Uniformity in Polynomial Szemerédi Theorem

VITALY BERGELSON
The Ohio State University, Columbus, OH 43210 U.S.A.

RANDALL McCUTCHEON
The Ohio State University, Columbus, OH 43210 U.S.A.

0. Introduction

The purpose of this paper is to give a concise proof and some combinatorial consequences of the following theorem.

Theorem 0.1 Assume that $(X, \mathcal{A}, \mu, T)$ is an invertible probability measure preserving system, $k \in \mathbf{N}$, $A \in \mathcal{A}$ with $\mu(A) > 0$, and $p_i(x) \in \mathbf{Q}[x]$ are polynomials satisfying $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$, $1 \leq i \leq k$. Then

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

Being less general than the polynomial Szemerédi theorem obtained in [BL1] (see Theorem 1.19 in [B2], these Proceedings) in that it deals with one rather than a commuting family of invertible measure preserving transformations of $(X, \mathcal{A}, \mu)$ and thereby has combinatorial applications in $\mathbf{Z}$ and $\mathbf{R}$ rather than in $\mathbf{Z}^t$ and $\mathbf{R}^t$, Theorem 0.1 has nevertheless a novel stronger although subtle feature: the limit appearing there is a uniform limit, whereas the main theorem of [BL1] would merely give, in the case of a single measure preserving transformation,

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

The reasons for undertaking the task of presenting a proof of Theorem 0.1 are twofold. First of all, the argument that we are going to give has some new and in our opinion promising features developed for our proof of an IP polynomial Szemerédi theorem (see section 4 in [B2]). These features, which allow for the attainment of uniformity of the limit, are more general than the methods of [BL1]. What one would like to have, of course, is an extension of Theorem 0.1 to a multi-operator situation, namely, one would like to show (for example) that for commuting invertible measure preserving transformations $T_1, \cdots, T_k$ of $(X, \mathcal{A}, \mu)$ one has

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T_1^{p_1(n)}A \cap \cdots \cap T_k^{p_k(n)}A) > 0.$$

Indeed, with extra effort we could, combining the techniques we present here with those of [BL1], give a full generalization of [BL1], namely we could show that for polynomials $p_{ij}(x) \in \mathbf{Q}[x]$ with $p_{ij}(\mathbf{Z}) \subset \mathbf{Z}$ and $p_{ij}(0) = 0$, $1 \le i \le s$, $1 \le j \le k$, and $A \in \mathcal{A}$ with $\mu(A) > 0$ one has

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu\Big(\bigcap_{i=1}^{s} \Big(\prod_{j=1}^{k} T_j^{p_{ij}(n)}\Big) A\Big) > 0.$$

However, the proof in this general case would be much more cumbersome than that of [BL1], to the point of obscuring the new features. Therefore, we choose to confine ourselves to the single operator case.

The other reason for presenting a proof of the uniformity of the limit is that this seems to be the right or most desirable thing to have in any ergodic theorem dealing with weak or norm Cesàro convergence. For example, Furstenberg's ergodic Szemerédi theorem [F1] established uniformity of the limit appearing in Theorem 0.1 for first degree polynomials $p_i(n)$. We would like to say that

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{p_1(n)}A \cap \cdots \cap T^{p_k(n)}A)$$

exists, but this seems to be presently out of reach even in the linear case when $k \ge 4$ (for the non-uniform limit as well).

Two major tools which we use are:

(i) The structure theorem for measure preserving systems established by Furstenberg in [F1], and which will be used in the form appearing in [FKO], which, roughly speaking, tells us that $(X, \mathcal{A}, \mu, T)$ can be "exhausted" by a chain of factors so that at every link in the chain there is either *relative compactness* or *relative weak mixing.*

(ii) An elaboration of a special case of a *Polynomial Hales-Jewett Theorem*, recently obtained in [BL2] which plays in our treatment the role analogous to that of the polynomial van der Waerden theorem in the proof of the polynomial Szemerédi theorem in [BL1], and which allows us to push uniformity of the limit through compact extensions. See Theorem 3.1, Corollary 3.2, and the appendix (Section 4).

We wish to conclude this introduction by giving some of the combinatorial consequences of Theorem 0.1. Each is proved using a *correspondence principle* due to Furstenberg (see Theorem 0.2 below). In order to formulate this correspondence principle, as well as our applications, we will remind the reader of a few definitions.

Suppose that $E \subset \mathbf{Z}$ is a set. The *upper density*, $\overline{d}(E)$, and *lower density* $\underline{d}(E)$ of E are defined by

$$\overline{d}(E) = \limsup_{N\to\infty} \frac{\big|E \cap [-N, -N+1, , \cdots, N]\big|}{2N+1}, \quad \underline{d}(E) = \liminf_{N\to\infty} \frac{\big|E \cap [-N, -N+1, \cdots, N]\big|}{2N+1}.$$

The *upper Banach density* of E is given by

$$d^*(E) = \limsup_{N-M\to\infty} \frac{\big|E \cap [M, \cdots, N-1]\big|}{N-M}.$$

E is said to be *syndetic* if it has bounded gaps, or, more formally, if for some finite set $F \subset \mathbf{Z}$ one has

$$E + F = \{x + y : x \in E, y \in F\} = \mathbf{Z}.$$

Clearly, any syndetic set has positive lower density, any set of positive lower density has postive upper density, and any set of positive upper density has positive upper Banach density. It is also completely clear that these are all different notions.

Here now is Furstenberg's correspondence principle.

Theorem 0.2 Given a set $E \subset \mathbf{Z}$ with $d^*(E) > 0$ there exists a probability measure preserving system $(X, \mathcal{A}, \mu, T)$ and a set $A \in \mathcal{A}$, $\mu(A) = d^*(E)$, such that for any $k \in \mathbf{N}$ and any $n_1, \cdots, n_k \in \mathbf{Z}$ one has:

$$d^*\big(E \cap (E - n_1) \cap \cdots \cap (E - n_k)\big) \geq \mu(A \cap T^{n_1} A \cap \cdots \cap T^{n_k} A).$$

The first of our applications follows easily from Theorems 0.1 and 0.2.

Theorem 0.3 Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$. Then for any polynomials $p_i(x) \in \mathbf{Q}[x]$ with $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$, $1 \leq i \leq k$, the set

$$\big\{n \in \mathbf{Z} : \text{ for some } x \in \mathbf{Z}, \{x, x + p_1(n), \cdots, x + p_k(n)\} \subset E\big\}$$

is syndetic.

Theorem 0.4 Let $p_i(x) \in \mathbf{Q}[x]$ with $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$, $1 \leq i \leq k$. Suppose that $r \in \mathbf{N}$ and that $\mathbf{N} = \bigcup_{i=1}^r C_i$ is any partition of $\mathbf{N}$ into r cells. Then there exists some $L \in \mathbf{N}$ and some $\alpha > 0$ with the property that for any interval $I = [M, N] \subset \mathbf{Z}$ with $N - M \geq L$ there exists i, $1 \leq i \leq r$, and $n \in C_i \cap I$ such that

$$d^*\Big(C_i \cap \big(C_i - p_1(n)\big) \cap \cdots \cap \big(C_i - p_k(n)\big)\Big) \geq \alpha.$$

In particular, the system of polynomial equations

$$\begin{aligned} x_0 &= n, \\ x_2 - x_1 &= p_1(n), \\ x_3 - x_1 &= p_2(n), \\ &\vdots \\ x_{k+1} - x_1 &= p_k(n) \end{aligned}$$

has monochromatic solutions $\{x_0, \cdots, x_{k+1}\}$ with $n = x_0$ choosable from any long enough interval.

Proof. Renumbering the sets C_i if needed, let $(C_i)_{i=1}^s$, where $s \leq r$, be those C_i for which $d^*(C_i) > 0$. For all i, $1 \leq i \leq s$, let $(X_i, \mathcal{A}_i, \mu_i, T_i)$ and $A_i \in \mathcal{A}_i$, $\mu(A_i) = d^*(C_i)$

be measure preserving systems and sets having the property that for any $k \in \mathbf{N}$ and any $n_1, \cdots, n_k \in \mathbf{Z}$ one has:

$$d^*\big(C_i \cap (C_i - n_1) \cap \cdots \cap (C_i - n_k)\big) \geq \mu_i(A_i \cap T_i^{n_1} A_i \cap \cdots \cap T_i^{n_k} A_i).$$

This is possible by Theorem 0.2. Applying Theorem 0.1 to $X = X_1 \times \cdots \times X_s$, $\mathcal{A} = \mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_s$, $\mu = \mu_1 \times \cdots \times \mu_s$, $A = A_1 \times \cdots \times A_s$, and $T = T_1 \times \cdots \times T_s$ we have:

$$\begin{aligned}&\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{p_1(n)} A \cap \cdots \cap T^{p_s(n)} A) \\ &= \liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \prod_{i=1}^{s} \mu_i(A_i \cap T_i^{p_1(n)} A_i \cap \cdots \cap T_i^{p_s(n)} A_i) = 3\alpha > 0.\end{aligned} \tag{1}$$

Let

$$S = \big\{n : \mu(A \cap T^{p_1(n)} A \cap \cdots \cap T^{p_s(n)} A) \geq \alpha\big\}.$$

Let L now be large enough that if $I = [M, N] \subset \mathbf{Z}$ is any interval with $N - M > L$ then

$$\big|S \cap I\big| > \alpha\big|I\big| \text{ and } \Big|\Big(\bigcup_{i=1}^{s} C_i\Big) \cap I\Big| > (1-\alpha)\big|I\big|.$$

(The former is possible by (1) which, as one may check, cannot hold unless $|S \cap I| > \alpha|I|$ for large intervals I. The latter is possible since C_i, $1 \leq i \leq s$ consist of *all* C_i for which $d^*(C_i) > 0$.) Of course, it follows that for any interval I with $|I| > L$, there exists i, $1 \leq i \leq s$, and $n \in C_i \cap I \cap S$. For this n we have

$$\begin{aligned}d^*\Big(C_i \cap (C_i - p_1(n)) \cap \cdots \cap \big(C_i - p_k(n)\big)\Big) &\geq \mu_i(A_i \cap T_i^{p_1(n)} A_i \cap \cdots \cap T_i^{p_k(n)} A_i) \\ &\geq \mu(A \cap T^{p_1(n)} A \cap \cdots \cap T^{p_k(n)} A) \geq \alpha.\end{aligned}$$

□

1. Measure theoretic preliminaries.

In this section we collect the facts concerning measure spaces and their factors which we will be using. For more details, the reader may wish to consult [FKO]. First of all, we remark that for the proof of Theorem 0.1 it suffices to assume that the measure space $(X, \mathcal{A}, \mu)$ is a Lebesgue space. This is a result of the fact that we may always pass to the σ-algebra generated by $(T^n A)_{n \in A}$, which is separable, the fact that μ may clearly be assumed non-atomic, and the fact that Theorem 0.1 is a result about the measure algebra induced by $(X, \mathcal{A}, \mu)$ (with no reference to the points of X). Therefore, as any separable non-atomic probability measure algebra is isomorphic to that induced by Lebesgue measure on the unit interval, one may freely choose $(X, \mathcal{A}, \mu)$ to be any measure space having a separable, non-atomic measure algebra; in particular, one may assume that $(X, \mathcal{A}, \mu)$ is Lebesgue. Furthermore, we may, in view of ergodic decomposition, assume that T is ergodic.

Suppose that $(X, \mathcal{A}, \mu, T)$ is an ergodic measure preserving system, where $(X, \mathcal{A}, \mu)$ is a Lebesgue space, and that $\mathcal{B} \subset \mathcal{A}$ is a complete, T-invariant σ-algebra. Then $\mathcal{B}$ determines a factor $(Y, \mathcal{B}_1, \nu, S)$ of $(X, \mathcal{A}, \mu, T)$, the construction of which we now indicate. Let $(B_i)_{i=1}^{\infty} \subset \mathcal{B}$ be a T-invariant sequence of sets which is dense in $\mathcal{B}$ (in the sense that for any $B \in \mathcal{B}$ and $\epsilon > 0$ there exists $i \in \mathbf{N}$ such that $\mu(B \triangle B_i) < \epsilon$), and denote by Y the set of equivalence classes under the equivalence relation which identifies x_1 with x_2, $x_1 \approx x_2$, when for all $i \in \mathbf{N}$, $x_1 \in B_i$ if and only if $x_2 \in B_i$. Let $\pi : X \to Y$ be the natural projection and let $\mathcal{B}_1 = \{B \subset Y : \pi^{-1}(B) \in \mathcal{B}\}$. For $B \in \mathcal{B}_1$, let $\nu(B) = \mu(\pi^{-1}B)$. Finally, write $S\pi(x) = \pi(Tx)$. Then $(Y, \mathcal{B}_1, \nu, S)$ is a factor of $(X, \mathcal{A}, \mu, T)$.

Since any complete, T-invariant σ-algebra $\mathcal{B} \subset \mathcal{A}$ determines such a factor, we will simply say that $\mathcal{B}$ is a factor of $\mathcal{A}$, or that $\mathcal{A}$ is an extension of $\mathcal{B}$, and will identify $\mathcal{B}_1$ with $\mathcal{B}$ when referring to the induced system, which we now write as $(Y, \mathcal{B}, \nu, S)$. If $x \in X$ and $y \in Y$, with $y = \pi(x)$, we will say that "x is in the fiber over y."

If $(Y, \mathcal{B}, \nu)$ is a factor of $(X, \mathcal{A}, \mu)$, then there is a uniquely (up to null sets in Y) determined family of probability measures $\{\mu_y : y \in Y\}$ on X with the property that μ_y is supported on $\pi^{-1}(y)$ for a.e. $y \in Y$ and such that for every $f \in L^1(X, \mathcal{A}, \mu)$ we have

$$\int_X f(x)\, d\mu(x) = \int_Y \Big(\int_X f(z)\, d\mu_y(z) \Big)\, d\nu(y).$$

Sometimes we write μ_x for μ_y when x is in the fiber over y. The decomposition gives, for any $\mathcal{A}$-measurable function f, the *conditional expectation* $E(f|\mathcal{B})$:

$$E(f|\mathcal{B})(y) = \int_X f(x) d\mu_y(x) \;\; a.e.$$

Equivalently, the conditional expectation $E(\cdot|\mathcal{B}) : L^2(X, \mathcal{A}, \mu) \to L^2(X, \mathcal{B}, \mu)$ is the orthogonal projection onto $L^2(X, \mathcal{B}, \mu)$. In particular, $E\big(E(f|\mathcal{B})|\mathcal{B}\big) = E(f|\mathcal{B})$.

Let $\mathcal{A} \otimes \mathcal{A}$ be the completion of the σ-algebra of subsets of $X \times X$ generated by all rectangles $C \times D$, $C, D \in A$. Now define a $T \times T$-invariant measure $\tilde{\mu}$ on $(X \times X, A \otimes A)$ by letting, for $f_1, f_2 \in L^\infty(X, \mathcal{A}, \mu)$

$$\int_{X \times X} f_1(x_1) f_2(x_2)\, d\tilde{\mu} = \int_Y \int_X \int_X f_1(x_1) f_2(x_2)\, d\mu_y(x_1) d\mu_y(x_2) d\nu(y).$$

(It is clear that there is one and only one measure which satisfies this condition.) We write $X \times_Y X$ for the set of pairs $(x_1, x_2) \in X \times X$ with $x_1 \approx x_2$. One checks that $X \times_Y X$ is the support of $\tilde{\mu}$, and we speak of the measure preserving system $(X \times_Y X, \mathcal{A} \otimes_{\mathcal{B}} \mathcal{A}, \tilde{\mu}, \tilde{T})$, where $\mathcal{A} \otimes_{\mathcal{B}} \mathcal{A}$ is the $\tilde{\mu}$-completion of the σ-algebra $\{\big(X \times_Y X\big) \cap C : C \in \mathcal{A} \otimes \mathcal{A}\}$ and $\tilde{T}$ is the restriction of $T \times T$ to $X \times_Y X$.

We now procede to introduce the basic elements of the Furstenberg structure theory. The specific format we adopt is from [FKO].

Definition 1.1 Suppose that $(Y, \mathcal{B}, \nu, S)$ is a factor of an ergodic system $(X, \mathcal{A}, \mu, T)$ arising from a complete, T-invariant σ-algebra $\mathcal{B} \subset \mathcal{A}$, and $f \in L^2(X, \mathcal{A}, \mu)$. We will say that f is *almost periodic over* Y, and write $f \in AP$, if for every $\delta > 0$ there exist functions

$g_1, \cdots, g_k \in L^2(X, \mathcal{A}, \mu)$ such that for every $n \in \mathbf{Z}$ and a.e. $y \in Y$ there exists some $s = s(n, y)$, $1 \leq s \leq k$, such that $||T^n f - g_s||_{L^2(X, \mathcal{B}, \mu_y)} < \delta$. If the set AP of almost periodic over Y functions is dense in $L^2(X, \mathcal{A}, \mu)$, we say that $(X, \mathcal{A}, \mu, T)$ is a *compact extension* of $(Y, \mathcal{B}, \nu, S)$, or simply that $\mathcal{A}$ is a compact extension of $\mathcal{B}$. If $(X \times_Y X, \mathcal{A} \otimes_{\mathcal{B}} \mathcal{A}, \tilde{\mu}, \tilde{T})$ is ergodic, then $(X, \mathcal{A}, \mu, T)$ is said to be a *weakly mixing extension* of $(Y, \mathcal{B}, \nu, S)$, or, $\mathcal{A}$ is said to be a weakly mixing extension of $\mathcal{B}$.

For proofs of the following two propositions, see [FKO].

Proposition 1.2 Suppose that an ergodic system $(X, \mathcal{A}, \mu, T)$ is a weakly mixing extension of $(Y, \mathcal{B}, \nu, S)$. Then $(X \times_Y X, \mathcal{A} \otimes_{\mathcal{B}} \mathcal{A}, \tilde{\mu}, \tilde{T})$ is also a weakly mixing extension of $(Y, \mathcal{B}, \nu, S)$.

Proposition 1.3 Suppose that $(X, \mathcal{A}, \mu, T)$ is a compact extension of $(Y, \mathcal{B}, \nu, S)$. Then for every $A \in \mathcal{A}$ with $\mu(A) > 0$ there exists some $A' \subset A$ with $\mu(A') > 0$ and $1_{A'} \in AP$.

The notions of relative weak mixing and relative compactness are mutually exclusive. Moreover, one may show that $(X, \mathcal{A}, \mu, T)$ is a weakly mixing extension of $(Y, \mathcal{B}, \nu, S)$ if and only if there is no intermediate factor $(Z, \mathcal{C}, \gamma, U)$ between $(X, \mathcal{A}, \mu, T)$ and $(Y, \mathcal{B}, \nu, S)$ which is a proper compact extension of $(Y, \mathcal{B}, \nu, S)$. (This is the relativized version of the fact that a system is weakly mixing if and only if it has no non-trivial compact factor.) The structure theorem (see Theorem 6.17 in [F2] and remarks following) we need may now be formulated as follows:

Theorem 1.4 Suppose that $(X, \mathcal{A}, \mu, T)$ is a separable measure preserving system. There is an ordinal η and a system of T-invariant sub-σ algebras $\{\mathcal{A}_\xi \subset \mathcal{A} : \ \xi \leq \eta\}$ such that:

(i) $\mathcal{A}_0 = \{\emptyset, X\}$

(ii) For every $\xi < \eta$, $\mathcal{A}_{\xi+1}$ is a compact extension of $\mathcal{A}_\xi$.

(iii) If $\xi \leq \eta$ is a limit ordinal then $\mathcal{A}_\xi$ is the completion of the σ-algebra generated by $\bigcup_{\xi' < \xi} \mathcal{A}_{\xi'}$.

(iv) Either $\mathcal{A}_\eta = \mathcal{A}$ or else $\mathcal{A}$ is a weakly mixing extension of $\mathcal{A}_\eta$.

The factor $\mathcal{A}_\eta$ appearing in the structure theorem is called the *maximal distal factor* of $\mathcal{A}$. In the next section, we show that in order to prove Theorem 0.1 for the system $(X, \mathcal{A}, \mu, T)$, it suffices to establish that the conclusion holds when A is taken from its maximal distal factor.

2. Weakly mixing extensions.

In this section, we will prove the following relativized version of Theorem 3.1 from [B1].

Theorem 2.1 Suppose an ergodic system $(X, \mathcal{A}, \mu, T)$ is a weakly mixing extension of $(Y, \mathcal{B}, \nu, S)$, and that $p_1(x), \cdots, p_k(x) \in \mathbf{Q}[x]$ are non-zero, pairwise distinct polynomials with $p_i(\mathbf{Z}) \subset \mathbf{Z}$ and $p_i(0) = 0$, $1 \leq i \leq k$. Then for any $f_1, \cdots, f_k \in L^\infty(X, \mathcal{A}, \mu)$,

$$\lim_{N-M \to \infty} \left|\left| \frac{1}{N-M} \sum_{n=M}^{N-1} \Big(\prod_{i=1}^{k} T^{p_i(n)} f_i - \prod_{i=1}^{k} S^{p_i(n)} E(f_i | \mathcal{B}) \Big) \right|\right| = 0.$$

Using Theorem 2.1, we will then show (Corollary 2.5) that if the conclusion of Theorem 0.1 holds for the maximal distal factor of a system, then it also holds for the full system. In other words, the validity of Theorem 0.1 passes through weakly mixing extensions.

We will be using the following concept of *convergence in density.*

Definition 2.2 Suppose that $(x_h)_{h\in\mathbf{N}} \subset \mathbf{R}$. If for every $\epsilon > 0$ the set $\{h \in \mathbf{Z} : |x_h - x| < \epsilon\}$ has (lower) density 1, we write

$$D\text{-}\lim_{h\to\infty} x_h = x.$$

Equivalently, $x_h \to x$ as $h \to \infty$, $h \notin E$ for some $E \subset \mathbf{Z}$ with $\overline{d}(E) = 0$.

We call the following lemma a "van der Corput type trick", because it is motivated by van der Corput's fundamental inequality.

Lemma 2.3 Suppose that $\{x_n : n \in \mathbf{Z}\}$ is a bounded sequence of vectors in a Hilbert space $\mathcal{H}$. If

$$D\text{-}\lim_{h\to\infty} \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \langle x_n, x_{n+h}\rangle = 0,$$

then

$$\lim_{N-M\to\infty} \left|\left| \frac{1}{N-M} \sum_{n=M}^{N-1} x_n \right|\right| = 0.$$

Proof. Let $\epsilon > 0$. Fix H large enough that

$$\sum_{r=-H}^{H} \frac{H-|r|}{H^2} \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{u=M}^{N-1} \langle x_u, x_{u+r}\rangle < \epsilon.$$

We have

$$\frac{1}{N-M} \sum_{n=M}^{N-1} x_n = \frac{1}{N-M} \sum_{n=M}^{N-1} \left(\frac{1}{H} \sum_{h=1}^{H} x_{n+h} \right) + \Psi'_{M,N} = \Psi_{M,N} + \Psi'_{M,N},$$

where $\limsup_{N-M\to\infty} ||\Psi'_{M,N}|| = 0$. We will show that $\limsup_{N-M\to\infty} ||\Psi_{M,N}|| < \epsilon$. We have

$$\begin{aligned}
||\Psi_{M,N}||^2 &\le \frac{1}{N-M} \sum_{n=M}^{N-1} \left|\left| \frac{1}{H} \sum_{h=1}^{H} x_{n+h} \right|\right|^2 \\
&= \frac{1}{N-M} \sum_{n=M}^{N-1} \frac{1}{H^2} \sum_{h,k=1}^{H} \langle x_{n+h}, x_{n+k}\rangle \\
&= \sum_{r=-H}^{H} \frac{H-|r|}{H^2(N-M)} \sum_{u=M}^{N-1} \langle x_u, x_{u+r}\rangle + \Psi''_{M,N},
\end{aligned}$$

where $\Psi''_{M,N} \to 0$ as $N - M \to \infty$. By choice of H the last expression is less than ϵ when $N - M$ is sufficiently large.

$\square$

The following lemma will serve as a starting point in our proof of Theorem 2.1. Here and throughout (for the sake of convenience and without loss of generality) we take $L^\infty(X, \mathcal{A}, \mu)$, $L^2(X, \mathcal{A}, \mu)$, etc. to consist of real-valued functions only. Also, if $f, g \in L^2(X, \mathcal{A}, \mu)$, we will write $f \otimes g(x_1, x_2) = f(x_1)g(x_2)$.

Lemma 2.4 Suppose that an ergodic system $(X, \mathcal{A}, \mu, T)$ is a weakly mixing extension of $(Y, \mathcal{B}, \nu, S)$ and that $f, g \in L^\infty(X, \mathcal{A}, \mu)$. If either $E(f|\mathcal{B}) = 0$ or $E(g|\mathcal{B}) = 0$ then

$$D\text{-}\lim_{h\to\infty} \left|\left|E(fT^h g|\mathcal{B})\right|\right| = 0.$$

Proof. Note that it suffices to show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left|\left|E(fT^n g|\mathcal{B})\right|\right|^2 = 0.$$

We have

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left|\left|E(fT^n g|\mathcal{B})\right|\right|^2 \\ =&\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int (f \otimes f)\tilde{T}^n (g \otimes g) d\tilde{\mu} \\ =&\Big(\int (f \otimes f) d\tilde{\mu}\Big)\Big(\int (g \otimes g) d\tilde{\mu}\Big) \\ =&\Big(\int E(f|\mathcal{B})^2 \, d\nu\Big)\Big(\int E(g|\mathcal{B})^2 \, d\nu\Big) = 0. \end{aligned}$$

$\square$

We now describe the inductive technique whereby we will prove Theorem 2.1. This technique was also utilized in [B1]. First of all we may and shall assume without loss of generality that all polynomials are in $\mathbf{Z}[x]$. The induction will be on $P = \{p_1(x), \cdots, p_k(x)\} \subset \mathbf{Z}[x]$, the set of polynomials appearing in the theorem, using a partial ordering on the family of all such sets which we now describe.

Suppose that $P = \{p_1(x), \cdots, p_k(x)\} \subset \mathbf{Z}[x]$ is a finite set of pairwise distinct polynomials having zero constant term. We associate with P the (infinite) *weight vector* $(a_1, \cdots, a_m, \cdots)$ where, for each $i \in \mathbf{N}$, a_i is the number of distinct integers which occur as the leading coefficient of some polynomial from P which is of degree i. For example,

$$P = \{4x, 9x, 3x^2 - 5x, 3x^2 + 12x, -7x^2, 2x^4, 2x^4 + 3x, 2x^4 - 10x^3, 17x^5\}$$

has weight vector $(2, 2, 0, 1, 1, 0, 0, \cdots)$. (Notice that the weight vector ends in zeros.) If P has weight vector $(a_1, \cdots, a_m, \cdots)$ and Q has weight vector $(b_1, \cdots, b_m, \cdots)$, we write

$P < Q$, and say that "P precedes Q", if for some n, $a_n < b_n$, and $a_m = b_m$ for all $m > n$. This partial order comes from a well-ordering on the set of weight vectors

$$\big\{(a_1, \cdots, a_m, 0, 0, \cdots) : m \in \mathbf{N}, a_i \in \mathbf{N} \cup \{0\}, 1 \leq i \leq m, a_m > 0\big\},$$

and therefore gives rise to the inductive technique we are after. This technique, which we call *PET-induction*, works as follows: in order to prove any assertion $\mathcal{W}(P)$ for all finite sets of pairwise distinct polynomials P having zero constant term, it suffices to show that

(i) $\mathcal{W}(P)$ holds for all P having minimal weight vector $(1, 0, 0, \cdots)$, and

(ii) If $\mathcal{W}(P)$ holds for all $P < Q$ then $\mathcal{W}(Q)$ holds as well.

We now proceed to prove Theorem 2.1 via PET-induction.

Proof of Theorem 2.1. First we show that the conclusion holds if the weight vector of $P = \{p_1(x), \cdots, p_k(x)\}$ is $(1, 0, 0, \cdots)$. In this case $k = 1$ and $p_1(x) = jx$ for some non-zero integer j. We may write f_1 as the sum of two functions, one of which has zero conditional expectation over $\mathcal{B}$ and the other of which is $\mathcal{B}$-measurable, namely $f_1 = (f_1 - E(f_1|\mathcal{B})) + E(f_1|\mathcal{B})$. Since the conclusion obviously holds when f_1 is replaced by $E(f_1|\mathcal{B})$ (recall that E is idempotent), we need only show that the conclusion holds when f_1 is replaced by $\big(f_1 - E(f_1|\mathcal{B})\big)$, i.e. we may assume without loss of generality that $E(f_1|\mathcal{B}) = 0$. What we must show, then, is that

$$\lim_{N-M\to\infty} \left|\left| \frac{1}{N-M} \sum_{n=M}^{N-1} T^{jn} f_1 \right|\right| = 0.$$

However, by the uniform mean ergodic theorem,

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} T^{jn} f_1 = P f_1,$$

in norm, where P is the projection onto the set of T^j-invariant functions. Since $(X, \mathcal{A}, \mu, T)$ is a weakly mixing extension of $(Y, \mathcal{B}, \nu, S)$, and $E(f_1|\mathcal{B}) = 0$, we have $P f_1 = 0$. This completes the minimal weight vector case.

Suppose now that $Q = \{p_1(x), \cdots, p_k(x)\}$ is a family of non-zero, pairwise distinct polynomials having zero constant term, and that the conclusion holds for all P with $P < Q$. Reindexing if necessary, we may assume that $1 \leq \deg p_1 \leq \deg p_2 \leq \cdots \leq \deg p_k$. Let $f_1, \cdots, f_k \in L^\infty(X, \mathcal{A}, \mu)$. Suppose that $E(f_a|\mathcal{B}) = 0$ for some a, $1 \leq a \leq k$. We then must show that

$$\lim_{N-M\to\infty} \left|\left| \frac{1}{N-M} \sum_{n=M}^{N-1} \Big(\prod_{i=1}^{k} T^{p_i(n)} f_i \Big) \right|\right| = 0.$$

To see that the supposition is made without loss of generality, consider the identity

$$\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i = (a_1 - b_1) b_2 b_3 \cdots b_k + a_1 (a_2 - b_2) b_3 b_4 \cdots b_k + \cdots + a_1 a_2 \cdots a_{k-1} (a_k - b_k)$$

with $a_i = T^{p_i(n)}f_i$ and $b_i = S^{p_i(n)}E(f_i|\mathcal{B})$, noting that on the right hand side we have a sum of terms each of which has at least one factor with zero expectation relative to $\mathcal{B}$.

We use Lemma 2.3. Let $x_n = \prod_{i=1}^k T^{p_i(n)}f_i$. Then

$$\begin{aligned}
&\limsup_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1}\langle x_n, x_{n+h}\rangle \\
&= \limsup_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1}\int\left(\prod_{i=1}^k T^{p_i(n)}f_i\right)\left(\prod_{i=1}^k T^{p_i(n+h)}f_i\right)\,d\mu \\
&= \limsup_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1}\int f_1\left(\prod_{i=2}^k T^{p_i(n)-p_1(n)}f_i\right) \\
&\qquad\qquad \left(\prod_{i=1}^k T^{p_i(n+h)-p_1(n)-p_i(h)}(T^{p_i(h)}f_i)\right)\,d\mu.
\end{aligned} \tag{2}$$

For any $h\in\mathbf{Z}$ let

$$P_h = \{p_i(n)-p_1(n) : 2\le i\le k\}\cup\{p_i(n+h)-p_1(n)-p_i(h) : \deg p_i\ge 2, 1\le i\le k\}.$$

P_h consist of polynomials with zero constant term. Furthermore, the equivalence class of polynomials in Q with degree and leading coefficient the same as $p_1(n)$ has been annihilated in P_h. All other equivalence classes consisting of polynomials in Q of the same degree as $p_1(n)$ have been preserved (although the leading coefficients of these classes have changed). Equivalence classes of higher degree are completely intact. New equivalence classes may exist, but if so they will be of lesser degree than $p_1(n)$. It follows that $P_h < Q$. We now consider two cases:

Case 1. $\deg p_1\ge 2$. Then $\deg p_i\ge 2$, $1\le i\le k$, and one may check that for all h outside of some finite set, P_h consists of $2k-1$ distinct polynomials. For these h, we use our induction hypothesis for the validity of the theorem conclusion for the family P_h (utilizing weak convergence only) and continue from (2):

$$\begin{aligned}
&= \limsup_{N-M\to\infty} \frac{1}{N-M}\sum_{n=M}^{N-1}\int E(f_1|\mathcal{B})\left(\prod_{i=2}^k S^{p_i(n)-p_1(n)}E(f_i|\mathcal{B})\right) \\
&\qquad\qquad \left(\prod_{i=1}^k S^{p_i(n+h)-p_1(n)-p_i(h)}E(T^{p_i(h)}f_1|\mathcal{B})\right)\,d\nu = 0.
\end{aligned}$$

This since $E(f_a|\mathcal{B})=0$.

Case 2. $\deg p_1 = \deg p_2 = \cdots = \deg p_t = 1 < \deg p_{t+1}$. (Of course, if all the p_i are of degree 1 then $t=k$ and there is no p_{t+1}.) In this case $p_1(n+h)-p_1(n)-p_1(h)=0$, and $p_i(n+h)-p_1(n)-p_i(h) = p_i(n)-p_1(n)$, $2\le i\le t$, so that P_h will consist of $2k-t-1$

elements (again, excepting a finite set of h's for which other relations might hold). In this case we write $p_i(n) = c_i n$, $1 \le i \le t$, and proceed from (2):

$$= \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int E(f_1 T^{c_1 h} f_1|\mathcal{B}) \Big(\prod_{i=2}^{t} S^{p_i(n)-p_1(n)} E(f_i T^{c_i h} f_i|\mathcal{B}) \Big)$$

$$\Big(\prod_{i=t+1}^{k} S^{p_i(n)-p_1(n)} E(f_i|\mathcal{B}) S^{p_i(n+h)-p_1(n)-p_i(h)} E(T^{p_i(h)} f_i|\mathcal{B}) \Big) \, d\nu.$$

If $t+1 \le a \le k$, this is zero. If $1 \le a \le t$, however, it will still be at most

$$\Big(||E(f_a T^{c_a h} f_a|\mathcal{B})||_{L^2(Y,\mathcal{B},\nu)} \Big) \prod_{l \ne a} ||f_l||_\infty^2,$$

so that, by Lemma 2.4,

$$D\text{–}\lim_{h\to\infty} \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \langle x_n, x_{n+h} \rangle \le D\text{–}\lim_{h\to\infty} ||E(f_a T^{c_a h} f_a|\mathcal{B})||_{L^2(Y,\mathcal{B},\nu)} \cdot \prod_{l\ne a} ||f_l||_\infty^2 = 0.$$

In either of these two cases, Lemma 2.3 gives

$$\lim_{N-M\to\infty} \Big|\Big| \frac{1}{N-M} \sum_{n=M}^{N-1} \Big(\prod_{i=1}^{k} T^{p_i(n)} f_i \Big) \Big|\Big| = 0.$$

□

The following corollary is what we have been aiming for in this section.

Corollary 2.5 Suppose that $(X, \mathcal{A}, \mu, T)$ is an ergodic measure preserving system and denote by $(Y, \mathcal{A}_\eta, \nu, S)$ its maximal distal factor. If for all $A \in \mathcal{A}_\eta$ with $\nu(A) > 0$, we have

$$\limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \nu(A \cap S^{p_1(n)} A \cap \cdots \cap S^{p_k(n)} A)$$
$$= \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int 1_A S^{-p_1(n)} 1_A \cdots S^{-p_k(n)} 1_A \, d\nu > 0,$$

then for all $A \in \mathcal{A}$ with $\mu(A) > 0$,

$$\limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{p_1(n)} A \cap \cdots \cap T^{p_k(n)} A)$$
$$= \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int 1_A T^{-p_1(n)} 1_A \cdots T^{-p_k(n)} 1_A \, d\mu > 0.$$

Proof. By Theorem 1.4, $(X, \mathcal{A}, \mu, T)$ is either isomorphic to, or is a non-trivial weakly mixing extension of, $(Y, \mathcal{A}_\eta, \nu, S)$. In the former case there is nothing to prove, so we assume the latter. If $A \in \mathcal{A}$, then for some $\delta > 0$ we have

$$\nu(A_\delta) = \nu\big(\{y \in Y : \ \mu_y(A) \geq \delta\}\big) > 0.$$

We have $E(1_A|\mathcal{A}_\eta) > \delta 1_{A_\delta}$, so that by Theorem 2.1 (utilizing only weak convergence),

$$\begin{aligned}
&\limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int 1_A T^{-p_1(n)} 1_A \cdots T^{-p_k(n)} 1_A \, d\mu \\
=&\limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int 1_A S^{-p_1(n)} E(1_A|\mathcal{A}_\eta) \cdots S^{-p_k(n)} E(1_A|\mathcal{A}_\eta) \, d\nu \\
=&\limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int E(1_A|\mathcal{A}_\eta) S^{-p_1(n)} E(1_A|\mathcal{A}_\eta) \cdots S^{-p_k(n)} E(1_A|\mathcal{A}_\eta) \, d\nu \\
\geq&\delta^{k+1} \limsup_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int 1_{A_\delta} S^{-p_1(n)} 1_{A_\delta} \cdots S^{-p_k(n)} 1_{A_\delta} \, d\nu > 0.
\end{aligned}$$

□

3. Uniform polynomial Szemerédi theorem for distal systems.

According to Corollary 2.5, in order to establish Theorem 0.1 for an arbitrary system $(X, \mathcal{A}, \mu, T)$, it suffices to establish that the conclusion holds for its maximal distal factor $(X, \mathcal{A}_\eta, \mu, T)$. That is what we shall do in this section, using transfinite induction on the set of ordinals $\{\xi : \xi \leq \eta\}$ appearing in Theorem 1.4. As Theorem 0.1 trivially holds for the one point system, there are two cases to check: that the validity of the theorem is not lost in the passage to successor ordinals, namely, through compact extensions, and that the validity of the theorem is not lost in the passage to limit ordinals. Again, in this section we assume without loss of generality that all polynomials are in $\mathbf{Z}[x]$.

In order to show that the validity of Theorem 0.1 passes through compact extensions, we will use a combinatorial result, the polynomial Hales-Jewett theorem obtained in [BL2]. A special case of it is given as Theorem 3.1. If A is a set, we denote by $F(A)$ the set of all finite subsets of A. (Of course if A is finite this is just $\mathcal{P}(A)$.)

Theorem 3.1 Suppose numbers $k, d, r \in \mathbf{N}$ are given. Then there exists a number $N = N(k, d, r) \in \mathbf{N}$ having the property that whenever we have an r-cell partition

$$F\Big(\{1, \cdots, k\} \times \{1, \cdots, N\}^d\Big) = \bigcup_{i=1}^{r} C_i,$$

one of the sets C_i, $1 \leq i \leq r$ contains a configuration of the form

$$\Big\{A \cup (B \times S^d) : B \text{ ranges over subsets of } \{1, \cdots, k\}\Big\}$$

for some $A \subset \big(\{1,\cdots,k\}\times\{1,\cdots,N\}^d\big)$ and some non-empty set $S \subset \{1,\cdots,N\}$ satisfying

$$A \cap \big(\{1,\cdots,k\} \times S^d\big) = \emptyset.$$

Theorem 3.1 is a set-theoretic version and generalization of the polynomial van der Waerden theorem proved in [BL1]. To give some of the flavor of how one uses this theorem to help with polynomial dealings, we show first, as an example, how Theorem 3.1 guarantees that for any finite partition of $\mathbf{Z}$, $\mathbf{Z} = \bigcup_{i=1}^r D_i$, we may find i, $1 \le i \le r$ and $x, n \in \mathbf{Z}$, $n \neq 0$, with $\{x, x+n^2\} \subset D_i$. Namely, let $N = N(1,2,r)$ as in Theorem 3.1 and create a partition

$$F\Big(\{1\} \times \{1,\cdots,N\}^2\Big) = \bigcup_{i=1}^r C_i$$

according to the rule:

$$A \in C_i \text{ if and only if } \sum_{(1,t,s)\in A} ts \in D_i, 1 \le i \le r.$$

According to Theorem 3.1, there exists i, $1 \le i \le r$, $A \subset \big(\{1\} \times \{1,\cdots,N\}^d\big)$, and a non-empty set $S \subset \{1,\cdots,N\}$, satisfying $A \cap \big(\{1\} \times S^2\big) = \emptyset$, such that

$$\big\{A, A \cup \big(\{1\} \times S^2\big)\big\} \subset C_i.$$

Letting $x = \sum_{(1,t,s)\in A} ts$ and $n = \sum_{t\in S} t$, we then have $\{x, x+n^2\} \subset D_i$.

This example uses very little of the strength of Theorem 3.1, in particular it only needs the case $k = 1$ there. By considering general k, one may prove in a completely analogous fashion that for any finite set of polynomials $p_1(x), \cdots, p_k(x) \in \mathbf{Z}[x]$ with $p_i(0) = 0$, $1 \le k$, and any finite partition of $\mathbf{Z}$, $\mathbf{Z} = \bigcup_{i=1}^r D_i$, one may find i, $1 \le i \le r$, and $x, n \in \mathbf{N}$, $n \neq 0$, with $\{x, x+p_1(n), \cdots, x+p_k(n)\} \subset D_i$. The following consequence of Theorem 3.1 is all we shall need from it and is a still further elaboration of the method introduced in the previous paragraph. In an appendix we will show how it can be derived from Theorem 3.1.

Corollary 3.2 Suppose that $r, k, t \in \mathbf{N}$ and that

$$p_1(x_1,\cdots,x_t), \cdots, p_k(x_1,\cdots,x_t) \in \mathbf{Z}[x_1,\cdots,x_t]$$

with $p_i(0,\cdots,0) = 0$, $1 \le i \le k$. Then there exist numbers $w, N \in \mathbf{N}$, and a set of polynomials

$$Q = \big\{q_1(y_1,\cdots,y_N), \cdots, q_w(y_1,\cdots,y_N)\big\} \subset \mathbf{Z}[y_1,\cdots,y_N]$$

with $q_i(0,\cdots,0) = 0$, $1 \le i \le w$, such that for any r-cell partition $Q = \bigcup_{i=1}^r C_i$, there exist some i, $1 \le i \le r$, $q \in Q$, and non-empty, pairwise disjoint subsets $S_1, \cdots, S_t \subset \{1,\cdots,N\}$, such that, under the symbolic substitution $x_m = \sum_{n\in S_m} y_n$, $1 \le m \le t$, we have

$$\Big\{q(y_1,\cdots,y_N), q(y_1,\cdots,y_N) + p_1(x_1,\cdots,x_t), \cdots, q(y_1,\cdots,y_N) + p_k(x_1,\cdots,x_t)\Big\} \subset C_i.$$

We now make some definitions.

Definition 3.3 A subset $E \subset \mathbf{Z}$ will be called *thick* if for every $M \in \mathbf{N}$, there exists $a \in \mathbf{Z}$ such that $\{a, a+1, a+2, \cdots, a+M\} \subset E$.

Note that a set is thick if and only if it contains arbitrarily large intervals, and a set is syndetic if and only if it intersects every thick set non-trivially.

The following definition is tailored to fit into the framework of the usage of Corollary 3.2 in passing to compact extensions. Here for any $n_1, \cdots, n_t \in \mathbf{N}$ (or any additive group) we write

$$FS(n_1, \cdots, n_t) = \big\{n_{i_1} + \cdots + n_{i_m} : 1 \leq m \leq t, 1 \leq i_1 < \cdots < i_m \leq t\big\}.$$

Definition 3.4 Suppose $(X, \mathcal{A}, \mu, T)$ is an invertible measure preserving system and that $\mathcal{B} \subset \mathcal{A}$ is a complete T-invariant sub-σ-algebra. $\mathcal{B}$ is said to have the *PSZ property* if for every $A \in \mathcal{B}$ with $\mu(A) > 0$, $t \in \mathbf{Z}$, and polynomials $p_1(x_1, \cdots x_t), \cdots, p_k(x_1, \cdots, x_t) \in \mathbf{Z}[x_1, \cdots, x_t]$ having zero constant term, there exists $\delta > 0$ such that in every thick set $E \subset \mathbf{Z}$, there exist $n_1, \cdots, n_t \in \mathbf{Z}$ such that $FS(n_1, \cdots, n_t) \subset E$ and

$$\mu(A \cap T^{p_1(n_1, \cdots, n_t)} A \cap \cdots \cap T^{p_k(n_1, \cdots, n_t)} A) > \delta.$$

The case $t = 1$, in particular, gives some $\delta > 0$ for which the set

$$\{n \in \mathbf{Z} : \ \mu(A \cap T^{p_1(n)} A \cap \cdots \cap T^{p_k(n)} A) > \delta\}$$

is syndetic, which insures that

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

Therefore, in light of Corollary 2.5, we will have established Theorem 0.1 if we are able to prove that the maximal distal factor of any system has the PSZ property. This is exactly what we shall do in the remainder of this section. The reader may wonder why this definition is stronger than appears necessary, that is, why we choose to deal with polynomials of many variables. The reason is that our method of proof requires the PSZ property in this strength in order to preserve itself under compact extensions. Corollary 3.2 is the key to this method, as we shall now see.

Theorem 3.5 Suppose that $(X, \mathcal{A}, \mu, T)$ is an ergodic measure preserving system and that $\mathcal{B} \subset \mathcal{A}$ is a complete, T-invariant sub-σ-algebra having the PSZ property. If $(X, \mathcal{A}, \mu, T)$ is a compact extension of the factor $(Y, \mathcal{B}, \nu, S)$ determined by $\mathcal{B}$, then $\mathcal{A}$ has the PSZ property as well.

Proof. Suppose that $A \in \mathcal{A}$, $\mu(A) > 0$. By Proposition 1.3 there exists a subset $A' \subset A$, $\mu(A') > 0$, such that $1_{A'} \in AP$. Therefore we may assume without loss of generality that $f = 1_A \in AP$. Suppose that $t, k \in \mathbf{N}$ and that

$$p_1(x_1, \cdots, x_t), \cdots, p_k(x_1, \cdots, x_t) \in \mathbf{Z}[x_1, \cdots, x_t]$$

have zero constant term. There exists some $c > 0$ and a set $B \in \mathcal{B}$, $\nu(B) > 0$, such that for all $y \in B$, $\mu_y(A) > c$. Let $\epsilon = \sqrt{\frac{c}{8k}}$. Since $1_A \in AP$, there exist functions $g_1, \cdots, g_r \in L^2(X, \mathcal{A}, \mu)$ having the property that for any $n \in \mathbf{N}$, and a.e. $y \in Y$, there exists $s = s(n, y)$, $1 \leq s \leq r$, such that $||T^n f - g_s||_y < \epsilon$. For these numbers r, k, t and polynomials p_i, let $w, N \in \mathbf{N}$ and

$$Q = \{q_1(y_1, \cdots, y_N), \cdots, q_w(y_1, \cdots, y_N)\} \subset \mathbf{Z}[y_1, \cdots, y_N],\ q_i(0, \cdots, 0) = 0,\ 1 \leq i \leq w$$

have the property that for any r-cell partition $Q = \bigcup_{i=1}^r C_i$, there exists i, $1 \leq i \leq r$, $q \in Q$, and pairwise disjoint subsets $S_1, \cdots, S_t \subset \{1, \cdots, N\}$ such that substituting $x_m = \sum_{n \in S_m} y_n$, $1 \leq m \leq t$, we have

$$\big\{q(y_1, \cdots, y_N), q(y_1, \cdots, y_N) - p_1(x_1, \cdots, x_t), \cdots, q(y_1, \cdots, y_N) - p_k(x_1, \cdots, x_t)\big\} \subset C_i.$$

(This is possible by Corollary 3.2.)

Since $\mathcal{B}$ has the PSZ property, there exists $\eta > 0$ such that for every thick set $E \subset \mathbf{Z}$, there exists $u_1, \cdots, u_N \in \mathbf{Z}$ such that $FS(u_1, \cdots, u_N) \subset E$ and

$$\nu\big(B \cap S^{q_1(u_1, \cdots, u_N)} B \cap \cdots \cap S^{q_w(u_1, \cdots, u_N)} B\big) > \eta.$$

Let D be the number of ways of choosing t non-empty, pairwise disjoint sets $S_1, \cdots, S_t \subset \{1, \cdots, N\}$, and set $\delta = \frac{c\eta}{2D}$. We want to show that in any thick set $E \subset \mathbf{Z}$ there exist $n_1, \cdots, n_t \in \mathbf{Z}$ such that $FS(n_1, \cdots, n_t) \subset E$ and

$$\mu\big(A \cap T^{p_1(n_1, \cdots, n_t)} A \cap \cdots \cap T^{p_k(n_1, \cdots, n_t)} A\big) > \delta.$$

Let E be any thick set. There exist $u_1, \cdots, u_N \in \mathbf{Z}$ such that $FS(u_1, \cdots, u_N) \subset E$ and

$$\nu\big(B \cap S^{q_1(u_1, \cdots, u_N)} B \cap \cdots \cap S^{q_w(u_1, \cdots, u_N)} B\big) > \eta.$$

Pick any $y \in \big(B \cap S^{q_1(u_1, \cdots, u_N)} B \cap \cdots \cap S^{q_w(u_1, \cdots, u_N)} B\big)$. Form an r-cell partition of Q, $Q = \bigcup_{i=1}^r C_i$, by the rule $q_a(y_1, \cdots, y_N) \in C_i$ if and only if $s\big(q_a(u_1, \cdots, u_N), y\big) = i$, $1 \leq a \leq w$. In particular, if $q_a \in C_i$ then $||T^{q_a(u_1, \cdots, u_N)} f - g_i||_y < \epsilon$. For this partition, there exists some i, $1 \leq i \leq r$, some $q \in Q$, and pairwise disjoint subsets $S_1, \cdots, S_t \subset \{1, \cdots, N\}$ such that, under the substitution $x_m = \sum_{n \in S_m} y_n$, $1 \leq m \leq t$, we have

$$\big\{q(y_1, \cdots, y_N), q(y_1, \cdots, y_N) - p_1(x_1, \cdots, x_t), \cdots, q(y_1, \cdots, y_N) - p_k(x_1, \cdots, x_t)\big\} \subset C_i.$$

In particular, making the analogous substitutions $n_m = \sum_{n \in S_m} u_n$, $1 \leq m \leq t$, we have $FS(n_1, \cdots, n_t) \subset E$, and furthermore, we have, setting $p_0(x_1, \cdots, x_t) = 0$,

$$\Big|\Big|T^{q(u_1, \cdots, u_N) - p_b(n_1, \cdots, n_t)} f - g_i\Big|\Big|_y < \epsilon; \quad 0 \leq b \leq k.$$

Setting $\tilde{y} = S^{-q(u_1, \cdots, u_N)} y$, we have

$$\Big|\Big|T^{-p_b(n_1, \cdots, n_t)} f - T^{-q(u_1, \cdots, u_N)} g_i\Big|\Big|_{\tilde{y}} < \epsilon; \quad 0 \leq b \leq k.$$

In particular, since this holds for $b=0$, we have by the triangle inequality

$$\left|\left|T^{-p_b(n_1,\cdots,n_t)}f - f\right|\right|_{\tilde{y}} < 2\epsilon, \quad 1 \le b \le k.$$

It follows that

$$\mu_{\tilde{y}}\big(A \setminus T^{p_b(n_1,\cdots,n_t)}A\big) \le \left|\left|T^{-p_b(n_1,\cdots,n_t)}f - f\right|\right|_{\tilde{y}}^2 \le 4\epsilon^2; \quad 1 \le b \le k.$$

Moreover, $\tilde{y} \in B$, so that $\mu_{\tilde{y}}(A) \ge c$, therefore, since $\epsilon = \sqrt{\frac{c}{8k}}$,

$$\mu_{\tilde{y}}\big(A \cap T^{p_1(n_1,\cdots,n_t)}A \cap \cdots \cap T^{p_k(n_1,\cdots,n_t)}A\big) \ge c - 4k\epsilon^2 = \frac{c}{2}.$$

The sets $S_1, \cdots, S_t$ depend measurably on y, and therefore the numbers $n_1, \cdots, n_t$ are measurable functions of y defined on the set $\big(B \cap S^{q_1(u_1,\cdots,u_N)}B \cap \cdots \cap S^{q_w(u_1,\cdots,u_N)}B\big)$, which, recall, is of measure greater than η. Hence, as there are only D choices possible for $S_1, \cdots, S_t$, we may assume that for all $y \in H$, where $H \in \mathcal{B}$ satisfies $\nu(H) > \frac{\eta}{D}$, $n_1, \cdots, n_t$ are constant. For this choice of $n_1, \cdots, n_t$ we have

$$\mu\big(A \cap T^{p_1(n_1,\cdots,n_t)}A \cap \cdots \cap T^{p_t(n_1,\cdots,n_t)}A\big) \ge \frac{c}{2}\nu(H) > \frac{c\eta}{2D} = \delta.$$

□

Looking again at Theorem 1.4, we see that, having proved that the PSZ property is preserved under compact extensions, we have left only to prove that, given a totally ordered chain of T-invariant sub-σ-algebras with the PSZ property, the completion of the σ-algebra generated by the chain again possesses the PSZ property.

Proposition 3.6 Suppose that $(X, \mathcal{A}, \mu, T)$ is a measure preserving system and that $\mathcal{A}_\xi$ is a totally ordered chain of sub-σ-algebras of $\mathcal{A}$ having the PSZ property. If $\bigcup_\xi \mathcal{A}_\xi$ is dense in $\mathcal{A}$, that is, if $\mathcal{A}$ is the completion of the σ-algebra generated by $\bigcup_\xi \mathcal{A}_\xi$, then $\mathcal{A}$ has the PSZ property.

Proof. Suppose $A \in \mathcal{A}$, $\mu(A) > 0$, $t, k \in \mathbf{N}$, and that

$$p_1(x_1, \cdots, x_t), \cdots, p_k(x_1, \cdots, x_t) \in \mathbf{Z}[x_1, \cdots, x_t]$$

are polynomials with zero constant term. There exists ξ and $B \in \mathcal{A}_\xi$ such that

$$\mu\big((A \setminus B) \cup (B \setminus A)\big) \le \frac{\mu(A)}{4(k+1)}.$$

Let $\int d\mu = \int_Y \int_X d\mu_y d\nu(y)$ be the decomposition of the measure μ over the factor $\mathcal{A}_\xi$. Let $C = \{y \in B : \ \mu_y(A) \ge 1 - \frac{1}{2(k+1)}\}$. It is easy to see that $\nu(C) > 0$. Since $\mathcal{A}_\xi$ has the PSZ property, there exists some $\alpha > 0$ having the property that in any thick set E we may find $n_1, \cdots, n_t \in \mathbf{Z}$ such that $FS(n_1, \cdots, n_t) \subset E$ and

$$\nu\big(C \cap S^{p_1(n_1,\cdots,n_t)}C \cap \cdots \cap S^{p_k(n_1,\cdots,n_t)}C\big) > \alpha. \tag{3}$$

Set $\delta = \frac{\alpha}{2}$ and let E be any thick set. Find $n_1, \cdots, n_t \in \mathbf{Z}$ satisfying (3) and with $FS(n_1, \cdots, n_t) \subset E$. For any $y \in \big(C \cap S^{p_1(n_1,\cdots,n_t)}C \cap \cdots \cap S^{p_k(n_1,\cdots,n_t)}C\big)$ we have $\mu_y(A), \mu_y(T^{p_1(n_1,\cdots,n_t)}A), \cdots, \mu_y(T^{p_k(n_1,\cdots,n_t)}A)$ all not less than $1 - \frac{1}{2(k+1)}$, from which it follows that

$$\mu_y\big(A \cap T^{p_1(n_1,\cdots,n_t)}A) \cap \cdots \cap T^{p_k(n_1,\cdots,n_t)}A\big) \geq \frac{1}{2}.$$

Therefore,

$$\mu\big(A \cap T^{p_1(n_1,\cdots,n_t)}A) \cap \cdots \cap T^{p_k(n_1,\cdots,n_t)}A\big) > \frac{\alpha}{2} = \delta.$$

□

4. Appendix: Proof of Theorem 3.2.

We will now derive Corollary 3.2 from Theorem 3.1. First, we derive from Theorem 3.1 its natural "multi-parameter" version.

Proposition 4.1 Suppose $k, d, r, t \in \mathbf{N}$ are given. Then there exists a number $N = N(k, d, r) \in \mathbf{N}$ having the property that for any r-cell partition

$$F\big(\{1, 2, \cdots, k\} \times \{1, 2, \cdots, tN\}^d\big) = \bigcup_{i=1}^{r} C_i,$$

one of the cells C_i, $1 \leq i \leq r$ contains a configuration of the form

$$\Big\{A \cup \big(B \times (S_1 \cup \cdots \cup S_t)^d\big) : B \subset \{1, \cdots, k\}\Big\},$$

where $A \subset \big(\{1, \cdots, k\} \times \{1, \cdots, tN\}^d\big)$ and $\emptyset \neq S_j \subset \{(j-1)N + 1, \cdots, jN\}$, $1 \leq j \leq t$ satisfy

$$A \cap \Big(\{1, \cdots, k\} \times (S_1 \cup \cdots \cup S_t)^d\Big) = \emptyset.$$

Proof. Let $N = N(k, d, r)$ be as in Theorem 3.1. Given any partition

$$F\big(\{1, 2, \cdots, k\} \times \{1, 2, \cdots, tN\}^d\big) = \bigcup_{i=1}^{r} C_i,$$

we will construct a partition

$$F\big(\{1, 2, \cdots, k\} \times \{1, 2, \cdots, N\}^d\big) = \bigcup_{i=1}^{r} D_i$$

in the following way: given any set $U \subset \big(\{1, 2, \cdots, k\} \times \{1, 2, \cdots, N\}^d\big)$, let

$$U_j = \big\{v \in \{1, \cdots, N\}^d : (j, v) \in U\big\}, 1 \leq j \leq k.$$

Also, for any set $E \subset \{1, \cdots, N\}^d$, let $\gamma(E) \subset \{1, \cdots, tN\}^d$ be the set which is obtained by taking the union of t^d shifts of E,

$$\gamma(E) = \bigcup_{0 \le i_1, \cdots, i_d < t} \big(E + (i_1 N, \cdots, i_d N)\big).$$

(Notice, in particular, that $\gamma(\{1, \cdots, N\}^d) = \{1, \cdots, tN\}^d$.) Now let

$$\gamma(U) = \bigcup_{j=1}^{k} \{j\} \times \gamma(U_j),$$

and put $U \in D_i$ if and only if $\gamma(U) \in C_i$. According to the property whereby N was chosen, one of the sets D_i, $1 \le i \le r$, say D_j, contains a configuration of the form

$$\Big\{ H \cup (B \times S^d) : B \subset \{1, \cdots, k\} \Big\},$$

for some $H \subset \big(\{1, \cdots, k\} \times \{1, \cdots, N\}^d\big)$ and some non-empty $S \subset \{1, \cdots, N\}$ satisfying

$$H \cap \big(\{1, \cdots, k\} \times S^d\big) = \emptyset.$$

Let $S_i = S + (i-1)N$, $1 \le i \le t$, and put $A = \gamma(H)$. Then $\emptyset \ne S_i \subset \{(i-1)N+1, \cdots, iN\}$, $1 \le i \le t$, and one may check that for every $B \subset \{1, \cdots, k\}$,

$$\gamma\big(H \cup (B \times S^d)\big) = A \cup \big(B \times (S_1 \cup \cdots \cup S_t)^d\big).$$

It follows that

$$\Big\{ A \cup \big(B \times (S_1 \cup \cdots \cup S_t)^d\big) : B \subset \{1, \cdots, k\} \Big\} \subset C_j.$$

Furthermore, $A \cap \big(\{1, \cdots, k\} \times (S_1 \cup \cdots \cup S_t)^d\big) = \emptyset$, so we are done.

$\square$

We are now in position to prove Corollary 3.2.

Proof of Corollary 3.2. Let d be the maximum degree of the polynomials p_j, $1 \le j \le k$, and let $N = N(k, d, r)$ be as in Proposition 4.1 above. We claim there exists a map

$$\varphi : \ F\big(\{1, \cdots, k\} \times \{1, \cdots, tN\}^d\big) \to \mathbf{Z}[y_1, \cdots, y_{tN}]$$

satisfying $\varphi(A \cup B) = \varphi(A) + \varphi(B)$ whenever $A \cap B = \emptyset$, and such that

$$\begin{aligned} &\varphi\big(\{j\} \times \{m_1^{(1)}, \cdots, m_{s_1}^{(1)}, \cdots, m_1^{(t)}, \cdots, m_{s_t}^{(t)}\}^d\big) \\ =&p_j\big(y_{m_1{}^{(1)}} + \cdots + y_{m_{s_1}{}^{(1)}}, \cdots, y_{m_1{}^{(t)}} + \cdots + y_{m_{s_t}{}^{(t)}}\big) \end{aligned}$$

whenever $1 \le j \le k$ and $\{m_1^{(l)}, \cdots, m_{s_l}^{(l)}\} \subset \{(l-1)N+1, \cdots, lN\}$, $1 \le l \le t$. We now proceed to define the function φ on singletons.

Fix j, $1 \le j \le k$. There exist polynomials

$$\alpha_{i_1,\cdots,i_t}(u_1, u_2, \cdots, u_{i_1+\cdots+i_t}) \in \mathbf{Z}[u_1, \cdots, u_{i_1+\cdots+i_t}], \ \ 1 \le i_1 + \cdots + i_t \le d,$$

such that whenever $\{m_1^{(l)}, \cdots, m_{s_l}^{(l)}\} \subset \{(l-1)N+1, \cdots, lN\}$, $1 \le l \le t$, we have

$$\begin{aligned} & p_j(u_{m_1^{(1)}} + \cdots + u_{m_{s_1}^{(1)}}, \cdots, u_{m_1^{(t)}} + \cdots + u_{m_{s_t}^{(t)}}) \\ = & \sum_{1 \le i_1+\cdots+i_t \le d} \ \sum_{\{l_1^{(l)}, \cdots, l_{i_l}^{(l)}\} \subset \{m_1^{(l)}, \cdots, m_{s_l}^{(l)}\}} \alpha_{i_1,\cdots,i_t}(u_{l_1^{(1)}}, \cdots, u_{l_{i_1}^{(1)}}, \cdots, u_{l_1^{(t)}}, \cdots, u_{l_{i_t}^{(t)}}). \end{aligned}$$

For example, say that $p_j(x_1, x_2) = x_1^2 x_2 + x_1^2 + x_2$. Then $\alpha_{1,0}(u_1) = u_1^2$, $\alpha_{0,1}(u_1) = u_1$, $\alpha_{2,0}(u_1, u_2) = 2u_1u_2$, $\alpha_{1,1}(u_1, u_2) = u_1^2 u_2$, and $\alpha_{2,1}(u_1, u_2, u_3) = 2u_1u_2u_3$. Now for each t-tuple of subsets

$$\{l_1^{(l)}, \cdots, l_{i_l}^{(l)}\} \subset \{(l-1)N+1, \cdots, lN\}, \ \ 1 \le l \le t, \ \ 1 \le i_1 + \cdots + i_t \le d,$$

we pick exactly one representative point $(j, a_1, \cdots, a_d) \in \big(\{j\} \times \{1, \cdots, tN\}^d\big)$ which satisfies

$$\{a_1, \cdots, a_d\} = \bigcup_{l=1}^{t} \{l_1^{(l)}, \cdots, l_{i_l}^{(l)}\},$$

and define

$$\varphi\Big(\big\{(j, a_1, \cdots, a_d)\big\}\Big) = \alpha_{i_1,\cdots,i_t}(y_{l_1^{(1)}}, \cdots, y_{l_{i_1}^{(1)}}, \cdots, y_{l_1^{(t)}}, \cdots, y_{l_{i_t}^{(t)}}) \in \mathbf{Z}[y_1, \cdots, y_{tN}].$$

φ is defined to be zero on all singletons in $\big(\{j\} \times \{1, \cdots, tN\}^d\big)$ not so chosen.

Repeat these steps for $1 \le j \le k$. Now extend φ to all of $F\big(\{1, \cdots, k\} \times \{1, \cdots, tN\}^d\big)$ according to the additivity condition we require of φ.

Now, whenever $1 \le j \le k$, and $\{m_1^{(l)}, \cdots, m_{s_l}^{(l)}\} \subset \{(l-1)N+1, \cdots, lN\}$, $1 \le l \le t$, we have

$$\begin{aligned} & p_j(y_{m_1^{(1)}} + \cdots + y_{m_{s_1}^{(1)}}, \cdots, y_{m_1^{(t)}} + \cdots + y_{m_{s_t}^{(t)}}) \\ = & \sum_{1 \le i_1+\cdots+i_t \le d} \ \sum_{\{l_1^{(l)}, \cdots, l_{i_l}^{(l)}\} \subset \{m_1^{(l)}, \cdots, m_{s_l}^{(l)}\}} \alpha_{i_1,\cdots,i_t}(y_{l_1^{(1)}}, \cdots, y_{l_{i_1}^{(1)}}, \cdots, y_{l_1^{(t)}}, \cdots, y_{l_{i_t}^{(t)}}) \\ = & \sum_{(a_1, \cdots, a_d) \in \{m_1^{(1)}, \cdots, m_{s_1}^{(1)}, \cdots, m_1^{(t)}, \cdots, m_{s_t}^{(t)}\}^d} \varphi\Big(\big\{(j, a_1, \cdots, a_d)\big\}\Big) \\ = & \varphi\big(\{j\} \times \{m_1^{(1)}, \cdots, m_{s_1}^{(1)}, \cdots, m_1^{(t)}, \cdots, m_{s_t}^{(t)}\}^d\big). \end{aligned}$$

Let $w = k(tN)^d$ and let $\big(q_i(y_1, \cdots, y_{tN})\big)_{i=1}^{w}$ be the images of the singletons under φ, so that

$$\mathcal{F} = \varphi\Big(F\big(\{1, \cdots, k\} \times \{1, \cdots, tN\}^d\big)\Big) = FS\big(q_i(y_1, \cdots, y_{tN})\big)_{i=1}^{w}.$$

If now $\mathcal{F} = \bigcup_{i=1}^{r} C_i$, then

$$F\big(\{1,\cdots,k\} \times \{1,\cdots,tN\}^d\big) = \bigcup_{i=1}^{r} \varphi^{-1}(C_i),$$

and by Proposition 4.1, there exists i, $1 \le i \le r$, $A \in \varphi^{-1}(C_i)$, and sets

$$S_l \subset \{(l-1)N+1,\cdots,lN\},\ \ 1 \le l \le t,$$

having the property that for all j, $1 \le j \le k$,

$$A \cap \Big(\{j\} \times \big(S_1 \cup \cdots \cup S_t\big)^d\Big) = \emptyset,\ \ A \cup \Big(\{j\} \times \big(S_1 \cup \cdots \cup S_t\big)^d\Big) \in \varphi^{-1}(C_i).$$

Let $q(y_1,\cdots,y_{tN}) = \varphi(A)$. Put $p_0(x_1,\cdots,x_t) = 0$. Then for all j, $0 \le j \le k$, we have, substituting $x_l = \sum_{n \in S_l} y_n$, $1 \le l \le t$,

$$q(y_1,\cdots,y_{tN}) + p_j(x_1,\cdots,x_t) = \varphi\Big(A \cup \big(\{j\} \times (S_1 \cup \cdots \cup S_t)^d\big)\Big) \in C_i.$$

□

REFERENCES

[B1] V. Bergelson, Weakly mixing PET, *Ergodic Theory and Dynamical Systems* **7** (1987), 337-349.

[B2] V. Bergelson, Ergodic Ramsey theory–an update.

[BL1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorem, *Journal of AMS* (to appear).

[BL2] V. Bergelson and A. Leibman, Polynomial Hales-Jewett theorem, preprint.

[F1] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.* **31** (1977), 204-256.

[F2] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, 1981.

[FKO] H. Furstenberg, Y. Katznelson and D. Ornstein, The ergodic theoretical proof of Szemerédi's theorem, *Bull. Amer. Math. Soc.* **7** (1982), 527-552.