual average. This plays a central role in establishing a density version of the Hales-Jewett theorem. See [FK85] and [BM00] for further details.

References

[BE87] V. Bergelson, *Weakly mixing PET*, Ergodic Theory and Dynamical Systems **7** (1987), 337–349.

[BE00] V. Bergelson, *The multifarious Poincaré recurrence theorem* in *Descriptive Set Theory and Dynamical Systems*, Eds. M. Foreman, A.S. Kechris, A. Louveau, B. Weiss. Cambridge Univ. Press, New York, 2000, 31–57.

[BE06] V. Bergelson, *Combinatorial and diophantine applications of ergodic theory*, in *Handbook of Dynamical Systems, vol. 1B* with appendix by A. Leibman, p. 844.

[BL96] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), 725–753.

[BLL08] V. Bergelson, A. Leibman and E. Lesigne, *Intersective polynomials and the polynomial Szemerédi theorem*, Adv. Math. **219** (2008), no. 1, p. 369.

[BM00] V. Bergelson and R. McCutcheon, *An ergodic IP polynomial Szemerédi theorem*, Mem. Amer. Math. Soc. **146** (2000), no. 695, viii+106 pp.

[CL84] J.P. Conze and E. Lesigne, *Théorèmes ergodiques pour des mesures diagonales*, Bull. Soc. Math. France **112** (1984), no. 2, 143–175.

[CL87] J.P. Conze and E. Lesigne, *Sur un théorème ergodique pour des mesures diagonales*, Publ. Inst. Recherch Math. Rennes 1987-1, Univ. Rennes I Rennes.

[DO53] J.L. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[EW10] M. Einsiedler and T. Ward, *Ergodic Theory with a view towards Number Theory*, Springer, Berlin, 2010, p. 178.

[FK78] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291 (1979).

[FK85] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Analyse Math. **45** (1985), 117–168.

[FU77] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[FU81] H. Furstenberg, *Recurrence in Ergodic theory and Combinatorial Number Theory*, Princeton University Press, Princeton, NJ, 1981.

[FKO82] H. Furstenberg, Y. Katznelson and D. Ornstein, *The ergodic theoretical proof of Szemerédi's theorem*, Bull. Amer. Math. Soc. (N.S.) **7** (1982), no. 2, 527–552.

[FW96] H. Furstenberg and B. Weiss, *A mean ergodic theorem for* $\frac{1}{N}\sum f(T^n x)g(T^{n^2}x)$ in *Convergence in Ergodic theory and Probability*, Columbus OH, Ohio State University Math. Res. Inst. Publicatins, vol. 5, p. 193.

[GL03] E. Glasner, *Ergodic Theory via Joinings*, Mathematical Surveys and Monographs, vol. **101**, A.M.S., Providence, RI, 2003.

[GO01] W.T. Gowers, *A new proof of Szemerédi's theorem*, Geom. and Functional Analysis **11** (2001), 465.

[HK01] B. Host and B. Kra, *Convergence of Conze-Lesigne average*, Ergodic Theory and Dynamical Systems **21** (2001) no. 2, 493–509.

[HK05] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Annals of Math. (2) **161** (2005), p. 397.

[LE98] A. Leibman, *Polynomial sequences in groups*, J. Algebra **201** (1998), 189–206.

[LE05] A. Leibman, *Convergence of multiple ergodic averages along polynomials of several variables*, Israel Journal of Mathematics **146** (2005), p. 303.

[LE05,1] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory and Dynamical Systems **25** (2005), part 1, p. 201.

[Le89] E. Lesigne, *Théorèmes ergodiques pour une tranlation sur une nilvariété*, Ergodic Theory and Dynamical Systems **9** (part 1, 1989), p. 115.

[PA69] W. Parry, *Ergodic properties of affine transformations and flows on nilmanifolds*, Amer. J. Math. 91 (1969), p. 757.

[SH96] N. Shah, *Invariant measures and orbit closures on homogeneous spaces for actions of subgroups*, Lie Groups and Ergodic Theory, (Mumbai 1996), 229, Tata Institute Studies in Mathematics **14**.

[TT09] T. Tao, *A Ratner-type theorem for nilmanifolds*, in *Poincaré's Legacies, Part I*, Amer. Math. Society, Providence, RI, 2009, p. 217.

[ZI05] T. Ziegler, *A non-conventional ergodic theorem for a nilsystem*, Ergodic Theory and Dynamical Systems **25** part 4, (2005), p. 1357.

[ZI07] T. Ziegler, *Universal Characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), p. 53.