Idempotent Ultrafilters, Multiple Weak Mixing and Szemerédi's Theorem for Generalized Polynomials

V. Bergelson and R. McCutcheon

It is possible to formulate the polynomial Szemerédi theorem as follows: Let $q_i(x) \in \mathbf{Q}[x]$ with $q_i(\mathbf{Z}) \subset \mathbf{Z}$, $1 \leq i \leq k$. If $E \subset \mathbf{N}$ has positive upper density then there are $a, n \in \mathbf{N}$ such that $\{a, a+q_1(n)-q_1(0), a+q_k(n)-q_k(0)\} \subset E$. Using methods of abstract ergodic theory, topological algebra in $\beta\mathbf{N}$, and some recently-obtained knowledge concerning the relationship between translations on nilmanifolds and the distribution of bounded generalized polynomials, we prove, among other results, the following extension, valid for generalized polynomials (functions obtained from regular polynomials via iterated use of the floor function). Let $q_i(x)$ be generalized polynomials, $1 \leq i \leq k$, and let $p \in \beta\mathbf{N}$ be an idempotent ultrafilter all of whose members have positive upper Banach density. Then there exist constants c_i, $1 \leq i \leq k$, such that if $E \subset \mathbf{Z}$ has positive upper Banach density then the set $\big\{n \in \mathbf{N} : \exists\, a \in \mathbf{Z} \text{ with } \{a, a+q_1(n)-c_1, a+q_k(n)-c_k\} \subset E\big\}$ belongs to p. As part of the proof, we also obtain a new ultrafilter polynomial ergodic theorem characterizing weak mixing.

1. Introduction.

1.1. In this paper we establish new results concerning multiple mixing for weakly mixing measure preserving systems. These results, in turn, lead to new extensions of Szemerédi's theorem on arithmetic progressions, while at the same time allowing for simplified presentations of some known extensions. We employ methods introduced in [BM3] involving *idempotent ultrafilters* and *IP systems*; here, however, we shall concern ourselves with polynomial structures as well, specifically *VIP systems* and *generalized polynomials*. Because some of this material (the methods, not the results) is semi-esoteric, we offer a brief review of it in this introductory section.

Furstenberg's ergodic-theoretic approach to Szemerédi's theorem.

1.2. Given a set $E \subset \mathbf{N}$, its *upper Banach density* is given by (Definition 2.20 below)

$$d^*(E) = \limsup_{N-M\to\infty} \frac{|E \cap \{M, M+1, \ldots, N-1\}|}{N-M}.$$

Szemerédi's theorem ([S]) states that if $d^*(E) > 0$ then E contains arbitrarily long arithmetic progressions. Furstenberg gave a new proof of this result based on a correspondence principle linking density combinatorics with (multiple) recurrence for measure preserving systems and the following theorem.

Convention: All measure preserving systems we consider in this paper will be assumed to be invertible probability measure preserving systems.

1.3. Theorem ([F1]). For any $k \in \mathbf{N}$, any measure preserving system $(X, \mathcal{A}, \mu, T)$ and any $A \in \mathcal{A}$ with $\mu(A) > 0$, one has

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

Combining the above theorem with a correspondence principle of Furstenberg, one gets the (at the time new) fact that if $d^*(E) > 0$ then the set of difference of progressions in E, that is, the set of n for which there exists a for which $\{a, a+n, \ldots, a+kn\} \subset E$, is *syndetic* (meaning it intersects any long enough interval).

Weakly mixing transformations and polynomial extensions.

1.4. Furstenberg's proof of Szemerédi's theorem utilized a new structure theorem for measure preserving systems based on *weak mixing* and its complementary notion, *compactness.* (An invertible probability measure preserving system $(X, \mathcal{A}, \mu, T)$ is weakly mixing if the unitary operator $T : L^2(X) \to L^2(X)$ defined by $Tf(x) = f(Tx)$ has no non-constant eigenfunctions.) The following theorem is representative of the important contribution weak mixing makes to proofs of this kind.

Theorem ([F1]). Let $(X, \mathcal{A}, \mu, T)$ be an invertible weakly mixing measure preserving system. For any $f_0, f_1 \ldots, f_k \in L^\infty(X)$, one has

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \int f_0 T^n f_1 T^{2n} f_2 \cdots T^{kn} f_k \, d\mu = \prod_{i=0}^{k} \int f_i \, d\mu.$$

Indeed, as an important extreme case of Theorem 1.3, one can derive from the above result that for $A_0, A_1, \ldots, A_k \in \mathcal{A}$, one has

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} |\mu(A_0 \cap T^{-n}A_1 \cap \cdots \cap T^{-kn}A_k) - \prod_{i=0}^{k} \mu(A_i)| = 0,$$

which in turn implies that for any $\epsilon > 0$ the set

$$R = \{n \in \mathbf{Z} : |\mu(A_0 \cap T^{-n}A_1 \cap \cdots \cap T^{-kn}A_k) - \prod_{i=0}^{k} \mu(A_i)| < \epsilon\}$$

has *uniform* Banach density one (meaning its complement has upper Banach density zero). An extension was proved in [B2]:

1.5. Theorem. Let $(X, \mathcal{A}, \mu, T)$ be a weakly mixing system, let $k \in \mathbf{N}$ and let $p_i \in \mathbf{Z}[n]$ be polynomials such that no p_i and no $p_i - p_j$ is constant, $1 \le i \ne j \le k$. Then for any $f_1, \ldots, f_k \in L^\infty(X)$,

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

The above theorem again provided a proof for a key extreme case of a polynomial extension of Szemerédi's theorem obtained in [BL1]. (We mention that results of [BL1] actually extend the multidimensional Szemerédi theorem obtained in [FK1]. In this paper, we are limiting ourselves to configurations in $\mathbf{Z}$.) A refinement of a special case of [BL1] is given by the following formulation, from [BM1]:

1.6. Theorem. For any invertible measure preserving system $(X, \mathcal{A}, \mu, T)$, any $k \in \mathbf{N}$, polynomials $q_i \in \mathbf{Z}[n]$ with $q_i(0) = 0$, $1 \leq i \leq k$, and any $A \in \mathcal{A}$ with $\mu(A) > 0$, one has

$$\liminf_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \mu(A \cap T^{q_1(n)}A \cap \cdots \cap T^{q_k(n)}A) > 0.$$

The conclusion of Theorem 1.6 implies in particular that the set

$$R = \left\{n \in \mathbf{Z} : \mu(A \cap T^{q_1(n)}A \cap \cdots \cap T^{q_k(n)}A) > 0\right\}$$

is syndetic. It is natural now to ask how much of the "largeness" of sets having the form of R is captured by the predicate "syndetic." At first blush, perhaps quite a bit; upon further reflection, however, one sees that the family of syndetic sets lacks the filter property, which the family of sets R can be checked to have quite easily (consider product systems). In this paper, we upgrade Theorem 1.6 first of all by making it apply to a wider class of functions q_i; even more significantly, however, we obtain, for a natural portion of this class, an improvement (having the filter property) upon syndeticity for the set R.

Generalized polynomials and weak mixing.

1.7. We denote the floor function by $[\cdot]$; that is, $[x]$ is the greatest integer not exceeding x. We write $\{\cdot\}$ for fractional part; that is, $\{x\} = x - [x]$, and $\langle x \rangle = |x - [x + \frac{1}{2}]|$ for the distance from x to the nearest integer. We denote by $\mathcal{G}$ the smallest family of functions $\mathbf{N} \to \mathbf{Z}$ containing $\mathbf{Z}[n]$ that forms an algebra under addition and multiplication and having the property that for every $f_1, \ldots, f_r \in \mathcal{G}$ and $c_1, \ldots, c_r \in \mathbf{R}$, $[\sum_{i=1}^r c_i f_i] \in \mathcal{G}$. (In other words, $\mathcal{G}$ contains all functions that can be obtained from regular polynomials with the help of the floor function and the usual arithmetic operations.)

The members of $\mathcal{G}$ are called *generalized polynomials*, and they appear quite naturally in diverse mathematical situations, from symbolic dynamics to Diophantine approximation to the theory of mathematical games. Unlike conventional polynomials, generalized polynomials needn't be eventually monotone (consider $[[n\alpha]n\beta] - [n^2\alpha\beta]$), may take only finitely many values (for example, $[(n+1)\alpha] - [n\alpha] - [\alpha]$ takes on only the values 0 and 1), and may vanish on sets of positive density while growing without bound on other such sets (multiply the previous example by n).

Comment. Alternatively, one could consider functions $\mathbf{Z} \to \mathbf{Z}$; we are restricting the domain of our generalized polynomials to $\mathbf{N}$, in part because it is convenient for us to deal with $\beta\mathbf{N}$ rather than with $\beta\mathbf{Z}$ (see below). However, any results we obtain for $\mathbf{N}$ can be extended to $\mathbf{Z}$ by considering the reflected functions $g(n) = f(-n)$ for $n < 0$. More generally, it is possible to define generalized polynomials to be functions $\mathbf{Z}^r \to \mathbf{Z}^t$, and obtain results for configurations in positive density subsets of $\mathbf{Z}^t$, with largeness conditions on the set of parameters in $\mathbf{Z}^r$. This is done at essentially no extra cost in [Mc2], where a non-classical IP structure theory must be used whether one restricts to the $\mathbf{Z}$ case or not. We are choosing not to do it here, as the classical Furstenberg structure theory for $\mathbf{Z}$ actions, which we do use, is simpler than the structure theory for general $\mathbf{Z}^t$ actions. (A similar choice for economy was made in [BM1], which like the present paper sought to present a new technique while avoiding minutiae where possible.)

1.8. Despite the oddities canvassed in the previous subsection, exciting new evidence has begun to emerge that generalized polynomials do possess certain strong regularities, as the following sample from [BL3] attests.

Theorem.

1. For any unitary operator U on a Hilbert space $\mathcal{H}$, any $f \in \mathcal{H}$ and any $g \in \mathcal{G}$, $\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} U^{g(n)} f$ exists in the norm topology.
2. There exists a unique invariant mean on the algebra of bounded generalized polynomials.
3. For any bounded generalized polynomial $g \in \mathcal{G}$, there exists a translation T on a nilmanifold X (that is, $X = N/\Gamma$, where N is nilpotent Lie group and Γ a cocompact lattice), and a Riemann integrable (actually, piecewise polynomial) function $f : X \to \mathbf{R}$ such that $g(n) = f(T^n x)$ for all $n \in \mathbf{Z}$.

Encouraged by 1. in the previous theorem, one may naturally hope for a version of Theorem 1.5 for generalized polynomials. However, one notices immediately that the situation with $\mathcal{G}$ is more complicated; indeed it is easy to show that a necessary condition on generalized polynomials $\{q_1, \dots, q_k\}$ satisfying the conclusion of the theorem is that no q_i nor $q_i - q_j$ be constant on a set of positive upper Banach density. (Non-trivial generalized polynomials that vanish on sets of positive density are abundant; e.g. $q(n) = [2(\pi n - [\pi n])]n$.) A serious attempt to obtain a satisfactory extension was undertaken in [BK], and for many special classes of generalized polynomials, was met with success. In the course of these investigations, no examples emerged showing that the obvious necessary condition is not also sufficient. This led to the following conjecture.

1.9. Conjecture. If p_i are generalized polynomials, $1 \leq i \leq k$, such that no p_i and no $p_i - p_j$, $1 \leq i \neq j \leq k$, is constant on a set of positive upper Banach density, then for any weakly mixing system $(X, \mathcal{A}, \mu, T)$ and any $A_i \in \mathcal{A}$, $0 \leq i \leq k$, one has

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} \left|\mu(A_0 \cap T^{p_1(n)} A_1 \cap \cdots \cap T^{p_k(n)} A_k) - \prod_{i=0}^{k} \mu(A_i)\right| = 0.$$

Moreover, for any $\epsilon > 0$, the set

$$\left\{n : \left|\mu(A_0 \cap T^{p_1(n)} A_1 \cap \cdots \cap T^{p_k(n)} A_k) - \prod_{i=0}^{k} \mu(A_i)\right| < \epsilon\right\}$$

has uniform upper Banach density one.

Unfortunately, we are unable to resolve the foregoing conjecture at this time. However, as we will show, a counterpart of the conjecture, in which uniform Cesàro convergence is replaced by convergence along certain types of ultrafilters, is true for weakly mixing systems. This result, in addition to enhancing our knowledge about weak mixing, also opens the door to new extensions of Szemerédi's theorem.

Ultrafilters: topological algebra in $\beta\mathbf{N}$.

1.10. In the next few subsections, we develop needed facts concerning $\beta\mathbf{N}$. For more information, see e.g. [HS] or [Be2].

A *filter* on $\mathbf{N}$ is a family $p \subset \mathcal{P}(\mathbf{N})$ satisfying the following conditions.

(i) $\emptyset \notin p$,

(ii) If $A \in p$ and $A \subset B$ then $B \in p$,

(iii) If $A \in p$ and $B \in p$ then $(A \cap B) \in p$.

If also

(iv) If $\mathbf{N} = A \cup B$ and $A \notin p$ then $B \in p$,

then p is an *ultrafilter*. We denote the family of ultrafilters on $\mathbf{N}$ by $\beta\mathbf{N}$. (By a routine application of Zorn's lemma, any filter is contained in an ultrafilter; however, one cannot construct non-principal (see below) ultrafilters without using some form of the axiom of choice.)

Comment. We find it intuitively useful to identify $p \in \beta\mathbf{N}$ with the $\{0,1\}$-valued, finitely additive probability measure μ_p on the power set of $\mathbf{N}$, defined by $\mu_p(A) = 1$ if and only if $A \in p$.

1.11. Given $A \subset \mathbf{N}$, let $\overline{A} = \{p \in \beta\mathbf{N} : A \in p\}$. The family $\{\overline{A} : A \subset \mathbf{N}\}$ forms a basis for a compact Hausdorff topology on $\beta\mathbf{N}$ having the following properties:

1. Identify each $n \in \mathbf{N}$ with the *principal* ultrafilter $p_n = \{A \subset \mathbf{N} : n \in A\}$. This gives an embedding for which one has $\overline{\mathbf{N}} = \beta\mathbf{N}$.

2. For $A \subset \mathbf{N}$ and $n \in \mathbf{N}$, put $A - n = \{m \in \mathbf{N} : m + n \in A\}$. Define a binary operation $+$ on $\beta\mathbf{N}$ by the rule $A \in p + q$ if and only if $\{n : (A - n) \in p\} \in q$. (When ultrafilters are viewed as measures, this is just convolution.) Then $+$, so defined, is associative (though not commutative), and is an extension of addition on $\mathbf{N}$.

3. For fixed $p \in \beta\mathbf{N}$, the map $q \to p + q$ is continuous. Thus we say that $(\beta\mathbf{N}, +)$ is a *compact left topological semigroup.*

1.12. By a theorem of Ellis ([El]), any compact Hausdorff left topological semigroup has an idempotent $p = p + p$. Note that $A \in p$ if and only if $\{n : (A - n) \in p\} \in p$. In other words, μ_p has a sort of shift-invariance property: $\mu_p(A) = 1$ if and only if for p-many n, $\mu_p(A - n) = 1$. Coupling this observation with iterated use of the Poincaré recurrence theorem (which holds for finitely additive measures), one obtains a simple proof of the following theorem of N. Hindman (the original proof was quite involved; see [H]).

1.13. Theorem. Let $r \in \mathbf{N}$ and let $\mathbf{N} = \bigcup_{i=1}^{r} C_i$. Then for some i, $1 \leq i \leq r$, C_i contains the set of finite sums of an infinite sequence (x_j), namely a set of the form

$$FS\big(\{x_j\}\big) = \{x_{j_1} + x_{j_2} + \cdots + x_{j_k} : j_1 < j_2 < \cdots < j_k, k \in \mathbf{N}\}.$$

What the ultrafilter proof of Hindman's theorem shows is that any cell C_i that belongs to any idempotent ultrafilter (an *IP-set*, for short), contains a finite-sums set. In this paper, however, we will be dealing with idempotent ultrafilters satisfying further properties; properties that guarantee an even richer combinatorial structure for their members.

The most well-studied such sub-class of idempotents are the *minimal idempotents*, that is, idempotent ultrafilters that are members of the smallest 2-sided ideal of $\beta\mathbf{N}$ ($I \subset \beta\mathbf{N}$ is a 2-sided ideal if $p+q$ belongs to I whenever either p or q does; the smallest being simply the intersection of all such), whose members are called *central sets.* (Central sets were introduced in [F2] under a different, dynamical definition. For equivalence to the characterization we have given, see [BH].) One indication of the combinatorial richness of central sets is that they are *piecewise syndetic*, that is, they are the intersection of a syndetic set with a set containing arbitrarily long intervals. Further evidence is given by Furstenberg's "central sets theorem". This theorem (see [F2, Proposition 8.21]) says, very roughly, that any central set contains an IP-set of arithmetic progressions of any finite length; among its many corollaries is the fact that so-called Rado systems (see [F2, Section 8.7]) are solvable in any central set. (For a combinatorial characterization of central sets, see [HMS]. Being a characterization, this account captures, in some sense, *all* of their combinatorial richness.)

1.14. In this paper, we will be dealing with a broader class of idempotent ultrafilters whose importance has more lately become apparent, namely those all of whose members have positive upper Banach density, called *essential* idempotents in [BD]. In [BBDF] it is shown that members of essential idempotents, called $\mathcal{D}$*-sets*, share much in the way of combinatorial richness with central sets (including satisfaction of the central sets theorem and solvability of Rado systems).

Notice that, since any piecewise syndetic set clearly has positive upper Banach density, every minimal idempotent is essential. On the other hand, one has the following (cf. [BM3], [BD]):

Theorem. There exist essential idempotents $p \in \beta\mathbf{N}$ that are not minimal.

Proof. First, observe that for any set E having positive upper Banach density, there are ultrafilters q containing E and having the property that every member of q has positive upper Banach density. To see this, just consider the family $\mathcal{I} = \{E \setminus F : d^*(F) = 0\}$. One easily checks that $\mathcal{I}$ is a filter containing E, and one may let q be any ultrafilter containing $\mathcal{I}$. Next, it is easy to show that the family D of ultrafilters all of whose members have positive upper Banach density forms a compact semigroup. We now proceed with the proof.

It is not difficult to construct a sequence (x_j) having the property that for every N, $S_N = FS\big(\{x_j\}_{j=N}^{\infty}\big)$ has positive upper density but is not piecewise syndetic. (Such a construction is carried out in [A].) Assuming this has been done, let $S = \bigcap_{N=1}^{\infty} \overline{S_N}$. S is a non-empty compact subsemigroup of $\beta\mathbf{N}$, and by the foregoing paragraph $\overline{S_N} \cap D \neq \emptyset$ for every N. This implies that $S \cap D$ is a non-empty compact semigroup and therefore contains an idempotent p. p is essential, being an idempotent member of D, but contains the non-piecewise syndetic sets S_N, and so is not minimal. $\square$

It is the essential idempotents that will allow us to formulate and prove a version of Theorem 1.5 for generalized polynomials. We use the following notion of limit along an ultrafilter:

1.15. Definition. Given a sequence (x_n) in a topological space and an ultrafilter $p \in \beta\mathbf{N}$, we write $p\text{-}\lim_n x_n = x$ if for any neighborhood U of x, $\{n : x_n \in U\} \in p$. Note that for sequences (x_n) in compact Hausdorff spaces, $p\text{-}\lim_n x_n$ always exists.

We are now in a position to formulate the aforementioned analogue of Theorem 1.5:

1.16. Theorem. $(X, \mathcal{A}, \mu, T)$ is a weakly mixing system if and only if whenever $k \in \mathbf{N}$, p is an essential idempotent and $v_1, \ldots, v_k \in \mathcal{G}$ such that neither v_i nor $v_i - v_j$ is constant on any member of p, $1 \leq i \neq j \leq k$. If $f_0, \ldots, f_k \in L^\infty(X)$,

$$p\text{-}\lim_n \int f_0 T^{v_1(n)} f_1 \cdots T^{v_k(n)} f_k \, d\mu = \prod_{i=0}^{k} \Big(\int f_i \, d\mu \Big).$$

1.17. The foregoing theorem suggests the question: for which $v \in \mathcal{G}$ are there no essential idempotents p such that v is constant on a member of p? A ready class of these are those $v \in \mathcal{G}$ having "non-zero leading coefficient" (the leading coefficient of a generalized polynomial $g(n)$ is the coefficient of the highest power of n in a formal expansion of g, treating greatest integer brackets as parentheses; for example, the leading coefficient of $[an][bn] + [cn^2]$ is $ab + c$). Somewhat more generally: it is easy to check (see Proposition 3.4 below) that every $v \in \mathcal{G}$ can be written as $v(n) = \sum_{i=0}^{k} B_i(n) n^i$, where each B_i is a bounded real-valued generalized polynomial. If B_k is bounded away from zero, then v will not be constant on any infinite set.

However, even for generalized polynomials v having zero leading coefficient, very strong algebraic conditions must typically be met in order for v to be constant on a member of an idempotent ultrafilter, as the following result from [MQ] attests.

Notation. We write $[\![x]\!] = [x + \frac{1}{2}]$ for the integer nearest to x. For a, b real, we denote by $a \otimes b$ the tensor product of a and b over the rationals. (So that $a \otimes b + c \otimes b = (a + c) \otimes b$, $a \otimes b + a \otimes c = a \otimes (b + c)$ and $r(a \otimes b) = ra \otimes b = a \otimes rb$ for $r \in \mathbf{Q}$.) $\mathbf{R} \otimes \mathbf{Q}$ denotes the span of tensors $a \otimes 1$ (over $\mathbf{Q}$).

1.18. Theorem. Let

$$F(n) = \sum_{i=1}^{n_1} [\![a_i n [\![b_i n]\!]]\!] + \sum_{i=1}^{n_2} [\![c_i n]\!][\![d_i n]\!] + \sum_{i=1}^{n_3} [\![e_i n]\!] + \sum_{i=1}^{n_4} [\![p_i n^2]\!].$$

There is an IP-set R on which F is constant if and only if the following three conditions are satisfied:

$$\sum_{i=1}^{n_1} a_i b_i + \sum_{i=1}^{n_2} c_i d_i + \sum_{i=1}^{n_4} p_i = 0; \tag{1}$$

$$\sum_{i=1}^{n_3} e_i = 0; \tag{2}$$

$$\sum_{i=1}^{n_1} a_i \otimes b_i + \sum_{i=1}^{n_2} (c_i \otimes d_i + d_i \otimes c_i) \in \mathbf{R} \otimes \mathbf{Q}. \tag{3}$$

Notice that (1) simply says that the leading coefficient of F is zero. So, if (1) holds but (2) and/or (3) fails, F is an example of a generalized polynomial having zero leading coefficient that is nevertheless not constant on any member of any idempotent p.

For degrees greater than 2, it appears to be very difficult to give necessary and sufficient conditions analogous to those of the foregoing theorem. On the other hand, it should in principle be possible to give (at least some) non-trivial necessary conditions, leading to broad classes of families of generalized polynomials for which the conclusion of Theorem 1.16 can be ensured. (Namely, those for which v_i and $v_i - v_j$ fail the conditions.) We make only one minor contribution to this project here (see Theorem 3.58 below), however we do think the question is an interesting one.

1.19. We also remark that, because of dependence on the ultrafilter p, Theorem 1.16 applies to many generalized polynomials that are nevertheless constant on sets of positive density (and hence do not satisfy the hypotheses of Conjecture 1.9). For example, let $q(n) = [2\{\sqrt{2}n\}]n$. The set $\{n : \{\sqrt{2}n\} > \frac{1}{2}\}$ is a central set (see, e.g. [HMS]), and in particular belongs to essential idempotents p. But for n belonging to this set, $q(n) = n$, so for such ultrafilters p, Theorem 1.16 may apply to families containing q. On the other hand, q takes on the constant value zero on the (again central) set $\{n : \{\sqrt{2}n\} < \frac{1}{2}\}$, which has density one-half.

Generalized polynomials, VIP systems and Szemerédi's theorem.

1.20. Recall that proofs of polynomial Szemerédi theorems (such as Theorem 1.6), generally have a mixing-of-all-orders component (such as Theorem 1.5). In like manner, here an appropriate version of Theorem 1.16 provides one piece of a proof of a general Szemerédi theorem for a class of functions that includes the generalized polynomials.

If q is a (conventional) polynomial of degree d, then for distinct integers $x_0, x_1, \ldots, x_d$ one has

$$\sum_{\emptyset \neq D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, q\Big(\sum_{i \in D} x_i\Big) = -a_0,$$

where a_0 is the constant term of q. More generally, if v is a generalized polynomial of degree d and p is an idempotent ultrafilter, then

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \neq D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, v\Big(\sum_{i \in D} x_i\Big) = -l_{v,p}, \tag{1.1}$$

where $l_{v,p}$ is some constant. (See Theorem 2.42 below.)

If $v : \mathbf{N} \to \mathbf{Z}$ is an arbitrary function satisfying (1.1) with $l_{v,p} = 0$, we say v is a p -VIP *system*. In this case, the *p-degree* $\deg_p v$ of v is the least d satisfying (1.1). Thus generalized polynomials are (up to a shift) p -VIP systems. However, there are many others. (By Proposition 3.48 below, for example, there exist p -VIP systems of superpolynomial growth.)

1.21. p -VIP systems are ultrafilter versions of *VIP systems* (see [BFM]), which are in turn polynomial variants of *IP-systems* (see, e.g., [FK2]). Briefly, an IP system is a function n taking finite subsets of n to a commutative group Γ in such a way that $n(\alpha \cup \beta) = n(\alpha)n(\beta)$ whenever $\alpha \cap \beta = \emptyset$. Furstenberg and Katznelson proved:

Theorem. Let $(X, \mathcal{A}, \mu)$ be a probability space and suppose T_i, $1 \leq i \leq k$, are IP systems into a commutative group Γ of measure preserving transformations of X. If $\mu(A) > 0$ then for some finite non-empty subset $\alpha \subset \mathbf{N}$, one has

$$\mu\big(A \cap T_1(\alpha)A \cap \cdots \cap T_k(\alpha)A\big) > 0.$$

1.22. It is a long-standing conjecture that Theorem 1.20 holds for VIP systems as well. (Some progress has been made in [BM2] and [Mc2].) In our current context of p -VIP systems in $\mathbf{Z}$, a version of this conjecture is as follows.

Conjecture. Let $(X, \mathcal{A}, \mu, T)$ be an invertible measure preserving system, let $p \in \beta\mathbf{N}$ be idempotent and let $v_1, \ldots, v_k$ be p -VIP systems. If $\mu(A) > 0$ and $B \in p$ then for some $n \in B$ one has

$$\mu\big(A \cap T^{v_1(n)}A \cap \cdots \cap T^{v_k(n)}A\big) > 0.$$

Although we are unable to settle this conjecture here, we do confirm it for essential ultrafilters p and a broad class of p -VIP systems containing, in particular, generalized polynomials v (more precisely, their shifts by the appropriate constants $l_{v,p}$). Restricting to the class of generalized polynomials, this leads to the following combinatorial application (see Corollary 3.41 below).

1.23. Theorem. Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$ and let $g_1, \ldots, g_k$ be generalized polynomials. Suppose p is an essential idempotent ultrafilter and $A \in p$. Then there exists $n \in A$ such that

$$d^*\big(E \cap (E + g_1(n) - l_{g_1,p}) \cap \cdots \cap (E + g_k(n) - l_{g_k,p})\big) > 0.$$

(In other words, the set of such n belongs to p.)

1.24. Discussion. To illustrate the dependence of $l_{v,p}$ on p, let $v(n) = [\pi n]$. The sets $E = \big\{n : \{\pi n\} \in [0, \frac{1}{2})\big\}$ and $E' = \big\{n : \{\pi n\} \in [\frac{1}{2}, 1)\big\}$ support essential idempotents p and p', respectively (they are central sets, again by the combinatorial characterization in [HMS]). An easy calculation now shows that $l_{v,p} = 0$ and $l_{v,p'} = -1$.

A natural class of generalized polynomials considered in [BKM] does not exhibit this dependence on p. As a first approximation, this class consists of all those generalized polynomials formed from polynomials of zero constant term via iterated use of the closest integer function $[\![\cdot]\!]$. More precisely (and more inclusively), the *admissible generalized polynomials* $\mathbf{N} \to \mathbf{Z}$ consist of the smallest subgroup $\mathcal{G}_a$ of the generalized polynomials that includes $n \to n$, is an ideal in the ring of all generalized polynomials, i.e. is such that if $p \in \mathcal{G}_a$ and q is a generalized polynomial then $pq \in \mathcal{G}_a$, and has the property that for all $m \in \mathbf{N}$, $c_1, \ldots, c_m \in \mathbf{R}$, $p_1, \ldots, p_m \in \mathcal{G}_a$, and $0 < k_1, \ldots, k_m < 1$, the mapping $n \to [\sum_{i=1}^m c_i p_i(n) + k_i]$ is in $\mathcal{G}_a$. For example, $p(n) = [\sqrt{3}[\sqrt{2}n^2]n^5 + \sqrt{17}n^3 + \frac{1}{2}][\sqrt{5}n]$ is admissible.

One can exploit the fact that $l_{v,p} = 0$ for v admissible to show that, for an arbitrary generalized polynomial g, $l_{g,p}$ can take only finitely many values as p ranges over the idempotents. Indeed, one can write down an expression for $l_{g,p}$ depending only on which cell of a given finite partition lies in p; here the partition is expressed in terms of inequalities involving fractional parts of generalized polynomials. For example, let $g(n) = 17[\sqrt{2}[\sqrt{3}n^2] + \sqrt{5}] + [\sqrt{7}n]$. Putting $C_1 = \{n : 0 \leq \{\sqrt{3}n^2\} < \frac{1}{2}\}$ and $C_2 = \{n : 0 \leq \{\sqrt{7}n\} < \frac{1}{2}\}$, one has

$$\begin{aligned} g(n) =& 17\big[\sqrt{2}[\sqrt{3}n^2 + \frac{1}{2}] + e_1(n) + \sqrt{5}\big] + [\sqrt{5}n + \frac{1}{2}] + e_2(n) \\ =& 17\Big(\big[\sqrt{2}v_1(n) + (\sqrt{5} - 2)\big] + 2 + e_1(n)\Big) + v_2(n) + e_2(n) \\ =& 17v_3(n) + 17\big(2 + e_1(n)\big) + v_2(n) + e_2(n), \end{aligned}$$

where $e_i(n) = 0$, $n \in C_i$, $e_i(n) = -1$ otherwise, and each v_i is admissible. Therefore:

$$l_{g,p} = \begin{cases} 16 \text{ if } (C_1 \cap C_2) \in p \\ 17 \text{ if } (C_1 \cap C_2^c) \in p \\ 33 \text{ if } (C_1^c \cap C_2) \in p \\ 34 \text{ if } (C_1^c \cap C_2^c) \in p \end{cases}$$

If one restricts to admissible generalized polynomials, Theorem 1.23 admits of a simpler formulation.

Notation. We denote the family of essential idempotents by $\mathcal{D}$ (Definition 2.21 a. below). We call a subset E of $\mathbf{N}$ a $\mathcal{D}$-set if it belongs to some $p \in \mathcal{D}$; E is a $\mathcal{D}^*$ set if it belongs to *every* $p \in \mathcal{D}$ (equivalently, if $E \cap F \neq \emptyset$ for every $\mathcal{D}$-set F). Notice that, since all minimal idempotents are essential, all central sets are $\mathcal{D}$ sets and all $\mathcal{D}^*$-sets are *central** (intersect every central set non-trivially). On the other hand, there are central* sets that are not $\mathcal{D}^*$ (for example, the complements of the sets S_N appearing in the proof of Theorem 1.14).

1.25. It is easily verified that the family of $\mathcal{D}^*$ sets possesses the filter property. The following theorem, therefore (modulo Furstenberg correspondence), makes good on the promise of subsection 1.6.

Theorem. Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$ and let $g_1, \ldots, g_k \in \mathcal{G}_a$. Then

$$\Big\{n \in \mathbf{N} : d^*\Big(E \cap \big(E + g_1(n)\big) \cap \cdots \cap \big(E + g_k(n)\big)\Big) > 0\Big\}$$

is $\mathcal{D}^*$ in $\mathbf{N}$.

Comments. (a) It follows immediately (by the consideration of reflected functions), that the set of $n \in \mathbf{Z}$ satisfying $d^*\big(E \cap \big(E + g_1(n)\big) \cap \cdots \cap \big(E + g_k(n)\big)\big) > 0$ is $\mathcal{D}^*$ in $\mathbf{Z}$ (see [BD]). Indeed, by keeping track of constants and just a bit better than we have (again, we don't do it in the interest of economy), one can find $a > 0$ such that $\big\{n \in \mathbf{Z} : d^*\big(E \cap \big(E+g_1(n)\big) \cap \cdots \cap \big(E+g_k(n)\big)\big) > a\big\}$ is $\mathcal{D}^*$ (this is implicit in [Mc2]–see Definition 4.2 there). Here a depends on E, k and $\max \deg g_i$; we do not see how to replace dependence on E by dependence on $d^*(E)$.

(b) If one is only interested in existence of "admissible generalized polynomial progressions", this can be achieved by finding "initial polynomial configurations" inside the ranges of admissible generalized polynomials. Indeed in [BHå], such a trick is used to show that the ranges of various generalized polynomials are good for single recurrence. For example, to see that $\{[\alpha n^2] : n \in \mathbf{N}\}$ is a set of recurrence, where α is an irrational number, it suffices by Sárközy's theorem ([Sá]) and the uniformity of recurrence (see e.g. [BHMP]) to establish that it contains a configuration of the form $\{i^2 m : 1 \leq i \leq L\}$ for every L. Fixing L, choose n with $\{\alpha n^2\}$ less than $\frac{1}{L^2}$. Then $\{[\alpha(in)^2] : 1 \leq i \leq L\}$ will have the required form with $m = [\alpha n^2]$. Though we will not supply details, a natural embellishment of this technique is sufficient to establish existence of the configurations guaranteed by Theorem 1.25. (That one only gets in this way that the set of good n has positive density provides part of the impetus for our more involved approach.)

1.26. As mentioned above, our methods actually allow one to obtain multiple recurrence for a class of p -VIP systems larger than the generalized polynomials. Indeed, what makes the methods work for generalized polynomials is, very roughly, the fact that once an essential idempotent p is fixed, any generalized polynomial g that is p-linear in the sense that $p\text{-}\lim_{x,y} \big(g(x+y) - g(x) - g(y)\big) = 0$ must also be p-linear in the stronger sense that $p\text{-}\lim_x \big(g(x) - [\alpha x + \frac{1}{2}]\big) = 0$, for some real α. (See Theorem 3.8 (a) below.) The reason this is significant is that our methods require $d^*\big(g(A)\big) > 0$ for every $A \in p$ when g is linear (in the first sense).

Generally speaking, then, once an essential idempotent p is fixed, our multiple recurrence conclusions will hold for any class of p -VIP systems that is closed under the difference and differentiation operations that occur in the various stages of the proof and whose linear members g have the property that $d^*\big(g(A)\big) > 0$ for every $A \in p$. Although our efforts surely don't begin to characterize the class of systems having this property, at the end of the paper (see subsections 3.43 to 3.67) we do discuss the construction of two natural examples, so-called *well spaced* and *densely packed* systems, not coming from generalized polynomials. These constructions are quite complicated, so we shall not attempt to summarize them here.

On a simpler note, a class to which our methods can easily be seen to apply is given by those p -VIP systems that are equivalent to generalized polynomials in their highest degree terms, i.e., systems of the form $g - l_{g,p} + v$, where g is a generalized polynomial and v is an *arbitrary* p -VIP system satisfying $\deg_p v < \deg_p(g - l_{g,p})$. This yields the following application (see Corollary 3.58 below), which can already be compared somewhat more favorably to Conjecture 1.21.

1.27. Theorem. Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$ and let $g_1, \ldots, g_k$ be generalized polynomials.

Suppose p is an essential idempotent ultrafilter. Let $v_1, \ldots, v_k$ be any p-VIP systems with $\deg_p v_i < \deg_p(g_i - l_{g_i,p})$ and $\deg_p(v_i - v_j) < \deg_p\big((g_i - l_{g_i,p}) - (g_j - l_{g_j,p})\big)$ $1 \le i \ne j \le k$. If $A \in p$ then there exists $n \in A$ such that

$$d^*\big(E \cap (E + g_1(n) - l_{g_1,p} + v_1(n)) \cap \cdots \cap (E + g_k(n) - l_{g_k,p} + v_k(n))\big) > 0.$$

1.28. The structure of the paper is as follows. In Section 2, we develop the general theory of p -VIP systems, introducing in particular the special subclasses mentioned in the previous paragraphs and proving, for those that are linear, a Hilbert space projection theorem (Theorem 2.25). This theorem is then used to prove Theorem 2.48, the multiple mixing result for weakly mixing systems upon which both Theorem 1.16 and the weakly mixing component of the proof of our multiple recurrence theorem are based. In Section 3, we give these two applications. First, the multiple mixing theorem for generalized polynomials (Theorem 3.9), then the multiple recurrence theorem (Theorem 3.10), the proof of which uses a "polynomialized" version of Theorem 2.25 (Theorem 3.11, which may also be seen as a version of the main result from [BFM]). This proof is carried out in subsections 3.11 through 3.40, its principal ingredients being a relativized version of the multiple mixing result (Theorem 3.20) and a VIP multiple recurrence theorem for distal measure preserving systems (Theorem 3.29). Finally, at the end of Section 3, we give the aforementioned further examples of p -VIP systems to which our results apply.

1.29. *Acknowledgement:* The authors express their gratitude to N. Frantzikinakis and to I. J. Håland-Knutson for helpful discussion and commentary.

2. p-VIP systems and multiple weak mixing.

In this section we develop the general theory of p-VIP systems, concluding with a general "weak mixing of all orders" result (Theorem 2.48 below) for essential idempotents p.

2.1. Notation. Let $H : \mathbf{N}^r \to X$ be a function, where X is a locally compact topological space. For an ultrafilter $p \in \beta\mathbf{N}$, we shall write $p\text{-}\lim_{x_1,\ldots,x_r} H(x_1, \ldots, x_r)$ for

$$p\text{-}\lim_{x_1} p\text{-}\lim_{x_2} p\text{-}\lim_{x_3} \cdots p\text{-}\lim_{x_r} H(x_1, \ldots, x_r).$$

2.2. Convention. In equations involving p - lim expressions on both sides, we interpret the "=" sign as meaning that if either side exists, so does the other and they are equal.

2.3. Lemma. Let $p \in \beta\mathbf{N}$ be idempotent, let X be a locally compact topological space and let $f : \mathbf{N} \to X$. Then

$$p\text{-}\lim_{x,y} f(x + y) = p\text{-}\lim_{x} f(x).$$

Proof. Suppose the right hand side exists and is equal to the value a. Let V be a neighborhood of a. Choose a pre-compact open set U containing a with $\overline{U} \subset V$. Now $A = \{x : f(x) \in \overline{U}\} \in p$, which implies by idempotence of p that $\{x : (A - x) \in p\} \in p$, which is to say $\{x : \{y : f(x + y) \in \overline{U}\} \in p\} \in p$. But this says precisely that $\{x :$

$p\text{-}\lim_y f(x+y) \in \overline{U}\} \in p$. Hence $\{x : p\text{-}\lim_y f(x+y) \in V\} \in p$. Since V was an arbitrary neighborhood of a, $p\text{-}\lim_{x,y} f(x+y) = a$.

Suppose now that the left hand side exists and is equal to a. Let V be a neighborhood of a and let $B = \{x : f(x) \in V\}$. One has $\{x : p\text{-}\lim_y f(x+y) \in V\} \in p$. Notice that $p\text{-}\lim_y f(x+y) \in V$ implies $B - x = \{y : f(x+y) \in V\} \in p$. Hence $\{x : (B-x) \in p\} \in p$, which implies $B \in p$. $\square$

2.4. Corollary. Let $p \in \beta\mathbf{N}$ be idempotent and let X be a locally compact topological space. Let $d < r \in \mathbf{N}$ and let $H : \mathbf{N}^{d+1} \to X$ be a function. Let $S_0, S_1, \ldots, S_d$ be non-empty subsets of $\mathbf{N}$ with $\bigcup_{i=0}^{d} S_i = \{1, 2, \ldots, r\}$ and such that if $0 \le i < j \le d$, $x \in S_i$ and $y \in S_j$, then $x < y$. Then

$$p\text{-}\lim_{y_0,\ldots,y_d} H(y_0, \ldots, y_d) = p\text{-}\lim_{x_1,\ldots,x_r} H\Big(\sum_{i \in S_0} x_i, \sum_{i \in S_1} x_i, \ldots, \sum_{i \in S_d} x_i\Big).$$

Proof. This is a routine induction using the previous lemma. Details are left to the reader. $\square$

2.5. Definition. Let $p \in \beta\mathbf{N}$ be an idempotent and let $(G, +)$ be a discrete commutative group with identity e. A function $v : \mathbf{N} \to G$ is a *p-VIP system* if for some $d \in \mathbf{N}$ one has

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, v\Big(\sum_{i \in D} x_i\Big) = e. \tag{2.1}$$

If v is a p-VIP system then the least d for which (2.1) holds is the *p-degree* of v, denoted $\deg_p v$.

2.6. Remarks. (a) It is easy to show that if $q(x) \in \mathbf{R}[x]$ with $q(0) = 0$ then q is a p-VIP system having p-degree equal to $\deg q$. Also, if v satisfies identity (2.1) for some d, then it satisfies the corresponding identity for any $d' > d$.

(b) The above definition should be construed as applying to any commutative group (even ones for which the usual topology is not discrete); however, the limit appearing in (2.1) must be construed as existing in the discrete topology. (For example, if $G = \mathbf{R}$, the sum appearing in the limit must actually be equal to zero for appropriately filtered $(x_0, x_1, \ldots, x_d)$; it is not sufficient for this sum to tend to zero in the usual topology.)

2.7. Definition. Let $(G, +)$ be a commutative group. For $f : \mathbf{N} \to G$ and $h \in \mathbf{N}$, we define the *derivative of f with step h* to be the function $D_h f : \mathbf{N} \to G$ given by $D_h f(n) = f(n+h) - f(n) - f(h)$.

2.8. Theorem. Let $p \in \beta\mathbf{N}$ be idempotent, let $(G, +)$ be a commutative group and let $v : \mathbf{N} \to G$ be a p-VIP system of degree d. Then for p-many h, $D_h v$ is a p-VIP system, and $\{h : \deg_p D_h v = d - 1\} \in p$.

Proof. For notational convenience we extend v to $\mathbf{N} \cup \{0\}$ by setting $v(0) = e$. One has the identity

$$\begin{aligned}
&\sum_{\emptyset\neq D\subset\{0,1,2,\dots,t\}} (-1)^{|D|}\, v\Big(\sum_{i\in D} x_i\Big)\\
&= -\,v(x_0) + \sum_{\emptyset\neq D\subset\{1,2,\dots,t\}} (-1)^{|D|} \left(v\Big(\sum_{i\in D} x_i\Big) - v\Big(x_0+\sum_{i\in D} x_i\Big)\right)\\
&= \sum_{D\subset\{1,2,\dots,t\}} (-1)^{|D|} \left(v\Big(\sum_{i\in D} x_i\Big) - v\Big(x_0+\sum_{i\in D} x_i\Big) + v(x_0)\right)\\
&= -\sum_{\emptyset\neq D\subset\{1,2,\dots,t\}} (-1)^{|D|}\, D_{x_0}v\Big(\sum_{i\in D} x_i\Big),
\end{aligned}$$

from which it follows immediately that v is p-VIP of degree at most t if and only if for p-many x_0, $D_{x_0}v$ is p-VIP of degree at most $t-1$. $\square$

In the following lemma, 1_A is the characteristic function of A.

2.9. Lemma. Let $p\in\beta\mathbf{N}$ be an idempotent and suppose that $A\in p$. Then

$$p\text{-}\lim_{x_0,\dots,x_d} 1_A(x_0+x_1+x_2+\cdots+x_d)=1. \tag{2.2}$$

Proof. This is a special case of Corollary 2.4. $\square$

We will use the next lemma often, usually without explicit mention. We omit the proof.

2.10. Lemma. Let $p\in\beta\mathbf{N}$ be idempotent, let X be a topological space, let $f:\mathbf{N}^r\to X$ and suppose that $p\text{-}\lim_{x_1,\dots,x_r} f(x_1,\dots,x_r)=a$. Then for any $1\le i_1<i_2<\cdots<i_r\le t$, one has $p\text{-}\lim_{x_1,\dots,x_t} f(x_{i_1},\dots,x_{i_r})=a$.

2.11. Proposition. Let $p\in\beta\mathbf{N}$ be idempotent and let $(G,+)$ be a Hausdorff topological commutative group with identity e. If $v:\mathbf{N}\to G$ is a p-VIP system for which $p\text{-}\lim_n v(n)=g$, then $g=e$.

Proof. Suppose not. Let U be a closed neighborhood of g not containing e. Choose a neighborhood V of g such that $2^dV-(2^d-1)V\subset U$. Let $A=\{n: v(n)\in V\}$. Then $A\in p$ so that (2.2) holds, contradicting (2.1). $\square$

2.12. Remark. It follows that when G is compact, $p\text{-}\lim_n v(n)=e$ for every p-VIP system $v:\mathbf{N}\to G$.

2.13. Proposition. Let $p\in\beta\mathbf{N}$ be idempotent, let $\pi:G\to H$ be a homomorphism of commutative groups, and let $v:\mathbf{N}\to G$ be a p-VIP system. Then $\pi\circ v:\mathbf{N}\to H$ is a p-VIP system into H.

2.14. Corollary. Let $p\in\beta\mathbf{N}$ be idempotent and suppose that $v:\mathbf{N}\to\mathbf{R}$ is p-VIP. Then $\{v\}$ is a p-VIP system into $\mathbf{T}$.

2.15. Definition. Let G be a commutative group and suppose $\phi, \psi : \mathbf{N} \to G$ are functions. For $p \in \beta\mathbf{N}$, we write $\phi \approx_p \psi$ if $\{n : \phi(n) = \psi(n)\} \in p$.

2.16. Proposition. $\approx_p$ is an equivalence relation.

2.17. Definition. Given an idempotent ultrafilter $p \in \beta\mathbf{N}$, we denote by $\mathcal{V}_p$ the set of equivalence classes of p-VIP systems $\mathbf{N} \to \mathbf{Z}$ under $\approx_p$.

2.18. Remark. Clearly, $\mathcal{V}_p$ is a group under addition. Often we may speak of a p-VIP system $f : \mathbf{N} \to \mathbf{Z}$ as being a member of $\mathcal{V}_p$. Obviously we mean by this the equivalence class of f under $\approx_p$.

2.19. Proposition. Let $f \in \mathcal{V}_p$ and let $p \in \beta\mathbf{N}$ be idempotent. The map $\pi_f : \mathbf{N} \to \mathcal{V}_p$ defined by $\pi_f(h) = D_h f$ is a p-VIP system into $\mathcal{V}_p$. Moreover $\deg_p \pi_f = \deg_p f - 1$.

Proof. Let $d = \deg_p f$. To show that $\deg_p \pi_f \leq d - 1$, we must show that

$$p\text{-}\lim_{x_0,\ldots,x_{d-1}} \sum_{\emptyset \neq E \subset \{0,1,\ldots,d-1\}} (-1)^{|E|} D_{\Sigma_{i\in E} x_i} f \approx_p 0. \tag{2.3}$$

(The 0 on the right here is of course a function, not a number.) To establish (2.3), one must show that

$$p\text{-}\lim_{x_0,\ldots,x_{d-1}} p\text{-}\lim_{x_d} \sum_{\emptyset \neq E \subset \{0,1,\ldots,d-1\}} (-1)^{|E|} D_{\Sigma_{i\in E} x_i} f(x_d) = 0,$$

which is equivalent to

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \neq E \subset \{0,1,\ldots,d-1\}} (-1)^{|E|} \big(f\big(x_d + \sum_{i\in E} x_i\big) - f\big(\sum_{i\in E} x_i\big) - f(x_d)\big) = 0.$$

However, this is just

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \neq E \subset \{0,1,\ldots,d\}} (-1)^{|E|+1} f\big(\sum_{i\in E} x_i\big) = 0.$$

We leave the very similar verification $\deg_p \pi_d \geq d - 1$ to the reader. $\square$

2.20. Definition. Let $A \subset \mathbf{N}$. The *upper Banach density* of A is the number

$$d^*(A) = \limsup_{N-M\to\infty} \frac{|A \cap \{M, M+1, \ldots, N-1\}|}{N-M}.$$

2.21. Definitions.

a. We denote by $\mathcal{D}$ the set of all idempotent ultrafilters $p \in \beta\mathbf{N}$ having the property that for every $A \in p$, one has $d^*(A) > 0$.

b. Let $p \in \beta\mathbf{N}$ be idempotent. We denote by $\mathcal{B}_{p,1}$ the set of equivalence classes, under $\approx_p$, of the family of p -VIP systems $\varphi : \mathbf{N} \to \mathbf{Z}$ satisfying $\deg_p \varphi = 1$ and such that $d^*\big(\varphi(A)\big) > 0$ for all $A \in p$.

2.22. Remark. Typically, we will be interested in the case $p \in \mathcal{D}$, where any p-VIP system φ with $\deg_p \varphi = 1$ and $d^*\big(\varphi(A)\big) > 0$ whenever $d^*(A) > 0$ is a member of $\mathcal{B}_{p,1}$.

2.23. Definition. Let $\mathcal{H}$ be a separable Hilbert space and let T be a unitary operator on $\mathcal{H}$. We let

$$\mathcal{K}_T = \{f \in \mathcal{H} : \{T^n f : n \in \mathbf{Z}\} \text{ is precompact in the norm topology}\}.$$

2.24. Theorem. Let $(X, \mathcal{A}, \mu, T)$ be an invertible measure preserving system.
1. The map $f \to Tf$ is unitary on $L^2(X)$.
2. $\mathcal{K}_T$ is the closed linear subspace of $L^2(X)$ generated by the eigenfunctions of T (that is, by those f for which there is $\lambda \in \mathbf{C}$ such that $Tf(x) = f(Tx) = \lambda f(x)$ a.e.).
3. T is weakly mixing if and only if $\mathcal{K}_T$ is spanned by the constant functions.

2.25. Theorem. Let $\mathcal{H}$ be a separable Hilbert space and let T be a unitary operator on $\mathcal{H}$. Suppose $p \in \beta\mathbf{N}$ is idempotent, let $\varphi \in \mathcal{B}_{p,1}$ and for $f \in \mathcal{H}$ write $p\text{-}\lim_n T^{\varphi(n)} f = Pf$, where the limit is taken in the weak topology. Then P is the orthogonal projection onto $\mathcal{K}_T$.

Proof. The limit in question exists because, restricted to closed bounded subsets of $\mathcal{H}$, the weak topology is compact and metrizable. It is well known that any continuous linear self-map P of a Hilbert space with $||P|| \leq 1$ and $P^2 = P$ is an orthogonal projection. We show now $P^2 = P$. Let $f \in \mathcal{H}$ with $||f|| \leq 1$, let $\epsilon > 0$ and let ρ be a metric for the weak topology on the unit ball of $\mathcal{H}$.

Let $A = \big\{x : \{y : \varphi(x+y) = \varphi(x) + \varphi(y)\} \in p\big\} \in p$. Let $A_1 = \{n : \rho(Pf, T^{\varphi(n)} f) < \epsilon\} \in p$ and $A_2 = \{n : \rho(P^2 f, T^{\varphi(n)} Pf) < \epsilon\} \in p$. Fix $x \in A \cap A_2 \cap \big\{n : (A_1 - n) \in p\big\}$. Let

$$A_x = (A_1 - x) \cap \big\{y : \rho\big(T^{\varphi(y)} T^{\varphi(x)} f, PT^{\varphi(x)} f\big) < \epsilon\big\} \cap \{y : \varphi(x+y) = \varphi(x) + \varphi(y)\} \in p.$$

Now choose $y \in A_x$. One has

$$\rho(P^2 f, Pf) \leq \rho(P^2 f, T^{\varphi(x)} Pf) + \rho(PT^{\varphi(x)} f, T^{\varphi(y)} T^{\varphi(x)} f) + \rho(T^{\varphi(x+y)} f, Pf) \leq 3\epsilon,$$

where we have used the facts that P commutes with T (an easy exercise), $x \in A_2$, $x+y \in A_1$ and $y \in A_x$. Since ϵ and f were arbitrary, this shows that $P^2 = P$ and hence that P is an orthogonal projection.

Since range P is closed (another easy exercise), in order to show that $\mathcal{K}_T \subset$ range P it suffices to show that linear combinations of eigenfunctions are in range P. Let c_i be constants and suppose we are given eigenfunctions f_i for T, $1 \leq r$. Let α_i be the eigenvalue corresponding to f_i, so that $Tf_i = \alpha_i f_i$, and let $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_r) \in \mathbf{T}^r$. $\mathbf{T}^r$ is compact so $p\text{-}\lim_n \alpha^{\varphi(n)}$ exists. By Proposition 2.11, the limit must be $(0, 0, \ldots, 0)$. From this it easily follows that $p\text{-}\lim_n ||T^{\varphi(n)}\big(\sum_{i=1}^r c_i f_i\big) - \sum_{i=1}^r c_i f_i|| = 0$.

Finally we show that range $P \subset \mathcal{K}_T$. Let $f \in$ range P and put $B = \{n : ||T^{\varphi(n)}f - f|| < \epsilon\} \in p$. Since $\varphi \in \mathcal{B}_{p,1}$, one has $d^*\big(\varphi(B)\big) > 0$. In particular, $\big(\varphi(B) - \varphi(B)\big)$ is syndetic. Choose $R \in \mathbf{N}$ such that every $n \in \mathbf{Z}$ can be written as $n = \varphi(b_1) - \varphi(b_2) + r$, where $b_1, b_2 \in B$ and $0 \leq r < R$. We claim that $\{f, Tf, T^2 f, \ldots, T^{R-1}f\}$ is a 2ϵ-net for $\{T^n f : n \in \mathbf{Z}\}$. To see this, let $n \in \mathbf{Z}$ and write $n = \varphi(b_1) - \varphi(b_2) + r$, where $b_1, b_2 \in B$ and $0 \leq r < R$. Then $||T^n f - T^r f|| = ||T^{\varphi(b_1)}f - T^{\varphi(b_2)}f|| \leq ||T^{\varphi(b_1)}f - f|| + ||f - T^{\varphi(b_2)}f|| < 2\epsilon$. □

From now on, we will assume that functions in $L^2(X)$, $L^\infty(X)$, etc. are real valued.

2.26. Corollary. Let $(X, \mathcal{A}, \mu, T)$ be a weakly mixing measure preserving system, let $p \in \beta\mathbf{N}$ be idempotent and let $\varphi \in \mathcal{B}_{p,1}$. Then for any $f, g \in L^2(X)$, one has

$$p\text{-}\lim_n \int f T^{\varphi(n)} g \, d\mu = \Big(\int f \, d\mu\Big)\Big(\int g \, d\mu\Big).$$

Proof. Since $(X, \mathcal{A}, \mu, T)$ is weakly mixing, $\mathcal{K}_T$ consists of the constant functions. Hence

$$\begin{aligned} p\text{-}\lim_n \int f T^{\varphi(n)} g &= \int f\big(p\text{-}\lim_n T^{\varphi(n)} g\big) \, d\mu \\ &= \int f\big(Pg\big) \, d\mu = \int f\Big(\int g \, d\mu\Big) \, d\mu = \Big(\int f \, d\mu\Big)\Big(\int g \, d\mu\Big). \end{aligned}$$

2.27. Definition. We denote by $\mathcal{S}$ the smallest family of functions $\mathbf{N} \to \mathbf{Z}$, containing the integer-valued constants and the inclusion map $n \to n$, that is closed under products, and having the property that for every $f \in \mathcal{S}$ and $c \in \mathbf{R}$, $[cf] \in \mathcal{S}$. The members of $\mathcal{S}$ will be called *simple generalized polynomials.*

2.28. Definition. We define a function Deg on $\mathcal{S}$ inductively as follows. Deg $0 = -\infty$, Deg $k = 0$, where 0 and $0 \neq k \in \mathbf{Z}$ are constant functions, Deg $(n \to n) = 1$, Deg $(f_1 f_2) =$ Deg $f_1 +$ Deg f_2, and Deg $[cf] =$ Deg f for $c \neq 0$.

2.29. Definition. We denote by $\mathcal{G}$ the smallest family of functions $\mathbf{N} \to \mathbf{Z}$, containing the integer-valued constants and the inclusion map $n \to n$, that is closed under sums and products, and having the property that for every $f_1, \ldots, f_r \in \mathcal{G}$ and $c_1, \ldots, c_r \in \mathbf{R}$, $[\sum_{i=1}^r c_i f_i] \in \mathcal{G}$. The members of $\mathcal{G}$ will be called *generalized polynomials.*

2.30. Theorem. Let $p \in \beta\mathbf{N}$ be an idempotent ultrafilter and let $f \in \mathcal{G}$. Then there exist simple generalized polynomials $g_1, g_2, \ldots, g_k$ such that $f \approx_p \sum_{i=1}^k g_i$.

Proof. The proof, by induction on the complexity of f, is immediate. Suppose $f_1, f_2, \ldots, f_r$ satisfy the conclusion. Write $f_j \approx_p \sum_{i=1}^{k_j} g_{ji}$. Then $f_1 f_2 \approx_p \sum_{i=1}^{k_1} \sum_{t=1}^{k_2} g_{1i} g_{2t}$, $f_1 + f_2 \approx_p \sum_{i=1}^{k_1} g_{1i} + \sum_{i=1}^{k_2} g_{2i}$, and for real numbers $c_1, c_2, \cdots, c_r$, $[\sum_{i=1}^r c_i f_i] \approx_p K + \sum_{i=1}^r [c_i f_i]$ for a constant $0 \leq K < r$. □

2.31. Definition. For $p \in \beta\mathbf{N}$ idempotent and $f \in \mathcal{G}$, we let $\text{Deg}_p\, f$ be the minimum d for which there exist $g_1, \ldots, g_k \in \mathcal{S}$ with $d = \max_i \text{Deg}\, g_i$ and $f \approx_p \sum_{i=1}^k g_i$.

2.32. Remark. Let $p \in \beta\mathbf{N}$ be idempotent. Below, we will show that in fact, up to a shift, any $g \in \mathcal{G}$ is a p-VIP system. For such a g, there are therefore two kinds of p-degree. We believe it can happen that $\mathrm{Deg}_p\ g \neq \deg_p\ g$; see however Theorem 3.8 a. below for an important case in which they coincide.

2.33. Proposition. Let $p \in \beta\mathbf{N}$ be idempotent. If $f \approx_p g$ then $\{h \in \mathbf{N} : D_h f \approx_p D_h g\} \in p$.

Proof. Let $A = \{n : f(n) = g(n)\} \in p$. For $h \in (A \cap \{t : (A - t) \in p\}) \in p$, one has, for every $n \in \big((A-h) \cap A\big) \in p$, $f(h) = g(h)$, $f(n) = g(n)$ and $f(n+h) = g(n+h)$, whence $D_h f(n) = D_h g(n)$. □

2.34. Lemma. Let $p \in \beta\mathbf{N}$ be idempotent and let $f_1, \ldots, f_r \in \mathcal{G}$. Then $\mathrm{Deg}_p\left(\sum_{i=1}^r f_i\right) \leq \max_{1\leq i\leq r} \mathrm{Deg}_p\ f_i$.

2.35. Theorem. Let $p \in \beta\mathbf{N}$ be idempotent and let $f \in \mathcal{G}$. Then $\{h : \mathrm{Deg}_p\ D_h f < \mathrm{Deg}_p\ f\} \in p$.

Proof. By Theorem 2.30 and Proposition 2.33, it suffices to show this for simple f, which we do by induction on f's complexity. Suppose the result holds for f_1 and f_2. Choose a set $A \in p$ such that for all $h \in A$, one has $\mathrm{Deg}_p\ D_h f_1 < \mathrm{Deg}_p\ f_1$ and $\mathrm{Deg}_p\ D_h f_2 < \mathrm{Deg}_p\ f_2$. Now for $h \in A$,

$$\begin{aligned}
D_h(f_1 f_2)(n) =& f_1(n+h) f_2(n+h) - f_1(n) f_2(n) - f_1(h) f_2(h) \\
=& \big(D_h f_1(n) + f_1(n) + f_1(h)\big)\big(D_h f_2(n) + f_2(n) + f_2(h)\big) \\
=& D_h f_1(n) D_h f_2(n) + D_h f_1(n) f_2(n) + D_h f_1(n) f_2(h) + f_1(n) D_h f_2(n) \\
&+ f_1(n) f_2(h) + f_1(h) D_h f_2(n) + f_1(h) f_2(h).
\end{aligned}$$

Thus $D_h(f_1 f_2)$ is a sum of terms, each of which has p-Degree less than $\mathrm{Deg}_p\ f_1 f_2 = \mathrm{Deg}_p\ f_1 + \mathrm{Deg}_p\ f_2$. Now apply Lemma 2.34. Next, for $c \in \mathbf{R}$ and $h \in A$,

$$\begin{aligned}
D_h[c f_1](n) =& [c f_1(n+h)] - [c f_1(n)] - [c f_1(h)] \\
=& [c D_h f_1(n) + c f_1(n) + c f_1(h)] - [c f_1(n)] - [c f_1(h)] = [c D_h f_1(n)] + B,
\end{aligned}$$

where $0 \leq B < 3$ may be considered constant on an appropriate member of p. Hence $\mathrm{Deg}_p\ D_h[c f_1] = \mathrm{Deg}_p\ [c D_h f_1] = \mathrm{Deg}_p\ D_h f_1 < \mathrm{Deg}_p\ f_1 = \mathrm{Deg}_p\ [c f_1]$. □

2.36. Definition. Denote by $\mathcal{F}$ the family of non-empty, finite subsets of $\mathbf{N}$. For $d \in \mathbf{N}$, let $\mathcal{F}_d = \{\alpha \in \mathcal{F} : |\alpha| \leq d\}$.

2.37. Theorem. Let $p \in \beta\mathbf{N}$ be idempotent, let G be an additive, discrete commutative group with identity e, let $v : \mathbf{N} \to G$ and $d \in \mathbf{N}$. Then v is a p-VIP system of degree at most d if and only if there exists a function $u : \mathcal{F}_d \to G$ such that for every $r \in \mathbf{N}$, one has

$$p\text{-}\lim_{x_1,\ldots,x_r} \Big(v\big(\sum_{i=1}^r x_i\big) - \sum_{\{x_1,\ldots,x_r\} \supset \gamma \in \mathcal{F}_d} u(\gamma) \Big) = e. \tag{2.4}$$

Proof. Suppose that (2.4) holds. We must show

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset\neq D\subset\{0,1,2,\ldots,d\}} (-1)^{|D|}\, v\Big(\sum_{i\in D} x_i\Big) = e. \tag{2.5}$$

Applying (2.4) on the inside, what we must show is that

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset\neq D\subset\{0,1,2,\ldots,d\}} (-1)^{|D|} \sum_{\{x_i:i\in D\}\supset\gamma\in\mathcal{F}_d} u(\gamma) = e,$$

which in turn may be rewritten as

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset\neq D\subset\{x_0,x_1,x_2,\ldots,x_d\}} (-1)^{|D|} \sum_{D\supset\gamma\in\mathcal{F}_d} u(\gamma) = e. \tag{2.6}$$

We claim that the expression inside the limit is identically equal to e. To see this, we simply fix some $\gamma\in\mathcal{F}_d$ with $\gamma\subset\{x_0,\ldots,x_d\}$ and count occurrences of $u(\gamma)$ in the expression. Letting $|\gamma|=k$, for $1\le t\le d+1$ the number of sets $D\subset\{x_0,\ldots,x_d\}$ with $|D|=t$ and $\gamma\subset D$ is 0 if $t<k$ and $\binom{d+1-k}{t-k}$ otherwise. Therefore the net number of times $u(\gamma)$ is counted is $\sum_{t=k}^{d+1}(-1)^t\binom{d+1-k}{t-k}$. Substituting $i=t-k$, this is $\sum_{i=0}^{d-k+1}(-1)^{i+k}\binom{d+1-k}{i}$, which is $(1-1)^{d+1-k}=0$ by the binomial theorem. Thus each $u(\gamma)$ is counted zero times (net) and we are done.

Now suppose that v is p-VIP of degree at most d; that is, suppose that (2.5) holds. We must exhibit a function $u:\mathcal{F}_d\to G$ such that (2.4) holds. For $\gamma\in\mathcal{F}_d$, we set

$$u(\gamma)=\sum_{\emptyset\neq\beta\subset\gamma}(-1)^{|\gamma|-|\beta|}v\big(\sum_{i\in\beta} i\big).$$

First we show that (2.4) holds for $r\le d$, which amounts to showing

$$p\text{-}\lim_{x_1,\ldots,x_r}\Big(v\big(\sum_{i=1}^r x_i\big)-\sum_{\{x_1,\ldots,x_r\}\supset\gamma\in\mathcal{F}_d}\Big(\sum_{\emptyset\neq\beta\subset\gamma}(-1)^{|\gamma|-|\beta|}v\big(\sum_{i\in\beta} i\big)\Big)\Big)=0.$$

We count occurrences of $v\big(\sum_{i\in\beta} i\big)$ in the double sum for $\beta\subset\{x_1,\ldots,x_r\}$. For $\beta=\{x_1,\ldots,x_r\}$ there is a single occurrence which, being subtracted, cancels the occurrence of $v\big(\sum_{i=1}^r x_i\big)$. If $|\beta|=k<r$, then for any t with $k\le t\le r$, there are $\binom{r-k}{t-k}$ sets $\gamma\subset\{x_1,\ldots,x_r\}$ containing β with $|\gamma|=t$. It follows that the net number of occurrences of $v\big(\sum_{i\in\beta} i\big)$ in the double sum is $\sum_{t=k}^{r}(-1)^{t-k}\binom{r-k}{t-k}$. Substituting $i=t-k$, this is $\sum_{i=0}^{r-k}(-1)^i\binom{r-k}{i}=(1-1)^{r-k}=0$.

For $r>d$ we use induction. Suppose (2.4) holds when r is replaced by anything less than r. Let $S_0,S_1,\ldots,S_d$ be the (non-empty) cells of a non-interlacing partition of $\{1,2,\ldots,r\}$: by this we mean that for all $x\in S_i$ and $y\in S_j$, $x<y$ if $i<j$. For $x_1,\ldots,x_r\in\mathbf{N}$, we adopt the notation $y_i=\sum_{j\in S_i}x_j$. (In particular, notice that y_0+

$y_1 + \cdots + y_d = x_1 + x_2 + \cdots + x_r$.) Our key observation is that for any function H of the variables $y_0, y_1, \ldots, y_d$, one has by Corollary 2.4 that

$$p\text{-}\lim_{y_0,\ldots,y_d} H(y_0,\ldots,y_d) = p\text{-}\lim_{x_1,\ldots,x_r} H(y_0,\ldots,y_d) = p\text{-}\lim_{x_1,\ldots,x_r} H\Big(\sum_{j\in S_0} x_j, \ldots, \sum_{j\in S_d} x_j\Big).$$

Now, since v is a p-VIP system of degree d, we have

$$\begin{aligned} &p\text{-}\lim_{x_1,\ldots,x_r} \sum_{\emptyset\neq D\subset\{0,1,2,\ldots,d\}} (-1)^{|D|}\, v\Big(\sum_{i\in D} y_i\Big) \\ =&p\text{-}\lim_{y_0,\ldots,y_d} \sum_{\emptyset\neq D\subset\{0,1,2,\ldots,d\}} (-1)^{|D|}\, v\Big(\sum_{i\in D} y_i\Big) = e. \end{aligned} \tag{2.7}$$

On the other hand, we claim that an analog of (2.6) above holds:

$$p\text{-}\lim_{x_1,\ldots,x_r} \sum_{\emptyset\neq D\subset\{0,1,2,\ldots,d\}} (-1)^{|D|} \sum_{\{x_j : j\in\bigcup_{i\in D} S_i\}\supset\gamma\in\mathcal{F}_d} u(\gamma) = e. \tag{2.8}$$

This can be established in the same way as (2.6), by showing the expression inside the limit to be identically zero. Some details: fix $\gamma \subset \{x_j : j \in \bigcup_{i\in D} S_i\} = \{x_1, \ldots, x_r\}$, $\gamma \in \mathcal{F}_d$. We count occurrences of $u(\gamma)$. Let k be the number of partition cells S_i for which $\{x_j : j \in S_i\} \cap \gamma) \neq \emptyset$ and proceed from here exactly as in the argument concerning (2.4) to show that the net number of times $u(\gamma)$ is counted is zero.

Next, for any non-empty D *properly* contained in $\{0, 1, \ldots, d\}$ our induction hypothesis yields

$$p\text{-}\lim_{x_1,\ldots,x_r} \Big(v\big(\sum_{i\in D} y_i\big) - \sum_{\{x_j : j\in\bigcup_{i\in D} S_i\}\supset\gamma\in\mathcal{F}_d} u(\gamma)\Big) = e. \tag{2.9}$$

Now (2.7), (2.8) and (2.9) combine to give

$$p\text{-}\lim_{x_1,\ldots,x_r} \Big(v\big(\sum_{i=0}^{d} y_i\big) - \sum_{\{x_j : j\in\bigcup_{i\in D} S_i\}\supset\gamma\in\mathcal{F}_d} u(\gamma)\Big) = e.$$

In other words,

$$p\text{-}\lim_{x_1,\ldots,x_r} \Big(v\big(\sum_{i=1}^{r} x_i\big) - \sum_{\{x_1,\ldots,x_r\}\supset\gamma\in\mathcal{F}_d} u(\gamma)\Big) = e,$$

as required. $\square$

A word of warning: although the previous theorem says that any p-VIP system v has a "generating" function u whose domain is $\mathcal{F}_d$, and also says that any function v that is generated in the right way by a function u whose domain is $\mathcal{F}_d$ is p-VIP, it does *not* say that an *arbitrary* function u whose domain is $\mathcal{F}_d$ will generate some v in the right way. In fact, we shall now see that this is in general not the case.

2.38. Theorem. Let $\mathbf{P} \in \beta\mathbf{N}$ be an idempotent and let $u : \mathcal{F}_1 \to \mathbf{Z}$ be given by $u(\{x\}) = 2^x$. There is no function $v : \mathbf{N} \to \mathbf{Z}$ for which

$$p\text{-}\lim_{x,y} \Big(v(x+y) - \big(u(\{x\}) + u(\{y\}) \big) \Big) = 0.$$

Proof. Suppose there were such a v. For $x \in \mathbf{N}$, put $A_x = \{y : v(x+y) = 2^x + 2^y\}$. Then $A = \{x : A_x \in p\} \in p$. Choose $x \in A$ such that $(A - x) \in p$ and let $B = A_x \cap (A - x) \in p$. Pick $b \in B$ such that $(B - b) \in p$. Then $b \in (A - x)$ and $(A_x - b) \in p$. Put $x' = b + x \in A$ (note $x' \neq x$). Note $\big(A_{x'} \cap (A_x - p)\big) \in p$. Choose some $y' \neq x$ in this set. Let $y = y' + b \in A_x$. Then $x + y = x' + y'$, $y \in A_x$, $y' \in A_{x'}$ and x is equal to neither x' nor y'. It follows that $2^{x'} + 2^{y'} = v(x' + y') = v(x + y) = 2^x + 2^y$, which is a contradiction. □

2.39. Lemma. Let $(x_i)_{i=1}^\infty \subset \mathbf{Z}$, $(y_i)_{i=1}^\infty \subset \mathbf{R}$, $0 < k < 1$ and suppose that $\sum_{i=1}^\infty |x_i| \langle y \rangle < \min\{k, 1-k\}$. Then for every $n \in \mathbf{N}$, $[\sum_{i=1}^n x_i y_i + k] = \sum_{i=1}^n x_i [y_i + \frac{1}{2}]$.

Proof.

$$\Big| \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i [y_i + \frac{1}{2}] \Big| \le \sum_{i=1}^n |x_i| \cdot |y_i - [y_i + \frac{1}{2}]| = \sum_{i=1}^n |x_i| \langle y_i \rangle \le \min\{k, 1-k\} < 1.$$

□

2.40. Corollary. Let $p \in \beta\mathbf{N}$ be idempotent, and let $v, f : \mathbf{N} \to \mathbf{Z}$ be p-VIP systems of degrees d and c respectively. Then fv is a p-VIP system of degree at most $d + c$.

Proof. By Theorem 2.37 there exist functions $u : \mathcal{F}_d \to \mathbf{Z}$ and $g : \mathcal{F}_c \to \mathbf{Z}$ such that for every $r \in \mathbf{N}$,

$$\begin{aligned} &p\text{-}\lim_{x_1,\ldots,x_r} \Big(v(\sum_{i=1}^r x_i) - \sum_{\{x_1,\ldots,x_r\} \supset \gamma \in \mathcal{F}_d} u(\gamma) \Big) \\ =&p\text{-}\lim_{x_1,\ldots,x_r} \Big(f(\sum_{i=1}^r x_i) - \sum_{\{x_1,\ldots,x_r\} \supset \gamma \in \mathcal{F}_d} g(\gamma) \Big) = 0. \end{aligned}$$

For $\beta \in \mathcal{F}_{d+c}$, set

$$h(\beta) = \sum_{\gamma \in \mathcal{F}_d, \alpha \in \mathcal{F}_c, \gamma \cup \alpha = \beta} u(\gamma) g(\alpha).$$

One now has

$$\begin{aligned} 0 =&p\text{-}\lim_{x_1,\ldots,x_r} \Big(v(\sum_{i=1}^r x_i) f(\sum_{i=1}^r x_i) - (\sum_{\{x_1,\ldots,x_r\} \supset \gamma \in \mathcal{F}_d} u(\gamma))(\sum_{\{x_1,\ldots,x_r\} \supset \alpha \in \mathcal{F}_c} g(\alpha)) \Big) \\ =&p\text{-}\lim_{x_1,\ldots,x_r} \Big(v(\sum_{i=1}^r x_i) f(\sum_{i=1}^r x_i) - \sum_{\{x_1,\ldots,x_r\} \supset \beta \in \mathcal{F}_{d+c}} (\sum_{\gamma \in \mathcal{F}_d, \alpha \in \mathcal{F}_c, \gamma \cup \alpha = \beta} u(\gamma) g(\alpha)) \Big) \\ =&p\text{-}\lim_{x_1,\ldots,x_r} \Big(v(\sum_{i=1}^r x_i) f(\sum_{i=1}^r x_i) - (\sum_{\{x_1,\ldots,x_r\} \supset \beta \in \mathcal{F}_{d+c}} h(\beta)) \Big). \end{aligned}$$

Now apply Theorem 2.37 again, this time in reverse. $\square$

2.41. Proposition. Let $p \in \beta\mathbf{N}$ be an idempotent, let $L \in \mathbf{Z}$ and suppose that $v_i : \mathbf{N} \to \mathbf{Z}$ are p-VIP systems, $1 \le i \le m$. Suppose also that $c_1, \ldots, c_m \in \mathbf{R}$, and let $0 < k < 1$. Then:

(a) $v_1 - v_2$ is a p-VIP system of p-degree at most $\max\{\deg_p v_1, \deg_p v_2\}$.
(b) $v_1(v_2 - n)$ is a p-VIP system of p-degree at most $\deg_p v_1 + \deg_p v_2\}$.
(c) $[k + \sum_{i=1}^m c_i v_i]$ is a p-VIP system of p-degree at most $\max\{\deg_p v_i : 1 \le i \le m\}$.

Proof. (a) is an easy consequence of Theorem 2.37, as is (b) with the help of (a) and Corollary 2.40. As for (c), write $f(n) = [k + \sum_{i=1}^m c_i v_i(n)]$ and note that for each i, $n \to \{c_i v_i(n)\}$ is a p-VIP system into $\mathbf{T}$, from which it follows by Remark 2.12 that $p\text{-}\lim_n \langle c_i v_i(n)\rangle = 0$. Let $d = \max\ \deg_p(v_i)$ and put $A = \{n : \langle c_i v_i(n)\rangle < \frac{\min\{k, 1-k\}}{2^d m}\} \in p$. Note that by Lemmas 2.9 and 2.10, for all non-empty $D \subset \{0, 1, \ldots, d\}$ and all i, $1 \le i \le m$, one has

$$p\text{-}\lim_{x_0,\ldots,x_d} 1_A\Big(\sum_{i\in D} x_i\Big) = 0.$$

From this it follows that

$$\begin{aligned}
&p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, f\Big(\sum_{i\in D} x_i\Big) \\
=&p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, \Big[k + \sum_{i=1}^m c_i v_i\Big(\sum_{i\in D} x_i\Big)\Big] \\
=&p\text{-}\lim_{x_0,\ldots,x_d} \Big[k + \sum_{i=1}^m c_i \sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|} v_i\Big(\sum_{i\in D} x_i\Big)\Big] = 0.
\end{aligned}$$

$\square$

2.42. Theorem. Let $p \in \beta\mathbf{N}$ be idempotent. For every generalized polynomial $f : \mathbf{N} \to \mathbf{Z}$ there is an $l_f = l_{f,p} \in \mathbf{Z}$ such that one has, for $d = \mathrm{Deg}_p f$,

$$p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, f\Big(\sum_{i\in D} x_i\Big) = -l_f.$$

In other words, $f - l_f$ is a p-VIP system of p-degree at most $\mathrm{Deg}_p f$.

Proof. Clearly it suffices to establish the result for simple generalized polynomials. Let $\mathcal{H}$ be the set of generalized polynomials in $\mathcal{S}$ for which the conclusion holds. $\mathcal{H}$ clearly contains the constants, the inclusion $n \to n$ and is closed under products (by Proposition 2.41 (b)). Finally, if $f \in \mathcal{H}$ then $[cf] \in \mathcal{H}$ follows by the fact that

$$\sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, \Big[cf\Big(\sum_{i\in D} x_i\Big)\Big] = \Big[c \sum_{\emptyset \ne D \subset \{0,1,2,\ldots,d\}} (-1)^{|D|}\, f\Big(\sum_{i\in D} x_i\Big)\Big] - J$$

for some $0 \leq J < 2^d$. (For appropriately filtered $x_0, \ldots, x_d$ J assumes a constant value, say J_0, so that the right-hand side is the constant $[-cl_f] - J_0$.) $\square$

From the previous theorem we get that $\deg_p(f - l_f) \leq \mathrm{Deg}_p f$. We give a special name to those f for which equality holds.

2.43. Definition. Let $p \in \beta\mathbf{N}$ be an idempotent and let f be a generalized polynomial. If $\deg_p(f - l_f) = \mathrm{Deg}_p f$ then we will say that f is *p-regular.*

2.44. Proposition. Let $p \in \beta\mathbf{N}$ be idempotent and suppose that f is a p-regular generalized polynomial. Then $\{h : D_h f \text{ is } p\text{-regular}\} \in p$.

Proof. Note that $D_h(f - l_f) = D_h f + l_f$. According to Theorem 2.42, for p-many h, $D_h(f - l_f)$ is a p-VIP system. Therefore, by the previous theorem,

$$\{h : \deg_p D_h(f - l_f) \leq \mathrm{Deg}_p D_h(f - l_f)\} \in p.$$

By Theorem 2.8,

$$\{h : \deg_p D_h(f - l_f) = \big(\mathrm{Deg}_p(f - l_f)\big) - 1\} = \{h : \deg_p D_h(f - l_f) = \big(\deg_p(f - l_f)\big) - 1\} \in p.$$

On the other hand, by Theorem 2.35,

$$\{h : \mathrm{Deg}_p\big(D_h(f - l_f)\big) < \mathrm{Deg}_p(f - l_f)\} \in p.$$

For h in the intersection of these three sets, clearly $\deg_p\big(D_h(f - l_f)\big) = \mathrm{Deg}_p\big(D_h(f - l_f)\big)$, which implies that $\deg_p(D_h f - l_{D_h f}) = \mathrm{Deg}_p D_h f$. $\square$

2.45. Theorem. (*Van der Corput lemma.*) Assume that $(u_g)_{g \in \mathbf{N}}$ is a bounded sequence in a Hilbert space. Let $p \in \beta\mathbf{N}$ be an idempotent. If $p\text{-}\lim_h p\text{-}\lim_g \langle u_{g+h}, u_g \rangle = 0$ then $p\text{-}\lim_g u_g = 0$ in the weak topology.

Proof. A bit of notation. Let $\mathcal{F}$ denote the family of finite, non-empty subsets of $\mathbf{N}$. For sets $\alpha, \beta \subset \mathcal{F} \cup \{\emptyset\}$, we write $\beta < \alpha$ if for every $b \in \beta$ and every $a \in \alpha$, one has $b < a$. If $\alpha = \{i_1, i_2, \cdots, i_k\}$, where $i_1 < i_2 < \cdots < i_k$, we will write g_α as shorthand for the sum $g_{i_k} + g_{i_{k-1}} + \cdots + g_{i_2} + g_{i_1}$. Finally we write $FS(g_1, \ldots, g_j) = \big\{g_\alpha : \emptyset \neq \alpha \subset \{1, 2, \ldots, j\}\big\}$.

Without loss of generality we will assume that $||u_g|| \leq 1$, $g \in \mathbf{N}$. Suppose to the contrary that $p\text{-}\lim_g u_g = \tilde{u} \neq 0$. Let $\delta = \frac{||\tilde{u}||^2}{2}$ and pick $k \in \mathbf{N}$ and $\epsilon > 0$ such that $\frac{1}{k} + \epsilon < \delta$. Inductively choose $g_1, \ldots, g_k \in \mathbf{N}$ such that for all j, $1 \leq j \leq k$, one has

(i) for every $\alpha, \beta \subset \{1, \ldots, j\}$ with $\alpha \neq \emptyset$, $\beta \neq \emptyset$ and $\beta < \alpha$, $|\langle u_{g_\alpha + g_\beta}, u_{g_\alpha} \rangle| < \epsilon$.

(ii) for every $\alpha, \beta \subset \{1, \ldots, j\}$ with $\beta \neq \emptyset$ and $\beta < \alpha$, $p\text{-}\lim_g |\langle u_{g + g_\alpha + g_\beta}, u_{g + g_\alpha} \rangle| < \epsilon$.

(iii) for all $r \in FS(g_1, \ldots, g_j)$, $\langle u_r, \tilde{u} \rangle > \delta$.

(iv) for all $r \in \{0\} \cup FS(g_1, \ldots, g_j)$, $\{g : \langle u_g, \tilde{u} \rangle > \delta\} - r \in p$.

(v) for all $r \in FS(g_1, \ldots, g_j)$, $\{h : p\text{-}\lim_g |\langle u_{g+h}, u_g \rangle| < \epsilon\} - r \in p$.

Having done this, we let $v_i = u_{g_k + g_{k-1} + \cdots + g_i}$, $1 \leq i \leq k$, and observe that $|\langle v_i, v_j \rangle| < \epsilon$ and $\langle v_i, \tilde{u} \rangle > \delta$, $1 \leq i \neq j \leq k$. From the former it follows that $\big\langle \sum_{i=1}^k v_i, \sum_{i=1}^k v_i \big\rangle < k +$

$k^2\epsilon < k^2\delta$, which implies that $|\big\langle \sum_{i=1}^k v_i, \tilde{u}\big\rangle| < k\delta$, contradicting the latter and completing the proof.

Suppose then that $0 \leq j < k$ and $g_1, \ldots, g_j$ have been chosen. By the induction hypothesis, for an $\epsilon' < \epsilon$ one has

$$B = \Big(\bigcap_{r \in \{0\} \cup FS(g_1,\ldots,g_j)} \big(\{g : \langle u_g, \tilde{u}\rangle > \delta\} - r\big)\Big)$$
$$\cap \Big(\bigcap_{\alpha,\beta \subset \{1,\ldots,j\}, \emptyset \neq \beta < \alpha} \{g : |\langle u_{g+g_\alpha+g_\beta}, u_{g+g_\alpha}\rangle| < \epsilon'\}\Big)$$
$$\cap \Big(\bigcap_{r \in \{0\} \cup FS(g_1,\ldots,g_j)} \big(\{h : p\text{-}\lim_g |\langle u_{g+h}, u_g\rangle| < \epsilon\} - r\big)\Big)$$

is a member of p. Therefore, we may choose $g_{j+1} \in B$ such that $B - g_{j+1} \in p$. It is now a routine (if tedious) matter to check that (i)-(v) hold for j replaced by $j+1$. $\square$

2.46. Theorem. Let $p \in \mathcal{D}$ and let $f \in \mathcal{G}$ with $\mathrm{Deg}_p f = 1$. Then $(f - l_f) \in \mathcal{B}_{p,1}$.

Proof. We have $f \approx_p \sum_i f_i$, where $f_i \in \mathcal{S}$ with $\mathrm{Deg} f_i = 1$. Each non-constant f_i has the form $f_i(n) = [c_k[c_{k-1}[c_{k-2}[\cdots[c_2[c_1 n]]\cdots]]]$. It is an easy exercise from here to show that $f(n) \approx_p [cn] + const$, $c \neq 0$. This clearly implies that if $d^*(A) > 0$, then $d^*\big((f - l_f)(A)\big) > 0$. $\square$

We remind the reader of the definition of $\mathcal{B}_{p,1}$ (see Definition 2.21).

2.47. Definitions. Let $p \in \beta\mathbf{N}$ be idempotent.

a. Define families of (equivalence classes of, under $\approx_p$) p-VIP systems $\mathbf{N} \to \mathbf{Z}$ $\mathcal{B}_{p,i}$, $i > 1$, inductively as follows.

$$\mathcal{B}_{p,i} = \big\{v \in \mathcal{V}_p : 1 \leq \deg_p v \leq i \text{ and } \{h : D_h v \in \mathcal{B}_{p,i-1}\} \in p\big\}.$$

Now write $\mathcal{B}_p = \bigcup_{i=1}^\infty \mathcal{B}_{p,i}$.

b. A family $\mathcal{E} \subset \mathcal{B}_p$ is said to be *p-acceptable* if it forms a group under addition and is essentially closed under taking of deriviatives; that is, if for every $v \in \mathcal{E}$ with $\deg_p v > 1$, $\{h : D_h v \in \mathcal{E}\} \in p$.

The next theorem uses an inductive scheme, called "PET-induction", introduced in [Be1]. In the present case, it is implemented as follows. Let $p \in \mathcal{D}$. Given two p-VIP systems v and w, we write $v \sim_p w$ if $\deg_p v = \deg_p w > \deg_p(v - w)$. One may check that $\sim_p$ is an equivalence relation. Given a finite set $A = \{v_1, \ldots, v_k\}$ of p-VIP systems, define the *weight* of A by $\mathbf{w}(A) = (w_1, w_2, \ldots)$, where w_i is the number of equivalence classes under $\sim_p$ of degree i p-VIP systems represented in A. Finally for distinct weights $\mathbf{w} = (w_1, w_2, \ldots)$ and $\mathbf{u} = (u_1, u_2, \ldots)$, one writes $\mathbf{w} > \mathbf{u}$ if $w_d > u_d$, where d is the largest j satisfying $w_j \neq u_j$. This is a well-ordering of the set of weights, and PET-induction is simply induction on this ordering.

2.48. Theorem. Suppose $(X, \mathcal{A}, \mu, T)$ is a weakly mixing system, $k \in \mathbf{N}$, $p \in \beta\mathbf{N}$ is idempotent and $\mathcal{E}$ is a p-acceptable family. Let $v_1, \ldots, v_k$ be distinct members of $\mathcal{E}$. If $f_0, \ldots, f_k \in L^\infty(X)$ then

$$p\text{ - }\lim_n \int f_0 T^{v_1(n)} f_1 \cdots T^{v_k(n)} f_k \, d\mu = \prod_{i=0}^{k} \Big(\int f_i \, d\mu \Big).$$

Proof. The proof is by induction on the weight vector $\mathbf{w}(A)$ of $A = \{v_1, \ldots, v_k\}$. By Corollary 2.26, the conclusion holds when the weight vector is $(1, 0, 0, \ldots)$. Suppose for induction that the result holds for families having weight vector $\mathbf{w} < \mathbf{w}(A)$.

We consider the case $\int f_a \, d\mu = 0$ for some a, $0 \le a \le k$, reducing the general case to this special one via the identity

$$\prod_{i=0}^{k} a_i - \prod_{i=0}^{k} b_i = (a_0 - b_0) \prod_{i=1}^{k} b_i + a_0(a_1 - b_1) \prod_{i=2}^{k} b_i + \cdots + \Big(\prod_{i=0}^{k-1} a_i \Big)(a_t - b_t) \tag{2.10}$$

under the integral, with $a_i = T^{v_i(n)} f_i$ and $b_i = \int f_i \, d\mu$, $0 \le i \le k$ (we take $v_0 \equiv 0$). Indeed, by composing through by $T^{-v_i(\alpha)}$, if necessary, where v_i is of minimal degree (this does not change the weight vector), we may in fact assume that $1 \le a \le k$. Also without loss of generality we may assume that $\|f_i\|_\infty \le 1$, $0 \le i \le k$.

We shall complete the proof by showing that $p\text{ - }\lim_n \prod_{i=1}^{k} T^{v_i(n)} f_i = 0$ weakly. We use Theorem 2.45 in the following way. For $n \in \mathbf{N}$, set $x_n = \prod_{i=1}^{k} T^{v_i(n)} f_i$. Then

$$\begin{aligned} & p\text{ - }\lim_m p\text{ - }\lim_n \langle x_{n+m}, x_n \rangle \\ =& p\text{ - }\lim_m p\text{ - }\lim_n \int \prod_{i=1}^{k} T^{v_i(n+m)} f_i \prod_{i=1}^{k} T^{v_i(n)} f_i \, d\mu \\ =& p\text{ - }\lim_m p\text{ - }\lim_n \int \prod_{i=1}^{k} T^{v_i(n)} f_i \prod_{i=1}^{k} T^{v_i(n) + D_m v_i(n)} (T^{v_i(m)} f_i) \, d\mu. \end{aligned} \tag{2.11}$$

By acceptability, for p-many m, $D_m v_i \in \mathcal{E}$. Also for p-many m, $\deg_p D_m v_i < \deg_p v_i$, so that in particular $v_i + D_m v_i \sim v_i$ for p-many m. By Propositions 2.11 and 2.19, for p-many m, $D_m v_i \ne v_j - v_i$ for all $i \ne j$. Similarly, for p-many m, $D_m(v_i - v_j) \ne (v_j - v_i)$ for $i \ne j$. Also, either for p-many m $D_m v_i$ is the identity (as will happen when v_i is of degree one), or for p-many m, $D_m v_i$ is not the identity. Let w be the number of indices for which the former occurs. Permuting indices so that $\deg_p v_i$ is non-decreasing with i, we may assume that for p-many m, $D_m v_i = I$ if $1 \le i \le w$ and $D_m v_i \ne I$ if $w < i \le k$.

We now write $A_m = \{v_1, \cdots, v_k, v_{w+1} + D_m v_{w+1}, \cdots, v_k + D_m v_k\}$. By the facts obtained in the previous paragraph, for p-many m, A_m consists of distinct elements of $\mathcal{E}$. Moreover, $\mathbf{w}(A_m) = \mathbf{w}(A)$.

We now rewrite the last line of (2.11) as

$$p\text{ -}\lim_m p\text{ -}\lim_n \int \prod_{i=1}^{w} T^{v_i(n)}(f_i T^{v_i(m)} f_i) \prod_{i=w+1}^{k} T^{v_i(n)} f_i T^{v_i(n)+D_m v_i(n)}(T^{v_i(m)} f_i)\ d\mu. \quad (2.12)$$

For p-many m, the set

$$B_m = \{v_2 - v_1, v_3 - v_1, \ldots, v_k - v_1, v_{w+1} + D_m v_{w+1} - v_1, \cdots, v_k + D_m v_k - v_1\}$$

is contained in $\mathcal{E}$ and precedes A_m. The reason for this is that v_1 is of minimal weight, so that subtracting throughout by v_1 will decrease the p-degree of every system u such that $u \sim_p v_1$, while failing to change the p-degrees of the others. Moreover, if neither $v_i \sim_p v_1$ nor $v_j \sim_p v_1$, then $v_i \sim_p v_j$ if and only if $(v_i - v_1) \sim_p (v_j - v_1)$. These considerations imply that B_m has one less equivalence class under $\sim_p$ of p-degree $\deg_p v_1$ and the same number of equivalence classes at any p-degree greater than $\deg_p v_1$.

Now by the induction hypothesis and the fact that T is measure preserving, we can rewrite (2.12) as

$$\begin{aligned} &p\text{ -}\lim_m p\text{ -}\lim_n \int \prod_{i=1}^{w} T^{v_i(n)-v_1(n)}(f_i T^{v_i(m)} f_i) \\ &\qquad\qquad \prod_{i=w+1}^{k} T^{v_i(n)-v_1(n)} f_i T^{v_i(n)+D_m v_i(n)-v_1(n)}(T^{v_i(m)} f_i)\ d\mu \\ =&p\text{ -}\lim_m \prod_{i=1}^{w} \Big(\int f_i T^{v_i(m)} f_i\ d\mu\Big) \prod_{i=w+1}^{k} \Big(\int f_i\ d\mu\Big)^2 = \prod_{i=1}^{k} \Big(\int f_i\ d\mu\Big)^2 = 0, \end{aligned}$$

as required. □

3. Measure-theoretic and combinatorial applications.

In this section we offer applications of Theorem 2.48. First is a sufficient condition (which is also clearly necessary) on a family of generalized polynomials for multiple p-mixing for weakly mixing systems, where p is idempotent and all the members of p have positive upper Banach density (Theorem 3.9). Next, we give a new extension of the polynomial Szemerédi theorem (restricted for ease of presentation to $\mathbf{Z}$) for p-acceptable families of p-VIP systems.

3.1. Definition. The *set of real valued generalized polynomials* is the smallest family $\mathcal{RG}$ of functions $\mathbf{N} \to \mathbf{Z}$ that contains the polynomial ring $\mathbf{R}[x]$ and is closed under sums, products and composition with the floor function.

We omit the proof of the following easy proposition.

3.2. Proposition. $\mathcal{RG} = \{\sum_{i=1}^{k} c_i g_i : c_i \in \mathbf{R}, g_i \in \mathcal{G}, 1 \le i \le k\}$.

3.3. Corollary. Let $p \in \beta\mathbf{N}$ be idempotent and suppose $g \in \mathcal{RG}$. Then for some constant $\alpha = \alpha_g \in \mathbf{R}$, $g - \alpha$ is a p -VIP system.

3.4. Proposition. Let $g \in \mathcal{RG}$. One has a representation

$$g(n) = \sum_{i=0}^{k} B_i(n) n^i, \tag{3.1}$$

where $B_i \in \mathcal{RG}$ is bounded, $1 \leq i \leq k$.

Proof. Anything in $\mathbf{R}[x]$ obviously admits of a representation of the form (3.1) (with the B_is constant). The sum or product of any two things having such representation is clearly again such. Finally if $[g]$ has such a representation then $[g] = g - \{g\}$ does as well, as $-\{g\}$ may be assimilated to B_0. $\square$

3.5. Lemma. Let $p \in \beta\mathbf{N}$ be idempotent, let G be an commutative group with identity e, and let $v : \mathbf{N} \to G$ be a p-VIP system. If $\big\{m : \{n : v(m+n) = v(n)\} \in p\big\} \in p$ then $v \approx_p e$.

Proof. Let $d = \deg_p v$. Then one has (in the discrete topology)

$$\begin{aligned} e =& p\text{-}\lim_{x_0,\ldots,x_d} \sum_{\emptyset \neq D \subset \{0,\ldots,d\}} (-1)^{|D|} v\big(\sum_{i \in D} x_i\big) \\ =& p\text{-}\lim_{x_0,\ldots,x_d} \Big(- v(x_0) + \sum_{\emptyset \neq D \subset \{1,\ldots,d\}} (-1)^{|D|+1} \Big(v\big(x_0 + \sum_{i \in D} x_i\big) - v\big(\sum_{i \in D} x_i\big)\Big)\Big) \\ =& p\text{-}\lim_{x_0} v(x_0). \end{aligned}$$

$\square$

3.6. Lemma. Let $g \in \mathcal{RG}$ and suppose $p \in \mathcal{D}$. If $|p\text{-}\lim_n ng(n)| = L < \infty$ then $g \approx_p 0$.

Proof. By [BL3, Theorem 3.1], there exists a set $\mathcal{L} \subset \mathbf{R}$ with non-empty interior that is defined by a system of polynomial inequalities, a polynomial map $\mathcal{P} : \mathcal{L} \to \mathbf{R}$, and a set $E \in p$ such that $g|_E$ is well-distributed on $\mathcal{S} = \mathcal{P}(\mathcal{L})$ with respect to the measure $\mu = \mathcal{P}_*(\lambda)$ on $\mathcal{S}$, defined by $\mu(A) = \frac{\lambda(\mathcal{P}^{-1}(A) \cap \mathcal{L})}{\lambda(\mathcal{L})}$ for Borel $A \subset \mathbf{R}$. If $\mathcal{S} = \{0\}$, we are done. Otherwise $\mu(\{0\}) = 0$; we shall obtain a contradiction. Let $C = E \cap \{n : |ng(n)| \leq 2L\}$. Then $C \in p$, so that $d^*(C) > 0$. Choose $\epsilon > 0$ such that $\mu([-\epsilon, \epsilon]) < \frac{d^*(C)}{2}$. Now $d^*\big(\{n \in E : g(n) \in [-\epsilon, \epsilon]\}\big) = d^*(E)\mu([-\epsilon, \epsilon]) < \frac{d^*(C)}{2}$. But $C \setminus \{n \in E : g(n) \in [-\epsilon, \epsilon]\}$ is finite. $\square$

3.7. Theorem. Let $p \in \mathcal{D}$, let $B \in \mathcal{RG}$ with $|B(n)| \leq T < \infty$ for all $n \in \mathbf{N}$, and put $f(n) = B(n)n^k$, where $k > 1$. If

$$\big|p\text{-}\lim_m p\text{-}\lim_n \frac{f(n+m) - f(n) - f(m)}{n^{k-1}}\big| = L < \infty \tag{3.2}$$

then $B \approx_p 0$.

Proof. Let

$$\begin{aligned} C =& \big\{m : \big|p\text{-}\lim_n \big(B(n+m)n + kmB(n+m) - B(n)n\big)\big| \le 2L\big\} \\ =& \big\{m : \big|p\text{-}\lim_n \frac{B(n+m)n^k + kmn^{k-1}B(n+m) - B(n)n^k}{n^{k-1}}\big| \le 2L\big\} \in p. \end{aligned}$$

Notice that

$$\begin{aligned} C \subset& \big\{m : \big|p\text{-}\lim_n \ n\big(B(n+m) - B(n)\big)\big| \le 2L + knT\big\} \\ \subset& \big\{m : \{n : B(n+m) = B(n)\} \in p\big\}. \end{aligned}$$

(The last containment utilizes Lemma 3.6.) Since the set in the final line is a member of p, and since by Corollary 3.3 $B - \alpha$ is a p-VIP system for some $\alpha \in \mathbf{R}$, by Lemma 3.5 one has $B - \alpha \approx_p 0$, i.e. $B \approx_p \alpha$. But as (4) holds, plainly $\alpha = 0$. □

The above prepares us for the following crucial theorem.

3.8. Theorem. Let $p \in \mathcal{D}$.

(a) If $g \in \mathcal{G}$ with $l_{g,p} = 0$ and $\deg_p g = 1$, then there exists $\alpha \in \mathbf{R}$ such that $g(n) \approx_p [\alpha n + \frac{1}{2}]$. In particular, g is p-regular.

(b) $\mathcal{G}_p = \{g - l_{g,p} : g \in \mathcal{G}\}$ is p-acceptable.

Proof. (a) Write $g(n) = \sum_{i=0}^k B_i(n)n^i$, where B_i are bounded generalized polynomials. Assuming $k > 1$, since $p\text{-}\lim_{m,n} \big(g(n+m) - g(n) - g(m)\big) = 0$ the hypotheses of Theorem 3.7 are met and we can conclude that $B_k \approx_p 0$. Iterating this argument we get $B_i \approx_p 0$ for all $i > 1$. Therefore, $g(n) \approx_p B_0(n) + B_1(n)n$. In order for this to hold, one must have:

$$\big|p\text{-}\lim_m p\text{-}\lim_n \big(B_1(n+m)(n+m) - B_1(n)n - B_1(m)m\big)\big| = L < \infty.$$

Let $C = \big\{m : |p\text{-}\lim_n \big(B_1(n+m)(n+m) - B_1(n)n - B_1(m)m\big)| < 2L\big\} \in p$. Then letting $T = \sup |B_1(n)|$,

$$\begin{aligned} C \subset& \big\{m : |p\text{-}\lim_n \ n\big(B_1(n+m) - B_1(n)\big)| < 2L + 2Tm\big\} \\ \subset& \big\{m : \{n : B_1(m+n) = B_1(n)\} \in p\big\}. \end{aligned}$$

It follows that for $m \in C$,

$$\begin{aligned} &\big|p\text{-}\lim_n \ m\big(B_1(n) - B_1(m)\big)\big| \\ =& \big|p\text{-}\lim_n \ m\big(B_1(n+m)(n+m) - B_1(n)n - B_1(m)m\big)\big| \le 2L. \end{aligned}$$

Since this holds for every m in a member of p, letting $\alpha = p\text{-}\lim_n B_1(n)$ one has

$$p\text{-}\lim_m \big|m\big(\alpha - B_1(m)\big)\big| \le 2L,$$

which by Lemma 3.6 yields $\big(\alpha - B_1\big) \approx_p 0$, i.e. $B_1 \approx_p \alpha$.

At this stage we have shown that $g(n) \approx_p \alpha n + B_0(n)$. Since B_0 is bounded, we quickly see that for some $k \in \mathbf{Z}$, $g(n) \approx_p [\alpha n + \frac{1}{2}] + k$. Letting $E = \{n : \langle \alpha n \rangle < \frac{1}{4}\}$, one sees that $E \in p$ and $g(n+m) - g(n) - g(m) = -k$ for all $n, m \in E$. Therefore $k = 0$.

Part (b) now follows easily. $\square$

We believe that, for $p \in \mathcal{D}$ and $g \in \mathcal{G}$, $g - l_g$ is always p-regular, even when $\deg_p g > 1$. However, we don't require this. At any rate, combining Theorem 3.8 (b) with Theorem 2.48 yields:

3.9. Theorem. Suppose $(X, \mathcal{A}, \mu, T)$ is a weakly mixing system, $k \in \mathbf{N}$, $p \in \mathcal{D}$ and let $v_1, \ldots, v_k \in \mathcal{G}$ such that neither v_i nor $v_i - v_j$ is constant on any member of p, $1 \leq i \neq j \leq k$. If $f_0, \ldots, f_k \in L^\infty(X)$ then

$$p\text{-}\lim_n \int f_0 T^{v_1(n)} f_1 \cdots T^{v_k(n)} f_k \, d\mu = \prod_{i=0}^{k} \Big(\int f_i \, d\mu \Big).$$

As mentioned in the introduction, it would be useful to develop criteria for verifying when generalized polynomials of certain forms meet the condition of the foregoing theorem; i.e. do not vanish on any member of a (perhaps given) idempotent. Theorem 1.18 provides a satisfactory answer for the degree 2 case, but beyond this our current knowledge is minimal. (See, however, Theorem 3.68 below.)

We now turn ourselves to the proof of the following multiple recurrence theorem for p-acceptable families.

3.10. Theorem. Let $p \in \beta\mathbf{N}$ be idempotent, let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system, with $\mu(X) = 1$, let $A \in \mathcal{A}$ with $\mu(A) > 0$ and let $\mathcal{E}$ be a p-acceptable family. If $v_1, \ldots, v_k \in \mathcal{E}$ then

$$p\text{-}\lim_n \mu\big(A \cap T^{v_1(n)} A \cap \cdots \cap T^{v_k(n)} A\big) > 0.$$

Our method of proof does not give any lower bound on the size of the foregoing expression (other than that it is positive). However, in the $k = 1$ case, the optimal lower bound $\mu(A)^2$ can easily be shown to apply. We show this before proceeding.

3.11. Theorem. Let $\mathcal{H}$ be a separable Hilbert space, let T be a unitary operator on $\mathcal{H}$, let $p \in \beta\mathbf{N}$ be idempotent and suppose $\varphi \in \mathcal{B}_p$. For $f \in \mathcal{H}$, write $p\text{-}\lim_n T^{\varphi(n)} f = Pf$, where the limit is taken in the weak topology. P is the orthogonal projection onto $\mathcal{K}_T$.

Proof. The proof is by induction on $d = \deg_p \varphi$. For $d = 1$, this is just Theorem 2.25. Suppose then that the conclusion holds for all members of $\mathcal{B}_p$ having p-degree less than d. Letting $Pf = p\text{-}\lim_n T^{\varphi(n)} f$, we must show that $P = P^2$.

Let $f \in \mathcal{H}$ and write $f = f_1 + f_2$, where $f_1 \in \mathcal{K}_T$ and $f_2 \in \mathcal{K}_T^\perp$. Let $A = \{m : \deg_p D_m \varphi = d - 1\} \in p$, and for $m \in A$ and $h \in L^2(X)$, put $P_m h = p\text{-}\lim_n T^{D_m \varphi(n)} h$. Then by the induction hypothesis, P_m is the orthogonal projection onto $\mathcal{K}_T$, $m \in A$. Hence taking $x_n = T^{\varphi(n)} f_2$,

$$\begin{aligned}
& p\text{-}\lim_m p\text{-}\lim_n \langle x_n, x_{n+m}\rangle \\
=& p\text{-}\lim_m p\text{-}\lim_n \int f_2 T^{\varphi(n+m)-\varphi(n)} f_2 \, d\mu \\
=& p\text{-}\lim_m p\text{-}\lim_n \int T^{-\varphi(m)} f_2 T^{D_m\varphi(n)} f_2 \, d\mu \\
=& p\text{-}\lim_m \int T^{-\varphi(m)} f_2 P_m f_2 \, d\mu = 0.
\end{aligned}$$

By Theorem 2.45, one has $p\text{-}\lim_n x_n = 0$ weakly; that is, $Pf_2 = 0$. On the other hand, just as in the proof of Theorem 2.25, one has $Pf_1 = f_1$ by Proposition 2.11. □

3.12. Corollary. Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system, let $p \in \mathcal{D}$, let $\varphi \in \mathcal{B}_p$ and suppose $\mu(A) > 0$. Then $p\text{-}\lim_n \mu(A \cap T^{\varphi(n)} A) \geq \mu(A)^2$.

Proof. Let $f = 1_A$. One has $p\text{-}\lim_n \mu(A \cap T^{\varphi(n)} A) = p\text{-}\lim_n \big\langle f, T^{-\varphi(n)} f\big\rangle = \big\langle f, Pf\big\rangle = \big\langle Pf, Pf\big\rangle \geq \mu(A)^2$. (For the final inequality, we used the fact that P is the orthogonal projection onto a space containing the constants.) □

3.13. Discussion: Lebesgue spaces. For the proof of Theorem 3.10 in general, we will make (without loss of generality, for reasons that are quite standard) the assumption that $(X, \mathcal{A}, \mu, T)$ is an ergodic system on a Lebesgue space. We now proceed to collect basic facts concerning such systems.

Let $\mathcal{B} \subset \mathcal{A}$ be a complete, T-invariant σ-algebra. $\mathcal{B}$ determines a factor $(Y, \mathcal{B}_1, \nu, S)$ of $(X, \mathcal{A}, \mu, T)$ in the following way. Let $(B_i)_{i=1}^\infty \subset \mathcal{B}$ be a dense (in $\mathcal{B}$ as a measure algebra), T-invariant sequence of sets; let Y be the set of equivalence classes under the equivalence relation $x \approx y$ iff $x \in B_i$ if and only if $y \in B_i$, $i \in \mathbf{N}$. 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 the desired factor. In a slight abuse of terminology that has become wholly usual, 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 $f \in L^1(X, \mathcal{A}, \mu)$ one has

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

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 $f_1, f_2 \in L^\infty(X, \mathcal{A}, \mu)$

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

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 as presented in [FKO].

3.14. Definition. 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 $f \in L^2(X, \mathcal{A}, \mu)$.

(a) 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 $s = s(n, y)$, $1 \le s \le k$, such that $||T^n f - g_s||_{L^2(X, \mathcal{A}, \mu_y)} < \delta$.

(b) If AP 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 $\mathcal{B}$.

(c) 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 $\mathcal{B}$.

3.15. Notation. In the future we will write $||f||_y$ for $||f||_{L^2(X, \mathcal{A}, \mu_y)}$.

For a proof of the following proposition, see [FKO].

3.16. Proposition. 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 $1_{A'} \in AP$.

3.17. Remark. 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.

3.18. Theorem. 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 $\{\mathcal{A}_\xi \subset \mathcal{A} :\ \xi \le \eta\}$ 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, $\mathcal{A}_\xi$ is the complete 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 $\mathcal{A}_\eta$.

3.19. Remark. The factor $\mathcal{A}_\eta$ appearing in the structure theorem is called the *maximal distal factor* of $\mathcal{A}$. Our present goal is to use a relativized version of Theorem 2.48 to show that in order to prove Theorem 3.10 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.

3.20. Theorem. Let $(X, \mathcal{A}, \mu, T)$ be an ergodic system. Suppose $\mathcal{B} \subset \mathcal{A}$ is a T-invariant σ-algebra such that $\mathcal{A}$ is a relatively weakly mixing extension of $\mathcal{B}$. Suppose $p \in \beta\mathbf{N}$ is idempotent and let $\mathcal{E}$ be a p-acceptable family. If $v_1(x), \cdots, v_k$ are distinct members of $\mathcal{E}$, then writing $v_0 \equiv 0$, for any $f_0, \cdots, f_k \in L^\infty(X, \mathcal{A}, \mu)$,

$$p\text{-}\lim_n \left|\left| \prod_{i=0}^k T^{v_i(n)} f_i - \prod_{i=0}^k T^{v_i(n)} E(f_i|\mathcal{B}) \right|\right| = 0.$$

Proof. Let $a_i = T^{v_i(n)} f_i$, $b_i = T^{v_i(n)} E(f_i|\mathcal{B})$ and employ (2.10). This allows one to assume without loss of generality that $E(f_a|\mathcal{B}) = 0$ for some a, whereupon one proceeds exactly as in the proof of Theorem 2.48. $\square$

3.21. Remark. The above theorem allows one to assume in the proof of Theorem 3.10 that the system one is working in is a *distal system* (that is, one that is equal to its maximal distal factor). However, subject to this restiction, an even more general statement than that appearing in the conclusion to Theorem 3.10 holds–and, owing to somewhat simpler notation, is easier to prove. Formulating this statement and proving that it is valid for distal systems is our next task.

3.22. Definition. Let $\mathcal{F}$ denote the family of finite, non-empty subsets of $\mathbf{N}$. Put also $\mathcal{F}_\emptyset = \mathcal{F} \cup \{\emptyset\}$. For $\alpha, \beta \in \mathcal{F}_\emptyset$, we write $\alpha < \beta$ if for all $x \in \alpha$ and $y \in \beta$, one has $x < y$.

Note that $\emptyset < \alpha < \emptyset$ for $\alpha \in \mathcal{F}$.

3.23. Notation. Suppose $\alpha_i \in \mathcal{F}$, $i \in \mathbf{N}$, with $\alpha_1 < \alpha_2 < \cdots$. Let $\mathcal{F}^{(1)}$ denote the set of non-empty finite unions of the α_i's. (Also $\mathcal{F}_\emptyset^{(1)} = \mathcal{F}^{(1)} \cup \{\emptyset\}$.) $\mathcal{F}^{(1)}$ is called an *IP ring*.

3.24. Definition. Let $k \in \mathbf{N}$ and let $\mathcal{F}^{(1)}$ be an IP ring. We denote by $(\mathcal{F}^{(1)})^k_<$ (respectively $(\mathcal{F}_\emptyset^{(1)})^k_<$) the set of all k-tuples $(\alpha_1, \ldots, \alpha_k) \in (\mathcal{F}^{(1)})^k$ (respectively $\in (\mathcal{F}_\emptyset^{(1)})^k$) satisfying $\alpha_1 < \alpha_2 < \cdots < \alpha_k$.

3.25. Theorem. (*Milliken-Taylor Theorem*: see [Mi], [T]). Let $k \in \mathbf{N}$, let $\mathcal{F}^{(1)}$ be an IP ring and suppose $(\mathcal{F}^{(1)})^k_< = \bigcup_{i=1}^r C_i$. Then for some IP ring $\mathcal{F}^{(2)} \subset \mathcal{F}^{(1)}$ and some i, $1 \leq i \leq r$, one has $(\mathcal{F}^{(2)})^k_< \subset C_i$.

3.26. Definition. A function $v : \mathcal{F} \to \mathbf{Z}$ is a *VIP system* if there exists $d \in \mathbf{N}$ such that for every $(\alpha_0, \ldots, \alpha_d) \in \mathcal{F}^{d+1}_<$ one has

$$\sum_{0 \leq i_i < i_2 < \cdots < i_k \leq d} (-1)^k v(\alpha_{i_1} \cup \cdots \cup \alpha_{i_k}) = 0.$$

The least such d is called the *degree* of the system.

3.27. Definition. Suppose that Ω is a topological space, $\omega : \mathcal{F} \to \Omega$ is a function, $z \in \Omega$ and $\mathcal{F}^{(1)}$ is an IP-ring. We write

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} \omega(\alpha) = z$$

if for any neighborhood U of z, there exists $\alpha_0 \in \mathcal{F}^{(1)}$ such that for all $\alpha \in \mathcal{F}^{(1)}$ with $\alpha > \alpha_0$, one has $\omega(\alpha) \in U$. Similarly, if $g : \mathcal{F}^t_< \to \Omega$ we write

$$\operatorname*{IP-lim}_{(\alpha_1,\cdots,\alpha_t)\in(\mathcal{F}^{(1)})^t_<} g(\alpha_1,\dots,\alpha_t) = z$$

if for any neighborhood U of z, there exists $\alpha_0 \in \mathcal{F}^{(1)}$ such that for all $(\alpha_1,\dots,\alpha_t) \in (\mathcal{F}^{(1)})^t_<$ with $\alpha_1 > \alpha_0$, one has $g(\alpha_1,\dots,\alpha_t) \in U$.

3.28. Remark. Using Hindman's theorem in place of the pigeonhole principle, one can prove, by mimicking the proof of the Bolzano-Weierstrass theorem, that if Ω is a compact metric space and $\omega : \mathcal{F} \to \Omega$ is an arbitrary function then one can find an IP ring $\mathcal{F}^{(1)}$ such that $\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} \omega(\alpha)$ exists. (Similarly for $\mathcal{F}^t_<$-valued functions, using the Milliken-Taylor theorem.)

3.29. Theorem. Let $(X,\mathcal{A},\mu,T)$ be a distal system. For any VIP systems $u_1,\dots,u_k$ and any $A \in \mathcal{A}$ with $\mu(A) > 0$, there exists an IP ring $\mathcal{F}^{(1)}$ such that

$$\operatorname*{IP-lim}_{\alpha\in\mathcal{F}^{(1)}} \mu\Big(\bigcap_{i=1}^{k} T^{u_i(\alpha)}A\Big) > 0.$$

We do not know whether Theorem 3.29 is true for arbitrary systems. At any rate, our first task is to use its validity for distal systems to derive Theorem 3.10, with the help of Theorem 3.20.

3.30. Theorem. Let $p \in \beta\mathbf{N}$ be idempotent and suppose $v_1,\dots,v_k$ are p-VIP systems. There exists an IP system $n : \mathcal{F} \to \mathbf{N}$ such that, writing $u_i(\alpha) = v_i\big(n(\alpha)\big)$, each u_i is a VIP system, $1 \le i \le k$.

Proof. Although this result is not deep, notation can be cumbersome. For convenience, we will give a proof for $k = 1$ and $\deg_p v_1 = 1$, leaving the more complicated, albeit uninterestingly so, general case to the reader.

Let $E = E_0 = \big\{x : \{y : v_1(x+y) = v_1(x) + v_1(y)\} \in p\big\} \in p$, and for $x \in E$ put $E_x = \{y > x : v_1(x+y) = v_1(x) + v_1(y)\} \in p$. We inductively generate a sequence $(x_i)_{i=1}^\infty$. Assume that $x_1,\dots,x_{t-1}$ have been chosen such that for every $\alpha,\beta \subset \{1,2,\dots,t-1\}$ with $\alpha \neq \emptyset$ and $\alpha < \beta$ (perhaps β is empty), one has, writing in a small abuse of notation $x_\gamma = \sum_{i\in\gamma} x_i$,

(1) $(E_{x_\alpha} - x_\beta) \in p$, and

(2) $(E - x_\alpha) \in p$.

Here is how x_t is to be chosen. Let

$$D = \Big(\bigcap_{\alpha\subset\{1,\dots,t-1\}} (E - x_\alpha)\Big) \cap \Big(\bigcap_{\alpha,\beta\subset\{1,\dots,t-1\},\emptyset\neq\alpha<\beta} (E_{x_\alpha} - x_\beta)\Big) \in p.$$

Note that α can be empty in the first intersection and β can be empty in the second. Now choose $x_t \in D \cap \{x : D - x \in p\}$. One checks that the induction hypothesis is carried forward and that $n(\alpha) = x_\alpha$ does the job. □

3.31. Remark. In our application of Theorem 3.30 we will need something more, which is an easy corollary of the proof. Namely, if $v_1, \dots, v_k$ are p-VIP systems and $r : \mathbf{N} \to \mathbf{R}$ is a function for which $p\text{-}\lim_n r(n) = x$, then there exists an IP system $n : \mathcal{F} \to \mathbf{N}$ such that, writing $u_i(\alpha) = v_i\big(n(\alpha)\big)$, each u_i is a VIP system, $1 \le i \le k$, with the additional property that $\text{IP-}\lim_{\alpha\in\mathcal{F}} r\big(n(\alpha)\big) = x$.

3.32. Proof of Theorem 3.10. Let $\mathcal{B} = \mathcal{A}_\eta \subset \mathcal{A}$ denote the maximal distal factor of the system $(X, \mathcal{A}, \mu, T)$. Take $A \in \mathcal{A}$ with $\mu(A) > 0$ and let $f = E(1_A|\mathcal{B})$. Choose $\delta > 0$ and $B \in \mathcal{B}$ such that $f \ge \delta$ on B. By Theorem 3.20 (using just weak convergence),

$$\begin{aligned} p\text{-}\lim_n \mu\big(B \cap T^{v_1(n)}B \cap \cdots \cap T^{v_k(n)}B\big) =& p\text{-}\lim_n \delta^{k+1} \int 1_B T^{-v_1(n)} 1_B \cdots T^{-v_k(n)} 1_B \\ \le& p\text{-}\lim_n \int f T^{-v_1(n)} f \cdots T^{-v_k(n)} f \\ =& p\text{-}\lim_n \int 1_A T^{-v_1(n)} 1_A \cdots T^{-v_k(n)} 1_A \\ & p\text{-}\lim_n \mu\big(A \cap T^{v_1(n)}A \cap \cdots \cap T^{v_k(n)}A\big). \end{aligned}$$

Let $r(n) = \mu\big(B \cap T^{v_1(n)}B \cap \cdots \cap T^{v_k(n)}B\big)$. By the above inequality, it suffices to show $p\text{-}\lim_n r(n) > 0$. Suppose that $p\text{-}\lim_n r(n) = 0$. We will obtain a contradiction. Choose by Remark 3.31 an IP system $n : \mathcal{F} \to \mathbf{N}$ such that, writing $u_i(\alpha) = v_i\big(n(\alpha)\big)$, $1 \le i \le k$, each u_i is a VIP system, and such that

$$\text{IP-}\lim_{\alpha\in\mathcal{F}} r\big(n(\alpha)\big) = \text{IP-}\lim_{\alpha\in\mathcal{F}} \mu\big(B \cap T^{u_1(\alpha)}B \cap \cdots \cap T^{u_k(\alpha)}B\big) = 0.$$

This contradicts Theorem 3.29. □

We now turn our attention to the proof of Theorem 3.29.

3.33. Definition. Let $d \in \mathbf{N}$. Put $\mathcal{F}_d = \{\alpha \in \mathcal{F} : |\alpha| \le d\}$.

3.34. Theorem. (See [Mc1].) A function $v : \mathcal{F} \to \mathbf{Z}$ is a VIP system if and only if there exists $d \in \mathbf{N}$ and a function $\gamma : \mathcal{F}_d \to \mathbf{Z}$ (the *generating function*) such that $v(\alpha) = \sum_{\beta\subset\alpha, 0<|\beta|\le d} \gamma(\beta)$ for all $\alpha \in \mathcal{F}$.

The foregoing characterization motivates the following definition.

3.35. Definition. Let $t \in \mathbf{N}$. A multivariable VIP-system (with t variables) of degree (at most) d into $\mathbf{Z}$ is a function $v : (\mathcal{F}_\emptyset)^t_< \to \mathbf{Z}$ of the form

$$v(\alpha_1, \ldots, \alpha_t) = \sum_{\gamma_i \subset \alpha_i, |\gamma_1 \cup \cdots \cup \gamma_t| \le d} f(\gamma_1, \ldots, \gamma_t),$$

where f takes on arbitrary values in $\mathbf{Z}$ with $f(\emptyset, \ldots, \emptyset) = 0$.

The following is a standard application of the polynomial Hales-Jewett theorem ([BL2]; cf. [BM1, Corollary 3.2] and [BM2, Theorem 2.12]).

3.36. Lemma. Let $k, r, t \in \mathbf{N}$ and let $v_j(\alpha_1, \ldots, \alpha_t)$ be multivariable VIP systems, $1 \le j \le k$. There exist $N, w \in \mathbf{N}$ and multivariable VIP systems $q_j(\alpha_1, \ldots, \alpha_N)$, $1 \le j \le w$, such that for any r-cell partition $Q = \{q_j : 1 \le j \le w\} = \bigcup_{j=1}^r C_j$, there is some i with $1 \le i \le r$, some $q \in Q$, and sets $S_j \subset \{1, \ldots, N\}$, $1 \le j \le t$, with $S_1 < S_2 < \cdots < S_t$, such that, utlizing the substitution $\beta_j = \bigcup_{n \in S_j} \alpha_n$, $1 \le j \le t$, one has

$$\big\{q(\alpha_1, \ldots, \alpha_N), q(\alpha_1, \ldots, \alpha_N) - v_1(\beta_1, \ldots, \beta_t), \ldots, q(\alpha_1, \ldots, \alpha_N) - v_k(\beta_1, \ldots, \beta_t)\big\} \subset C_i.$$

3.37. Definition. Suppose $(X, \mathcal{A}, \mu, T)$ is an invertible measure preserving system and $\mathcal{B} \subset \mathcal{A}$ is a complete T-invariant sub-σ-algebra. $\mathcal{B}$ is said to have the *VIPSZ property* if for every $A \in \mathcal{B}$ with $\mu(A) > 0$, $t, k \in \mathbf{N}$, and multivariable VIP systems $v_1(\alpha_1, \ldots \alpha_t), \ldots, v_k(\alpha_1, \ldots, \alpha_t)$ there is an IP ring $\mathcal{F}^{(1)}$ such that, writing $v_0 \equiv 0$,

$$\operatorname*{IP-lim}_{(\alpha_1, \cdots, \alpha_t) \in (\mathcal{F}^{(1)})^t_<} \mu\Big(\bigcap_{j=0}^k T^{v_j(\alpha_1, \ldots, \alpha_t)} A\Big) > 0.$$

3.38. Discussion. In order to prove Theorem 3.29, it suffices to show that for an arbitrary distal system $(X, \mathcal{A}, \mu, T)$, $\mathcal{A}$ has the VIPSZ property. This is accomplished via transfinite induction through the ordinals appearing in Theorem 3.18. As the trivial algebra $\mathcal{A}_0$ plainly has the property, there are two steps: one must show that the property passes to successor ordinals (compact extensions), and to limit ordinals.

First we handle successor ordinals.

3.39. Theorem. 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 VIPSZ 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 VIPSZ property as well.

Proof. Suppose that $A \in \mathcal{A}$, $\mu(A) > 0$. By Proposition 3.16 we may assume without loss of generality that $f = 1_A \in AP$. Suppose that $t, k \in \mathbf{N}$ and $v_i(\alpha_1, \ldots, \alpha_t)$ are multivariable VIP systems, $1 \le i \le k$. 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 $f \in AP$, there exist functions $g_1, \ldots, g_r \in L^2(X, \mathcal{A}, \mu)$ having the property that for any $n \in \mathbf{N}$, and a.e. $y \in Y$, there exists $s = s(n, y)$, $1 \le s \le r$, such that $\big|\big|T^n f - g_s\big|\big|_y < \epsilon$. For these numbers r, k, t and the systems v_i, let $w, N \in \mathbf{N}$ and $Q = \{q_j(\alpha_1, \ldots, \alpha_N) : 1 \leq j \leq w\}$ be as guaranteed by Lemma 3.36.

Write $q_0 \equiv 0$. Since $\mathcal{B}$ has the VIPSZ property, there exists an IP ring $\mathcal{F}^{(1)}$ such that

$$\underset{(\alpha_1,\cdots,\alpha_N)\in(\mathcal{F}^{(1)})^t_<}{\text{IP-lim}} \mu\Big(\bigcap_{j=0}^{N} S^{q_j(\alpha_1,\ldots,\alpha_N)} B\Big) > \eta > 0.$$

We may further assume that

$$\underset{(\alpha_1,\cdots,\alpha_t)\in(\mathcal{F}^{(1)})^t_<}{\text{IP-lim}} \mu\Big(\bigcap_{j=0}^{N} T^{v_j(\alpha_1,\ldots,\alpha_t)} A\Big) \tag{3.1}$$

exists. It is this limit we wish to show positive. Let D be the number of ways of choosing t non-empty sets $S_1, \cdots, S_t \subset \{1, \cdots, N\}$ with $S_1 < S_2 < \cdots < S_t$ and set $\delta = \frac{c\eta}{2D}$.

If the limit (3.1) were equal to zero, then by Milliken-Taylor we could, by passing to a subring of $\mathcal{F}^{(1)}$, assume that $\mu\Big(\bigcap_{j=0}^{N} T^{v_j(\alpha_1,\ldots,\alpha_t)} A\Big) < \delta$ for every $(\alpha_1, \ldots, \alpha_t) \in (\mathcal{F}^{(1)})^t_<$. We will show that this assumption is impossible by exhibiting $(\beta_1, \ldots, \beta_t) \in (\mathcal{F}^{(1)})^t_<$ for which $\mu\Big(\bigcap_{j=0}^{N} T^{v_j(\beta_1,\ldots,\beta_t)} A\Big) \geq \delta$. This will complete the proof of Theorem 3.40.

Choose $(\alpha_1, \ldots, \alpha_N) \in (\mathcal{F}^{(1)})^N_<$ such that

$$\mu\Big(\bigcap_{j=0}^{N} S^{-q_j(\alpha_1,\ldots,\alpha_N)} B\Big) > \eta.$$

Pick any $y \in \bigcap_{j=0}^{N} S^{-q_j(\alpha_1,\ldots,\alpha_N)} B$. Form an r-cell partition of Q, $Q = \bigcup_{i=1}^{r} C_i$, by $q_j \in C_i$ if and only if $s\big(q_j(\alpha_1, \ldots, \alpha_N), y\big) = i$, $1 \leq j \leq w$. In particular, if $q_j \in C_i$ then $\big|\big|T^{q_j(\alpha_1,\ldots,\alpha_N)} f - g_i\big|\big|_y < \epsilon$. For this partition, there exists some i, $1 \leq i \leq r$, some $q \in Q$, and sets $S_1, \ldots, S_t \subset \{1, \cdots, N\}$, $S_1 < S_2 < \cdots < S_t$, such that, utilizing the substitution $\beta_m = \bigcup_{n \in S_m} \alpha_n$, $1 \leq m \leq t$, one has

$$\big\{q(\alpha_1, \ldots, \alpha_N), q(\alpha_1, \ldots, \alpha_N) - v_1(\beta_1, \ldots, \beta_t), \ldots, q(\alpha_1, \ldots, \alpha_N) - v_k(\beta_1, \ldots, \beta_t)\big\} \subset C_i.$$

(We are exercising a slight abuse of notation here, of course; in the previous display as in Lemma 3.36 the α_i are variables, whereas in the flow of the proof they have been fixed. This obviously makes no difference.) This implies

$$\Big|\Big|T^{q(\alpha_1,\ldots,\alpha_N)-v_j(\beta_1,\ldots,\beta_t)} f - g_i\Big|\Big|_y < \epsilon,\ 0 \leq j \leq k.$$

Setting $\tilde{y} = S^{q(\alpha_1,\ldots,\alpha_N)} y$, one has

$$\Big|\Big|T^{-v_j(\beta_1,\ldots,\beta_t)} f - T^{-q(\alpha_1,\ldots,\alpha_N)} g_i\Big|\Big|_{\tilde{y}} < \epsilon,\ 0 \leq j \leq k.$$

In particular, since this holds for $j = 0$, by the triangle inequality

$$\left|\left|T^{-v_j(\beta_1,\ldots,\beta_t)}f - f\right|\right|_{\tilde{y}} < 2\epsilon, \quad 1 \le j \le k.$$

It follows that

$$\mu_{\tilde{y}}\big(A \setminus T^{v_j(\beta_1,\ldots,\beta_t)}A\big) \le \left|\left|T^{-v_j(\beta_1,\ldots,\beta_t)}f - f\right|\right|^2_{\tilde{y}} \le 4\epsilon^2; \quad 1 \le j \le k.$$

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

$$\mu_{\tilde{y}}\Big(\bigcap_{i=0}^{k} T^{v_i(\beta_1,\ldots,\beta_t)}A\Big) \ge c - 4k\epsilon^2 = \frac{c}{2}.$$

$S_1, \ldots, S_t$ depend measurably on y, therefore $\beta_1, \ldots, \beta_t$ are measurable functions of y defined on the set $\Big(\bigcap_{j=0}^{w} S^{-q_j(\alpha_1,\ldots,\alpha_N)}B\Big)$, which, recall, is of measure greater than η. Hence, as there are only D choices for $S_1, \ldots, S_t$, we may assume that for all $y \in H$, where $H \in \mathcal{B}$ satisfies $\nu(H) > \frac{\eta}{D}$, $\beta_1, \ldots, \beta_t$ are constant. For this choice of $\beta_1, \ldots, \beta_t$ one has

$$\mu\Big(\bigcap_{j=0}^{k} T^{v_i(\beta_1,\ldots,\beta_t)}A\Big) \ge \frac{c}{2}\nu(H) > \frac{c\eta}{2D} = \delta.$$

□

Next we handle passage to limit ordinals.

3.40. Theorem. Suppose that $(X, \mathcal{A}, \mu, T)$ is a measure preserving system and $\mathcal{A}_\xi$ is a totally ordered chain of sub-σ-algebras of $\mathcal{A}$ having the VIPSZ property. If $\bigcup_\xi \mathcal{A}_\xi$ is dense in $\mathcal{A}$ then $\mathcal{A}$ has the VIPSZ property.

Proof. Suppose $A \in \mathcal{A}$, $\mu(A) > 0$, $t, k \in \mathbf{N}$ and $v_i(\alpha_1, \cdots, \alpha_t)$ are VIP systems, $1 \le i \le k$. There exists an ordinal ξ and a set $B \in \mathcal{A}_\xi$ 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 μ over $\mathcal{A}_\xi$. We will speak of the system $(Y, \mathcal{A}_\xi, \nu, S)$, as usual. 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 VIPSZ property, there exists an IP ring $\mathcal{F}^{(1)}$ such that

$$\operatorname*{IP-lim}_{(\alpha_1,\cdots,\alpha_t)\in(\mathcal{F}^{(1)})^t_<} \nu\Big(\bigcap_{j=0}^{k} S^{v_j(\alpha_1,\cdots,\alpha_t)}C\Big) > \delta > 0.$$

By passing to a subring if necessary, we may assume that

$$\operatorname*{IP-lim}_{(\alpha_1,\cdots,\alpha_t)\in(\mathcal{F}^{(1)})^t_<} \mu\Big(\bigcap_{j=0}^{k} T^{v_j(\alpha_1,\cdots,\alpha_t)}A\Big)$$

exists. We claim that this limit is at least $\frac{\delta}{2}$. Otherwise, by Milliken-Taylor one could again pass to an IP subring (continue to call it $\mathcal{F}^{(1)}$) having the property that for all $(\alpha_1, \ldots, \alpha_t) \in (\mathcal{F}^{(1)})^t_<$, $\mu\Big(\bigcap_{j=0}^k T^{v_j(\alpha_1,\cdots,\alpha_t)}A\Big) < \frac{\delta}{2}$. But one may pick $(\alpha_1, \ldots, \alpha_t) \in (\mathcal{F}^{(1)})^t_<$ for which $\nu\Big(\bigcap_{j=0}^k S^{v_j(\alpha_1,\cdots,\alpha_t)}C\Big) > \delta$. Also, for every $y \in \bigcap_{j=0}^k S^{v_j(\alpha_1,\cdots,\alpha_t)}C$, one has $\mu_y\big(T^{v_j(\alpha_1,\cdots,\alpha_t)}A\big) \geq 1 - \frac{1}{2(k+1)}$, from which it follows that $\mu_y\Big(\bigcap_{j=0}^k T^{v_j(\alpha_1,\cdots,\alpha_t)}A\Big) \geq \frac{1}{2}$, whence $\mu\Big(\bigcap_{j=0}^k T^{v_j(\alpha_1,\cdots,\alpha_t)}A\Big) \geq \frac{\delta}{2}$. □

This completes the proof of Theorem 3.29, and hence also of Theorem 3.10. Combining this with Furstenberg correspondence, we get the following corollary.

3.41. Corollary. Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$ and let $g_1, \ldots, g_k \in \mathcal{G}$. Suppose $p \in \mathcal{D}$ and $A \in p$. Then there exists $n \in A$ such that $d^*\big(E\cap(E+g_1(n)-l_{g_1})\cap\cdots\cap(E+g_k(n)-l_{g_k})\big) > 0$.

Proof. Recall, we have shown in Lemma 3.8 b. that $\mathcal{G}_p = \{g - l_{g,p} : g \in \mathcal{G}\}$ is p-acceptable for $p \in \mathcal{D}$. Combining this with Theorem 3.10 and Furstenberg correspondence gives the desired result. □

3.42. Remark. Part (a) of the above theorem is, in effect, a special case of [Mc2, Theorem B], which essentially gives the same result for $\mathbf{Z}^l$-valued generalized polynomials of finitely many variables, and for (up to an equivalence of formulation) arbitrary idempotents p. (Only "in effect" because [Mc2, Theorem B] is restricted to admissible generalized polynomials g (see the introduction), for which $l_{g,p} = 0$ for every idempotent p. However, one can use [BKM, Theorem 2.8] to pass from these to arbitrary generalized polynomials.) Apart from the more general formulation (b), an advantage of the proof we have given here is that it uses the classical structure theory of [FKO], whereas [Mc2] (also [BM2]) employs a substantially more complicated structure theory analgous to that of [FK2].

However, there are further advantages as well; taking the VIP, degree 2 case as a paradigm, the methods of [BKM] and [Mc2] apply only to linear combinations of "basic" systems $v(\alpha) = \sum_{\{i,j\}\subset\alpha} d_{\{i,j\}}$ whose generators $(d_{\{i,j\}})$ satisfy a restrictive algebraic condition, say $d_{\{i,j\}} = n_i m_j$, where (n_i) and (m_j) are sequences of integers. The methods we employ here are sensitive rather to growth-rate and dependence conditions on the generators. We now illustrate this by way of two examples.

3.43. Definition. Let $(d_i)_{i=0}^\infty$ be a sequence in $\mathbf{Z}$ generating an IP system $v(\alpha) = \sum_{i\in\alpha} d_i$. Now, for $x \in \mathbf{N}$, write $x = \sum_{i=1}^\infty a_i(x)2^i$, with each $a_i(x) \in \{0,1\}$ (this is of course just the unique binary expansion of x). We put $f_v(x) = \sum_{i=0}^\infty a_i(x)d_i$.

3.44. Proposition. If $p \in \beta\mathbf{N}$ is any idempotent and v is an IP system then f_v is a p-IP system.

Proof. Let $x \in \mathbf{N}$ and put $m(x) = \max\{i : a_i(x) = 1\}$. One easily checks that $f_v(x+y) = f_v(x) + f_v(y)$ for any y divisible by 2^{m+1}. □

3.45. Definition. An IP system $v(\alpha) = \sum_{i\in\alpha} d_i$, $\min\alpha \geq i_0 \in \mathbf{N}$ is *well spaced* if for some $M \in \mathbf{N}$, either $\sum_{i=i_0}^{n-1} d_i < d_n < M2^n$ for all $n > i_0$ or $\sum_{i=i_0}^{n-1} d_i > d_n > -M2^n$ for all $n > i_0$.

Notice that non-zero multiples of well spaced systems are well spaced, but finite sums of well spaced systems needn't be.

3.46. Proposition. Let $v(\alpha) = \sum_{i\in\alpha} d_i$ be a well spaced IP system. If $E \subset \mathbf{N}$ with $d^*(E) = \epsilon > 0$ then $d^*\big(f_v(E)\big) > 0$.

Proof. We assume for convenience that $i_0 = 1$. Choose a large n and a k such that $\big|E \cap [k2^n, (k+1)2^n)\big| > \frac{1}{2}\epsilon 2^n$. If M is such that $\sum_{i=1}^{n-1} d_i < d_n < M2^n$ for all $n > 1$,

$$f_v\big([k2^n, (k+1)2^n)\big) \subset [f_v(k2^n), f_v(k2^n) + M2^n).$$

Moreover, f_v is injective, hence

$$\frac{\big|f_v(E) \cap [f_v(k2^n), f_v(k2^n) + M2^n)\big|}{M2^n} > \frac{\epsilon}{2M}.$$

Since n was chosen arbitrarily large, $d^*\big(f_v(E)\big) \geq \frac{\epsilon}{2M}$. $\square$

Recall that a function $g : \mathcal{F}_d \to \mathbf{Z}$ generates a VIP system by $v(\alpha) = \sum_{\gamma\subset\alpha} g(\gamma)$.

3.47. Definition.
(a) For $x \in \mathbf{N}$, write $\alpha(x) = \{i : a_i(x) = 1\}$.
(b) Let $v : \mathcal{F} \to \mathbf{Z}$ be a VIP system. For $x \in \mathbf{N}$, we put $f_v(x) = v\big(\alpha(x)\big)$.

3.48. Proposition. Let $v : \mathcal{F} \to \mathbf{Z}$ be VIP of degree at most d and let $p \in \beta\mathbf{N}$ be idempotent. Then f_v is a p-VIP system of degree at most d.

3.49. Definition. Let v be a VIP system. If $\deg v = d > 1$, v is said to be *well spaced* if (i) $v(\alpha) = \sum_{\gamma\subset\alpha, |\gamma|=d} g(\gamma) + \sum_{\gamma\subset\alpha,\gamma\in\mathcal{F}_{d-1}} h(\gamma)$, $\min\alpha \geq i_0$, where g is strictly positive; and (ii) for every $\alpha \in \mathcal{F}_{d-1}$, the sequence $d_i = g(\alpha \cup \{i\})$, $i > \max\alpha$, generates a well spaced IP system. f_v is said to be well spaced if v is well spaced.

3.50. Definition. Let v be a VIP system into $\mathbf{Z}$ and let $\alpha \in \mathcal{F}$. The *derivative of v with step α* is the function $D_\alpha v : \{\beta \in \mathcal{F} : \min\beta > \max\alpha\} \to \mathbf{Z}$ defined by $D_\alpha v(\beta) = v(\alpha \cup \beta) - v(\alpha) - v(\beta)$.

3.51. Proposition. Let $x, y \in \mathbf{N}$ and let $v : \mathcal{F} \to \mathbf{Z}$ be a VIP system. If $\alpha(y) < \alpha(x)$ then $D_y f_v(x) = D_{\alpha(y)} v\big(\alpha(x)\big)$.

We leave the easy proof of the following to the reader.

3.52. Proposition. Let $v : \mathcal{F} \to \mathbf{N}$ be a well spaced VIP system of degree $d > 1$ and fix $\alpha \in \mathcal{F}$. Then $D_\alpha v$ is well spaced.

3.53. Corollary. Let $p \in \mathcal{D}$ and let v be a well spaced VIP system. Then $f_v \in \mathcal{B}_p$.

Combining this with Corollary 3.12, one gets:

3.54. Corollary. Let $v : \mathcal{F} \to \mathbf{Z}$ be a well spaced VIP system and let $\epsilon > 0$.

a. Suppose $(X, \mathcal{A}, \mu, T)$ is a measure preserving system. If $\mu(A) > 0$ then there is some $\alpha \in \mathcal{F}$ such that $\mu(A \cap T^{v(\alpha)}A) \geq \mu(A)^2 - \epsilon$.

b. If $E \subset \mathbf{Z}$ with $d^*(E) > 0$ then there is some $\alpha \in \mathcal{F}$ such that

$$d^*\Big(E \cap \big(E - v(\alpha)\big)\Big) > d^*(E)^2 - \epsilon.$$

In order to obtain results for multiple recurrence along well spaced VIP systems, we need some additional conditions. At this stage, notation can become rather cumbersome, so we will restrict ourselves to the degree 2 case for convenience.

3.55. Theorem. Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system. Let

$$v(\alpha) = \sum_{i<j, \{i,j\} \subset \alpha} d_{\{i,j\}} + \sum_{i \in \alpha} c_i$$

be a well spaced VIP system of degree 2 with $d_{\{i,j\}}$ positive. For $i < j$, put $\mathrm{gap}(i, j) = d_{\{i,j\}} - \sum_{k=i+1}^{j-1} d_{\{i,k\}} > 0$, and suppose that for every $L, M > 0$, there exists $N = N(v, L, M) > 0$ such that for every $j > i \geq N$,

$$\mathrm{gap}(i, j) > M\big(\sum_{t=1}^{L} \mathrm{gap}(t, j)\big). \tag{$*$}$$

If $A \in \mathcal{A}$ with $\mu(A) > 0$ then there is some $\alpha \in \mathcal{F}$ such that $\mu\big(A \cap T^{v(\alpha)} A \cap T^{2v(\alpha)} A\big) > 0$.

3.56. Remark. If $d_{\{i,j\}}$ is negative, simply reverse the inequality in $(*)$. Notice that the requirement $\mathrm{gap}(i, j) > 0$ is just the well spacedness condition, whereas $(*)$ implies in particular that for any $\beta \in \mathcal{F}$ with $\min \beta \geq N$ and any $\alpha \in f$ with $\max \alpha \leq L$, $D_\beta v + M' D_\alpha v$ is well spaced, where $|M'| \leq M$. More generally, if $\alpha_1 < \alpha_2 < \cdots < \alpha_{r-1}$, with $\max \alpha_{r-1} \leq L$, $D_\beta v - \sum_{i=1}^{r-1} M_i D_{\alpha_i} v$ will be well spaced, where $|M_i| \leq M$.

Condition $(*)$ may be less artificial than it at first seems; e.g. degree two polynomials $cx^2 + bx$ can be realized as f_v, where v is well spaced and satisfies $(*)$. Indeed, put $w(\alpha) = \sum_{i<j, \{i,j\} \subset \alpha} d_{\{i,j\}}$, where $d_{\{i,j\}} = c2^{i+j+1}$, and let u be the IP system defined by $u(\alpha) = \sum_{i \in \alpha} (c2^{2i} + b2^i)$. One may easily check that $v = w + u$ is well spaced, $f_v(n) = cn^2 + bn$ and $(*)$ holds.

Proof of Theorem 3.55. Since well spaced VIP systems satisfying $(*)$ don't form a group, the current result does not follow as a corollary of Theorem 3.20 as stated. Instead, we must derive the current result as a corollary of its proof (actually, the proof of Theorem 2.48). Notice that the condition in Theorem 2.48 that all of the v_i come from a p-acceptable family $\mathcal{E}$ is stronger than necessary: all that is actually required is that the p-linear systems occuring at the various stages of the induction belong to $\mathcal{B}_{p,1}$ (the base case of weight vector $(1, 0, \ldots)$ requires this, as does the $p\text{-}\lim_m$ assertion in the final display of the proof).

Fix $p \in \mathcal{D}$. For the current proof, we start with the family $\{f_v, 2f_v\}$. After a single application of van der Corput, one gets to $\{D_x f_v, f_v, f_v{+}2D_x f_v\}$ (this is the derived family, called B_m in the proof of Theorem 2.48). The linear member of this family is $D_x f_v$, which for p-many x will be a member of $\mathcal{B}_{p,1}$, using just the fact that v is well spaced. Fixing x, a second application yields the family

$$\{f_v - D_x f_v, f_v + D_x f_v, f_v + D_y f_v - D_x f_v, f_v + D_y f_v + D_x f_v\}.$$

There are no linear members here, but we still must be careful about our choice of y. Note that y is chosen after x is, and can be taken from an arbitrary member of p; in particular, one may arrange that $\min \alpha(y)$ be large enough to guarantee that $D_y f_v - 2D_x f_v$ is well spaced (here is where we must utilize $(*)$). A third application of van der Corput leads to the family

$$\{2D_x f_v, D_y f_v, D_y f_v + 2D_x f_v, D_z f_v, D_z f_v + 2D_x f_v, D_z f_v + D_y f_v, D_z f_v + D_y f_v + 2D_x f_v\}.$$

All seven of these systems are linear. It is sufficient for the proof to go through that all of them, as well as all of their differences (which will occur as linear members of later families in the induction), be well spaced. Provided we require that $\min \alpha(z)$ be large enough to guaranteed that $D_z f_v - 2D_y f_v - 2D_x f_v$ is well spaced, this condition will be met. $\square$

3.57. Remark. We don't see any obstacle to generalizing the foregoing theorem to well spaced systems v of arbitrary degree satisfying appropriately formulated versions of $(*)$, and to more general configurations (certainly anything like $\mu\big(\prod_{i=1}^k T^{q_i(v(\alpha))}A\big) > 0$ should be okay, where the $q_i(x) \in \mathbf{Z}[x]$ have zero constant term). We haven't worked out the details.

A further analysis of the proof of Theorem 2.48 yields the following.

3.58. Corollary. Let $E \subset \mathbf{Z}$ with $d^*(E) > 0$, $p \in \mathcal{D}$, $A \in p$, let $\mathcal{E}$ be a p-acceptable family and let $g_1, \ldots, g_k \in \mathcal{E}$. If $h_1, \ldots, h_k$ are any p-VIP systems with $\deg_p h_i < \deg_p g_i$ and $\deg_p(h_i - h_j) < \deg_p(g_i - g_j)$, $1 \leq i \neq j \leq k$, then there exists $n \in A$ such that $d^*\big(E \cap (E + g_1(n) + h_1(n)) \cap \cdots \cap (E + g_k(n) + h_k(n))\big) > 0$.

Proof. Let $\mathcal{H}$ be the family of finite sets of p-VIP systems $\{v_1, \ldots, v_k\}$ such that $v_i = e_i + f_i$, $1 \leq i \leq k$, where $e_i \in \mathcal{E}$, f_i is a p-VIP system and $\deg_p f_i < \deg_p e_i$, $\deg_p(f_i - f_j) < \deg_p(e_i - e_j)$, $1 \leq i \neq j \leq k$. What we must show is that Theorem 2.48 holds for families $\{v_1, \ldots, v_k\} \in \mathcal{H}$.

Since it is obvious that any linear member v_i of such a family is in $\mathcal{E}$, and therefore in $\mathcal{B}_{p,1}$, all that is required in order for the proof to go through is that for any $\{v_1, \ldots, v_k\} \in \mathcal{H}$, for p-many m the family

$$B_m = \{v_2 - v_1, v_3 - v_1, \ldots, v_k - v_1, v_{w+1} + D_m v_{w+1} - v_1, \cdots, v_k + D_m v_k - v_1\}$$

again belongs to $\mathcal{H}$. We argue in steps:

(1) We have that, for $2 \leq i \leq k$, $v_i - v_1 = (e_i - e_1) + (f_i - f_1)$, and by hypothesis, $\deg_p(f_i - f_j) < \deg_p(e_i - e_j)$.

(2) For $w < i \leq k$,

$$v_i + D_m v_i - v_1 = (e_i + D_m e_i - e_1) + (f_i + D_m f_i - f_1).$$

Clearly $\deg_p(e_i+D_m e_i-e_1) > \deg_p(f_i+D_m f_i-f_1)$ if $e_i \not\sim_p e_1$, and if $e_i \sim_p e_1$, for p-many m one has $\deg_p(e_i+D_m e_i-e_1) = \deg_p e_i - 1 \geq \deg_p(e_i-e_1) > \deg_p(f_i-f_1)$. On the other hand, $\deg_p(e_i+D_m e_i-e_1) = \deg_p e_i - 1 > \deg_p f_i - 1 = \deg_p D_m f_i$. It follows that $\deg_p(e_i+D_m e_i-e_1) > \deg_p(f_i+D_m f_i-f_1)$.

(3) If $w < i \leq k$ and $2 \leq j \leq k$ then $(v_i+D_m v_i-v_1)-(v_j-v_1) = (v_i+D_m v_i-v_j)$; simply repeat argument (2) with 1 replaced by j.

(4) If $2 \leq i < j \leq k$ then $(v_i-v_1)-(v_j-v_1) = (v_i-v_j)$; repeat argument (1) with 1 replaced by j.

(5) If $w < j < i \leq k$ then

$$\begin{aligned}
&(v_i+D_m v_i-v_1)-(v_j+D_m v_j-v_1)\\
=&(v_i-v_j)+D_m(v_i-v_j)\\
=&\big((e_i-e_j)+D_m(e_i-e_j)\big)+\big((f_i-f_j)+D_m(f_i-f_j)\big).
\end{aligned}$$

But now clearly $\deg_p\big((e_i-e_j)+D_m(e_i-e_j)\big) = \deg_p(e_i-e_j) > \deg_p(f_i-f_j) = \deg_p\big((f_i-f_j)+D_m(f_i-f_j)\big)$. $\square$

Notice that in the foregoing theorem, one can take $\mathcal{E}$ to be the set of admissible generalized polynomials. This yields Theorem 1.27. We now move to the second of the aforementioned two examples.

3.59. Definition. An IP system $v(\alpha) = \sum_{n\in\alpha} d_n$, $\min\alpha \geq i_0 \in \mathbf{N}$, is *densely packed* if $d_n \in \mathbf{N}$, $n \geq i_0$, and $d_n = o\big(\sqrt{n/\log n}\big)$.

3.60. We are indebted to P. Balister for the condition $d_n = o\big(\sqrt{n/\log n}\big)$ appearing in the previous definition (in a prior draft, (d_n) was required to be bounded).

3.61. Proposition. Let $v(\alpha) = \sum_{n\in\alpha} d_n$ be a densely packed IP system. If $E \subset \mathbf{N}$ with $d^*(E) = \epsilon > 0$ then $d^*\big(f_v(E)\big) > 0$.

Proof. Let X_i be independent random variables with $P(X_i = 0) = \frac{1}{2} = P(X_i = d_i)$. For $N \in \mathbf{N}$, set $X = X^{(N)} = \sum_{i=1}^N X_i$. Our plan is to give our original proof, under the assumption (d_n) is bounded, then to state Balister's lemma, which allows the proof of the general case to go through in a similar fashion.

Choose large N and k such that $\frac{|E\cap[k2^N,(k+1)2^N)|}{2^N} > \frac{\epsilon}{10}$. The standard deviation of $X = X^{(N)}$ is at most $\frac{1}{2}M\sqrt{N}$ and its distribution is approximately normal, hence we may choose an interval I of length at most $\frac{100M\sqrt{N}}{\epsilon}$ such that $P(X \in I) > 1 - \frac{\epsilon}{20}$. From this is follows that $C = \{y \in [k2^N,(k+1)2^N) : f(y) \in I\}$ satifies $|C| \geq 2^N(1-\frac{\epsilon}{20})$, hence $|C \cap E| \geq 2^N\frac{\epsilon}{20}$.

Let $R = f_v(C \cap E)$. According to a theorem of Erdös ([Er]; see also Kleitman [K]), if d_i are non-zero integers, $1 \leq i \leq N$, and T is a positive integer then the number of distinct sets $\alpha \subset \{1,2,\ldots,N\}$ such that $\sum_{i\in\alpha} d_i = T$ is at most $\binom{N}{\lfloor\frac{N}{2}\rfloor}$. As a consequence of this theorem, each $x \in I$ has at most $\binom{N}{\lfloor\frac{N}{2}\rfloor}$ preimages under f; by Stirling's formula this is no more than, say, $\frac{100\cdot 2^N}{\sqrt{N}}$. It follows that $|R| \geq \frac{\epsilon\sqrt{N}}{2000}$. Since $I \approx \frac{100M\sqrt{N}}{\epsilon}$, this yields $\frac{|R|}{|I|} \geq \frac{\epsilon^2}{200000M}$. We may conclude, since N can be chosen as large as desired, that $d^*\big(f_v(E)\big) \geq \frac{\epsilon^2}{200000M}$.

For the general case, one substitutes the following lemma for the Erdös result, which changes things only slightly. For the assertion that for large N, $X^{(N)}$ is approximately normal, one needs, e.g., Liapounov's theorem. (Also note that Balister's lemma remains true when $d_n = 0$ for n less than some i_0, though the constant C appearing there will then depend on i_0.)

3.62. Lemma ([Ba]). Let X and $(X_n)_{n=1}^N$ be as above. There exist absolute constants $c, C > 0$ (independent of N) such that if $1 \leq d_n \leq c\sqrt{n/\log n}$, $1 \leq n \leq N$, then $P(X = k) \leq C(\sum_{n=1}^N d_n^2)^{-\frac{1}{2}}$ for all k. □

3.63. Definition. Let $v : \mathcal{F} \to \mathbf{Z}$ be a VIP system. If $\deg v = d > 1$, v is said to be *densely packed* if (i) $v(\alpha) = \sum_{\gamma\subset\alpha,|\gamma|=d} g(\gamma) + \sum_{\gamma\subset\alpha,\gamma\in\mathcal{F}_{d-1}} h(\gamma)$, where g has constant sign, and (ii) for every $\alpha \in \mathcal{F}_{d-1}$, the sequence $d_i = g(\alpha \cup \{i\})$, $i > \max\alpha$, generates a densely packed IP system. f_v is said to be densely packed if v is densely packed.

We leave the easy proof of the following to the reader.

3.64. Proposition. Let $v : \mathcal{F} \to \mathbf{N}$ be a densely packed VIP system of degree $d > 1$ and fix $\alpha \in \mathcal{F}$. Then $D_\alpha v$ is a densely packed VIP system.

3.65. Corollary. Let $p \in \mathcal{D}$ and let v be a densely packed VIP system. Then $f_v \in \mathcal{B}_p$.

3.66. Corollary. Let $v : \mathcal{F} \to \mathbf{Z}$ be a densely packed VIP system and let $\epsilon > 0$.

a. Suppose $(X, \mathcal{A}, \mu, T)$ is a measure preserving system. If $\mu(A) > 0$ then there is some $\alpha \in \mathcal{F}$ such that $\mu(A \cap T^{v(\alpha)}A) \geq \mu(A)^2 - \epsilon$.

b. If $E \subset \mathbf{Z}$ with $d^*(E) > 0$ then there is some $\alpha \in \mathcal{F}$ such that

$$d^*\Big(E \cap \big(E - v(\alpha)\big)\Big) > d^*(E)^2 - \epsilon.$$

As was the case with well spaced systems, we need a stronger assumption for multiple recurrence results along densely packed VIP systems. Again, we give only a special case but see no obstacle to a more general formulation.

3.67. Theorem. Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system. Let

$$v(\alpha) = \sum_{i<j,\{i,j\}\subset\alpha} d_{\{i,j\}} + \sum_{i\in\alpha} c_i$$

be a densely packed VIP system of degree 2. Suppose that for every $L, M > 0$, there exists $R = R(L, M) > 0$ such that for every $j > i \geq R$,

$$d_{\{i,j\}} > M\Big(\sum_{t=1}^{L} d_{\{t,j\}}\Big). \qquad (**)$$

If $A \in \mathcal{A}$ with $\mu(A) > 0$ then there is some $\alpha \in \mathcal{F}$ such that $\mu\big(A \cap T^{v(\alpha)} A \cap T^{2v(\alpha)} A\big) > 0$.

Proof. The proof is wholly analogous to that of Theorem 3.55, in light of the fact that Comment 3.56 generalizes to densely packed systems. In particular, $(**)$ implies that for any $\beta \in \mathcal{F}$ with $\min \beta \geq R$ and any $\alpha \in f$ with $\max \alpha \leq L$, the IP system $D_\beta v + M' D_\alpha v$, defined for ξ with $\min \xi > \max \beta$, is densely packed, where $|M'| \leq M$. More generally, if $\alpha_1 < \alpha_2 < \cdots < \alpha_{r-1}$, with $\max \alpha_{r-1} \leq L$, $D_\beta v - \sum_{i=1}^{r-1} M_i D_{\alpha_i} v$ is densely packed, where $|M_i| \leq M$. $\square$

Lastly, we state the condition (analogous to Theorem 1.18) promised in the introduction and after the statement of Theorem 3.9.

3.68. Theorem. Let p be an idempotent ultrafilter, let $u, v : \mathbf{N} \to \mathbf{Z}$ be p-VIP systems (in particular, u and v may be powers of n, or admissible generalized polynomials) having distinct, non-trivial rates of growth, in the sense that

$$\lim_{n\to\infty} |v(n)| = \lim_{n\to\infty} |u(n)| = \lim_{n\to\infty} \left|\frac{v(n)}{u(n)}\right| = \infty.$$

Put

$$F(n) = \sum_{i=1}^{n_1} [\![a_i u(n) [\![b_i v(n)]\!]]\!] + \sum_{i=1}^{n_2} [\![c_i v(n) [\![b_i u(n)]\!]]\!] + \sum_{i=1}^{n_3} [\![e_i u(n)]\!] [\![d_i v(n)]\!]$$
$$+ \sum_{i=1}^{n_4} [\![g_i u(n)]\!] + \sum_{i=1}^{n_5} [\![h_i v(n)]\!] + \sum_{i=1}^{n_6} [\![k_i u(n) v(n)]\!].$$

If $p\text{-}\lim_n |F(n)| < \infty$ then the following five conditions are satisfied:

$$\text{(1)} \qquad \sum_{i=1}^{n_1} a_i b_i + \sum_{i=1}^{n_2} c_i d_i + \sum_{i=1}^{n_3} e_i f_i + \sum_{i=1}^{n_6} k_i = 0;$$

$$\text{(2)} \qquad \sum_{i=1}^{n_5} h_i = 0;$$

$$\text{(3)} \qquad \sum_{i=1}^{n_4} g_i = 0;$$

$$\text{(4)} \qquad \sum_{i=1}^{n_2} c_i \otimes d_i + \sum_{i=1}^{n_3} f_i \otimes e_i \in \mathbf{R} \otimes \mathbf{Q}$$

$$\text{(5)} \qquad \sum_{i=1}^{n_1} a_i \otimes b_i + \sum_{i=1}^{n_3} e_i \otimes f_i \in \mathbf{R} \otimes \mathbf{Q}.$$

In particular, if any of (1)-(5) are not met, then $p\text{-}\lim_n \int f T^{F(n)} g \, d\mu = (\int f \, d\mu)(\int g \, d\mu)$ for any weakly mixing system (X, T) and $f, g \in L^\infty(X)$.

Proof. The last assertion follows from the first assertion and Theorem 3.9. For the proof of the body of the theorem, cf. [MQ, Theorem 3.2]. The present result can be obtained by essentially the same method. $\square$

References

[A] C. Adams, Largeness of the set of finite sums of sequences in **N**, Ph.D. dissertation, Howard University, 2006.

[Ba] P. Balister, personal communication.

[BBDF] M. Beiglböck, V. Bergelson, T. Downarowicz and A. Fish, D-sets and the central sets theorem. Manuscript.

[Be1] V. Bergelson, Weakly mixing PET, *Ergodic Theory and Dynamical Systems*, **7** (1987), 337-349.

[Be2] V. Bergelson, Ergodic Ramsey theory–an update, *Ergodic Theory of* Z^d*-actions* (editted by M. Pollicot and K. Schmidt), London Math. Soc. Lecture Note Series **228** (1996), 1-61.

[BD] V. Bergelson and T. Downarowicz, Large set of integers and hierarchy of mixing properties of measure preserving systems, *Colloq. Math.* To appear.

[BFM] V. Bergelson, H. Furstenberg, and R. McCutcheon, IP-sets and polynomial recurrence, *Ergodic Theory and Dynamical Systems* **16** (1996), 963-974.

[BHå] V. Bergelson and I. Håland, Sets of recurrence and generalized polynomials, *Convergence in ergodic theory and probability* (Columbus, OH, 1993), 91–110, Ohio State Univ. Math. Res. Inst. Publ., **5**, de Gruyter, Berlin, 1996.

[BH] V. Bergelson and N. Hindman, Nonmetrizable topological dynamics and Ramsey Theory, *Trans. Amer. Math. Soc.* **320** (1990), 293-320.

[BHMP] V. Bergelson, B. Host, R. McCutcheon and F. Parraeu, Aspects of uniformity in recurrence. Dedicated to the memory of Anzelm Iwanik. *Colloq. Math.* **84/85** (2000), 549-576.

[BK] V. Bergelson and I.J. Håland-Knutson, Higher order weak mixing along generalized polynomials. Manuscript.

[BKM] V. Bergelson, I.J. Håland-Knutson and R. McCutcheon, IP systems, generalized polynomials and recurrence. *Ergodic Theory and Dynamical Systems* **26** (2006), 999-1019.

[BL1] V. Bergelson and A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorem, *Journal of AMS* **9** (1996), 725-753.

[BL2] V. Bergelson and A. Leibman, Set-polynomials and polynomial extension of Hales-Jewett theorem, *Annals of Mathematics (2)* **150** (1999), 33-75.

[BL3] V. Bergelson and A. Leibman, Distribution of values of bounded generalized polynomials, *Acta. Math.* **198** (2007), 155-230.

[BM1] V. Bergelson and R. McCutcheon, Uniformity in the polynomial Szemerédi theorem, *Ergodic Theory of* $\mathbf{Z}^d$*-actions*, M. Pollicott and K. Schmidt, editors, London Mathematical Society Lecture Notes Series **228** (1996) 273-296.

[BM2] V. Bergelson and R. McCutcheon, An ergodic IP polynomial Szemeré di theorem, *Memoirs of AMS* **146** (2000), 1-106.

[BM3] V. Bergelson and R. McCutcheon, Central sets and a non-commutative Roth theorem, *American J. Math.* **129** (2007), 1251-1275.

[El] R. Ellis, *Lectures on topological dynamics*, Benjamin, New York, 1969.

[Er] P. Erdös, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898-902.

[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.

[FK1] H. Furstenber and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. d'Analyse Math.* **34** (1978), 275-291.

[FK2] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for IP-systems and combinatorial theory, *J. d'Analyse Math.* **45** (1985), 117-168.

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

[H] N. Hindman, Finite sums from sequences within cells of a partition of $\mathbf{N}$, *J. Combinatorial Theory* (Series A) **17** (1974), 1-11.

[HMS] N. Hindman, A. Maleki and D. Strauss, Central sets and their combinatorial characterization, *J. Combinatorial Theory* (Series A) **74** (1996), 188-208.

[HS] N. Hindman and D. Strauss, *Algebra in the Stone-Čech compactification: theory and applications*, de Gruyter, Berlin. 1998.

[K] D.J. Kleitman, On a lemma of Littlewood and Offord on the distribution of certain sums, *Math. Z.* **90** (1965), 251-259.

[Mc1] R. McCutcheon, An infinitary polynomial van der Waerden theorem *J. Combinatorial Theory* (Series A) **86** (1999), 214-231.

[Mc2] R. McCutcheon, FVIP systems and multiple recurrence, *Israel J. Math.* **146** (2005), 157-188.

[MQ] R. McCutcheon and A. Quas, Generalized polynomials and mild mixing systems, *Canadian J. Math.* To appear.

[Mi] K. Milliken, Ramsey's Theorem with sums or unions, *J. Combinatorial Theory* (Series A) **18** (1975), 276-290.

[Sá] A. Sárközy, On difference sets of integers III, *Acta. Math. Acad. Sci. Hungar.*, **31** (1978) 125-149.

[S] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta. Arith.* **27** (1975), 199-245.

[T] A. Taylor, A canonical partition relation for finite subsets of ω, *J. Combinatorial Theory* (Series A) **21** (1976), 137-146.

Vitaly Bergelson
Department of Mathematics
The Ohio State University
Columbus, OH 43210 USA
vitaly@math.ohio-state.edu

Randall McCutcheon
Dept. of Math. Sciences
University of Memphis
Memphis, TN 38152 USA
rmcctchn@memphis.edu