A NILPOTENT IP POLYNOMIAL MULTIPLE RECURRENCE THEOREM

PAVEL ZORIN-KRANICH

ABSTRACT. We generalize the IP-polynomial Szemerédi theorem due to Bergelson and McCutcheon and the nilpotent Szemerédi theorem due to Leibman. Important tools in our proof include a generalization of Leibman’s result that polynomial mappings into a nilpotent group form a group and a multiparameter version of the nilpotent Hales-Jewett theorem due to Bergelson and Leibman.

1. INTRODUCTION

Furstenberg’s ergodic theoretic proof [Fur77] of Szemerédi’s theorem on arithmetic progressions [Sze75] has led to various generalizations of the latter. Recall that Furstenberg’s original multiple recurrence theorem provides a syndetic set of return times. The IP recurrence theorem of Furstenberg and Katznelson [FK85], among other things, improves this to an IP* set. The idea to consider the limit behavior of a multicorrelation sequence not along a Følner sequence but along an IP-ring has proved to be very fruitful and allowed them to obtain the density Hales-Jewett theorem [FK91].

In a different direction, Bergelson and Leibman [BL96] have proved a polynomial multiple recurrence theorem. That result has been extended from commutative to nilpotent groups of transformations by Leibman [Lei98]. Many of the additional difficulties involved in the nilpotent extension were algebraic in nature and have led Leibman to develop a general theory of polynomial mappings into nilpotent groups [Lei02]. An important aspect of the proofs of these polynomial recurrence theorems, being present in all later extensions including the present article, is that the induction process involves “multiparameter” recurrence even if one is ultimately only interested in the “one-parameter” case.

More recently an effort has been undertaken to combine these two directions. Building on their earlier joint work with Furstenberg [BFM96], Bergelson and McCutcheon [BM00] have shown the set of return times in the polynomial multiple recurrence theorem is IP*. Joint extensions of their result and the IP recurrence theorem of Furstenberg and Katznelson have been obtained by Bergelson, Håland Knutson and McCutcheon for single recurrence [BHKM06] and McCutcheon for multiple recurrence [McC05]. The results of the last two papers also provide multiple recurrence along admissible generalized polynomials (Definition 4.24), a class that includes ordinary polynomials that vanish at the origin, and, more generally, along FVIP-systems (Definition 4.1), a generalization of IP-systems.

In this article we continue this line of investigation. Our main result, Theorem 5.32, generalizes McCutcheon’s IP polynomial multiple recurrence theorem to the nilpotent setting. Its content is best illustrated by the following generalization of Leibman’s nilpotent multiple recurrence theorem (here and throughout the article group actions on topological spaces and measure spaces are on the right and on function spaces on the left.).

Theorem 1.1. Let $T_1, \ldots, T_t$ be invertible measure-preserving transformations on a probability space $(X, \mathcal{A}, \mu)$ that generate a nilpotent group. Then for every $A \in \mathcal{A}$ with $\mu(A) > 0$, every
\( m \in \mathbb{N} \), and any admissible generalized polynomials \( p_{i,j} : \mathbb{Z}^m \to \mathbb{Z} \), \( i = 1, \ldots, t \), \( j = 1, \ldots, s \), the set
\[
\{ \vec{n} \in \mathbb{Z}^m : \mu \left( \bigcap_{j=1}^{s} A \left( \prod_{i=1}^{t} T_i^{p_{i,j}(\vec{n})} \right)^{-1} \right) > 0 \}
\]
is \( \text{FVIP}^* \) in \( \mathbb{Z}^m \), that is, it has nontrivial intersection with every \( \text{FVIP} \)-system in \( \mathbb{Z}^m \).

In particular, the set (1.2) is \( \text{IP}^* \), so that it is syndetic \([\text{Fur81}, \text{Lemma} \ 9.2]\).

**Contents**

1. Introduction 1
2. Polynomial mappings into nilpotent groups 2
   IP-polynomials 5
   IP-polynomials in several variables 5
   Polynomial-valued polynomials 6
   Monomial mappings 7
3. Topological multiple recurrence 8
   PET induction 8
   Nilpotent Hales-Jewett 9
   Multiparameter nilpotent Hales-Jewett 10
4. FVIP groups 12
   Hirsch length 12
   Partition theorems for IP-rings 13
   IP-limits 15
   Generalized polynomials and examples of FVIP groups 18
5. Measurable multiple recurrence 20
   Compact extensions 21
   Mixing and primitive extensions 22
   Existence of primitive extensions 23
   Almost periodic functions 23
   Multiple mixing 24
   Multiparameter multiple mixing 26
   Lifting multiple recurrence to a primitive extension 27
   Good groups of polynomial expressions 29
   References 31

**2. Polynomial mappings into nilpotent groups**

In this section we set up the algebraic framework for dealing with IP-polynomials in several variables with values in a nilpotent group.

We begin with a generalization of Leibman’s result that polynomial mappings into a nilpotent group form a group under pointwise operations \([\text{Lei02}, \text{Proposition} \ 3.7 \text{ and erratum}]\). Following an idea from the proof of that result by Green and Tao \([\text{GT12}, \text{Proposition} \ 6.5]\) we encode the information that is contained in Leibman’s vector degree in a prefiltration indexed by \( \mathbb{N} = \{ 0, 1, \ldots \} \) (see \([\text{GTZ12}, \text{Appendix B}]\) for related results regarding prefiltrations indexed by more general partially ordered semigroups).

A prefiltration \( G_* \) is a sequence of nested groups
\[
G_0 \geq G_1 \geq G_2 \geq \ldots \quad \text{such that} \quad [G_i, G_j] \subset G_{i+j} \quad \text{for every} \ i, j \in \mathbb{N}.
\]
A filtration (on a group $G$) is a prefiltration in which $G_0 = G_1$ (and $G_0 = G$). We will frequently write $G$ instead of $G_0$. Conversely, most groups $G$ that we consider in this article are endowed with a prefiltration $G_*$ such that $G_0 = G$. A group may admit several prefiltrations, and we usually fix one of them even if we do not refer to it explicitly.

A prefiltration is said to have length $d \in \mathbb{N}$ if $G_{t+1}$ is the trivial group and length $-\infty$ if $G_0$ is the trivial group. Arithmetic for lengths is defined in the same way as conventionally done for degrees of polynomials, i.e. $d - t = -\infty$ if $d < t$.

Let $G_*$ be a prefiltration of length $d$ and let $t \in \mathbb{N}$ be arbitrary. We denote by $G_{s+t}$ the prefiltration of length $d - t$ given by $(G_{s+t})_i = G_{i+t}$ and by $G_{s/t}$ the prefiltration of length $\min(d, t - 1)$ given by $G_{i/t} = G_i / G_t$ (this is understood to be the trivial group for $i \geq t$; note that $G_i$ is normal in each $G_t$ for $i \leq t$ by (2.1)). These two operations on prefiltrations can be combined: we denote by $G_{s/t+s}$ the prefiltration given by $G_{i/t+s} = G_{i+s}/G_t$, it can be obtained applying first the operation $/t$ and then the operation $+s$ (hence the notation).

If $G$ is a nilpotent group then the lower central series is a filtration. More generally, if $G_1 \leq G_0$ is a normal subgroup then $G_{i+1} = [G_i, G_1]$ defines a prefiltration (that has finite length if and only if $G_1$ is nilpotent), see for example [MKS66, Theorem 5.3]. If $G_*$ is a prefiltration and $\bar{d} = (d_i)_{i \in \mathbb{N}} \subset \mathbb{N}$ is a superadditive sequence (i.e. $d_{i+j} \geq d_i + d_j$ for all $i, j \in \mathbb{N}$; by convention $d_{-1} = -\infty$) then $G_{\bar{d}}$, defined by

\begin{equation}
G_{\bar{d}}^i = G_j \quad \text{whenever} \quad d_{j-1} < i \leq d_j,
\end{equation}

is again a prefiltration.

We define $G_*$-polynomial maps by induction on the length of the prefiltration.

**Definition 2.3.** Let $\Gamma$ be any set and $\mathcal{T}$ be a set of partially defined maps $T : \Gamma \ni \text{dom} (T) \rightarrow \Gamma$. Let $G_*$ be a prefiltration of length $d \in \{-\infty\} \cup \mathbb{N}$. A map $g : \Gamma \rightarrow G_0$ is called $G_*$-polynomial (with respect to $\mathcal{T}$) if either $d = -\infty$ (so that $g$ identically equals the identity) or for every $T \in \mathcal{T}$ there exists a $G_{*+1}$-polynomial map $D_T g$ such that

\begin{equation}
D_T g = g^{-1} T g := g^{-1} (g \circ T) \quad \text{on} \quad \text{dom} T.
\end{equation}

We write $P(\Gamma, G_*)$ for the set of $G_*$-polynomial maps, usually suppressing any reference to the set of maps $\mathcal{T}$ that will remain fixed for most of the article.

Informally, a map $g : \Gamma \rightarrow G_0$ is polynomial if every discrete derivative $D_T g$ is polynomial “of lower degree” (the “degree” of a $G_*$-polynomial map would be the length of the prefiltration $G_*$, but we prefer not to use this notion since it is necessary to keep track of the prefiltration $G_*$ anyway). The connection with Leibman’s notion of vector degree is provided by (2.2): a map has vector degree $\bar{d}$ with respect to a prefiltration $G_*$ if and only if it is $G_{\bar{d}}$-polynomial.

Note that if a map $g$ is $G_*$-polynomial then the map $g G_t$ is $G_{*+t}$-polynomial for any $t \in \mathbb{N}$ (but not conversely). We abuse the notation by saying that $g$ is $G_{*+t}$-polynomial if $g G_t$ is $G_{*+t}$-polynomial. In assertions that hold for all $T \in \mathcal{T}$ we omit the subscript in $D_T$.

The next theorem is the basic result about $G_*$-polynomials.

**Theorem 2.5.** For every prefiltration $G_*$ of length $d \in \{-\infty\} \cup \mathbb{N}$ the following holds.

1. Let $t_i \in \mathbb{N}$ and $g_i : \Gamma \rightarrow G$ be maps such that $g_i$ is $G_{*(d+1-t_i)_{-t_i}}$-polynomial for $i = 0, 1$. Then the commutator $[g_0, g_1]$ is $G_{*+t_0+t_1}$-polynomial.

2. Let $g_0, g_1 : \Gamma \rightarrow G$ be $G_*$-polynomial maps. Then the product $g_0 g_1$ is also $G_*$-polynomial.

3. Let $g : \Gamma \rightarrow G$ be a $G_*$-polynomial map. Then its pointwise inverse $g^{-1}$ is also $G_*$-polynomial.

**Proof.** We use induction on $d$. If $d = -\infty$ then the group $G_0$ is trivial and the conclusion hold trivially. Let $d \geq 0$ and assume that the conclusion holds for all smaller values of $d$. 
We prove part (1) using descending induction on $t = t_0 + t_1$. We clearly have $[g_0, g_1] \subset G_t$. If $t \geq d + 1$ there is nothing left to show. Otherwise it remains to show that $D[g_0, g_1]$ is $G_{t+1}$-polynomial. To this end we use the commutator identity

\begin{equation}
D[g_0, g_1] = [g_0, Dg_1] \cdot [[g_0, Dg_1], [g_0, g_1]] \\
\cdot [[[g_0, g_1], Dg_1], [[g_0, g_1], Dg_1], Dg_0] \cdot [Dg_0, g_1, Dg_1].
\end{equation}

We will show that the second to last term is $G_{t+1}$-polynomial, the argument for the other terms is similar. Note that $Dg_0$ is $G_{(d+1-t_0)+t_1}$-polynomial. By inner induction hypothesis it suffices to show that $[g_0, g_1, Dg_1]$ is $G_{(d-t_0)+t_1}$-polynomial. But the prefiltration $G_{(d-t_0)}$ has smaller length than $G_d$, and by the outer induction hypothesis we can conclude that $g_1 Dg_1$ is $G_{(d-t_0)+t_1}$-polynomial. Moreover, $g_0$ is clearly $G_{(d-t_0-t_1)}$-polynomial, and by the outer induction hypothesis its commutator with $g_1 Dg_1$ is $G_{(d-t_0)+t_1}$-polynomial as required.

Provided that each multiplicand in (2.6) is $G_{t+1}$-polynomial we can conclude that $D[g_0, g_1]$ is $G_{t+1}$-polynomial by the outer induction hypothesis.

Part (2) follows immediately by the Leibniz rule

\begin{equation}
D(g_0 g_1) = Dg_0 [Dg_0, g_1] Dg_1
\end{equation}

from (1) with $t_0 = 1$, $t_1 = 0$ and the induction hypothesis.

To prove part (3) notice that

\begin{equation}
D(g^{-1}) = g (Dg)^{-1} g^{-1} = [g^{-1}, Dg] (Dg)^{-1}.
\end{equation}

By the induction hypothesis the map $g^{-1}$ is $G_{d'}$-polynomial, the map $Dg$ is $G_{d+1}$-polynomial and the map $(Dg)^{-1}$ is $G_{d+1}$-polynomial. Thus also $D(g^{-1})$ is $G_{d+1}$-polynomial by (1) and the induction hypothesis. □

Discarding some technical information that was necessary for the inductive proof we can write the above theorem succinctly as follows.

**Corollary 2.9.** Let $G_d$ be a prefiltration of length $d$. Then the set $P(\Gamma, G_d)$ of $G_d$-polynomials on $\Gamma$ is a group under pointwise operations and admits a canonical prefiltration of length $d$ given by

\[ P(\Gamma, G_d) = P(\Gamma, G_{d+1}) \cdots \geq P(\Gamma, G_{d+d+1}). \]

Clearly, every subgroup $F \leq P(\Gamma, G_d)$ admits a canonical prefiltration $F_d$ given by

\begin{equation}
F_i := F \cap P(\Gamma, G_{d+i}).
\end{equation}

**Remark 2.11.** If $\Gamma$ is a group, then we recover [Lei02, Proposition 3.7] setting

\begin{equation}
\mathcal{T} = \{ T_b : n \mapsto nb, \text{dom}(T_b) = \Gamma, \text{where } b \in \Gamma \}.
\end{equation}

**Example 2.13.** If $\Gamma$ is a group and

\begin{equation}
\mathcal{T} = \{ T_{a,b} : \Gamma \rightarrow \Gamma, n \mapsto anb, \text{where } a, b \in \Gamma \},
\end{equation}

then every group homomorphism $g : \Gamma \rightarrow G_1$ is polynomial. In particular, every homomorphism to a nilpotent group is polynomial with respect to the lower central series.

This can be seen by induction on the length $d$ of the prefiltration $G_d$ as follows. If $d = -\infty$, then there is nothing to show. Otherwise write

\begin{equation}
D_{T_{a,b}} g(n) = g(n^{-1}) g(ab) = [g(n), g(a)^{-1}] g(ab).
\end{equation}

By the induction hypothesis $g G_d$ is $G_{d'}$-polynomial and the constant maps $g(a)^{-1}, g(ab)$ are $G_{d+1}$-polynomial since they take values in $G_1$. Hence $D_{T_{a,b}} g$ is $G_{d+1}$-polynomial by Theorem 2.5.
We will encounter further concrete examples of polynomials in Proposition 2.24 and Lemma 4.29.

**IP-polynomials.** In this article we are interested in the case $\Gamma = F_\emptyset$, where $F_\emptyset$ is the partial semigroup of finite subsets of $\mathbb{N}$ with the operation $\alpha \ast \beta = \alpha \cup \beta$ that is only defined if $\alpha$ and $\beta$ are disjoint. It is partially ordered by the relation $\alpha < \beta : \iff \max \alpha < \min \beta$.

Note that in particular $\emptyset < \alpha$ and $\alpha < \emptyset$ for any $\alpha \in F_\emptyset$.

The set $\mathcal{T}$ is then given by

$$\mathcal{T} = \{ T_\alpha : \beta \mapsto \alpha \ast \beta, \text{dom}(T_\alpha) = \{ \beta : \alpha \cap \beta = \emptyset \} \}, \text{ where } \alpha \in \Gamma.$$ 

If $T = T_\alpha$ then we also write $D_\alpha$ instead of $D_{T_\alpha}$. We write $\text{VIP}(G_\ast) \leq \mathcal{P}(F_\emptyset, G_\ast)$ for the subgroup of polynomials that vanish at $\emptyset$ and call its members VIP systems. For every $g \in \text{VIP}(G_\ast)$ and $\beta \in F_\emptyset$ we have

$$g(\beta) = g(0)D_\beta g(0) \in G_1.$$ 

Therefore the symmetric derivative $\tilde{D}$, defined by

$$\tilde{D}_\beta g(\alpha) := D_\beta g(\alpha)g(\beta)^{-1} = g(\alpha)^{-1}g(\alpha \cup \beta)g(\beta)^{-1},$$

maps $\text{VIP}(G_\ast)$ into $\text{VIP}(G_{\ast+1})$. Moreover, $\text{VIP}(G_\ast)$ admits the canonical prefiltration of length $d - 1$ given by

$$\text{VIP}(G_\ast) \geq \text{VIP}(G_{\ast+1}) \geq \cdots \geq \text{VIP}(G_{\ast+d}).$$

There is clearly no need to keep track of values of VIP systems at $\emptyset$, so we consider them as functions on $\mathcal{F} := F_\emptyset \setminus \{ \emptyset \}$.

The group $\text{VIP}(G_\ast)$ can be alternatively characterized by $\text{VIP}(G_\ast) = \{ 1_G \}$ for prefiltations $G_\ast$ of length $d = -\infty, 0$ and

$$g \in \text{VIP}(G_\ast) \iff g : \mathcal{F} \to G_1 \text{ and } \forall \beta \in \mathcal{F} \tilde{D}_\beta g \in \text{VIP}(G_{\beta+1}).$$

This characterization shows that if $G$ is an abelian group with the standard filtration $G_0 = G_1 = G$, $G_2 = \{ 1_G \}$, then $\text{VIP}(G_\ast)$ is just the set of IP systems in $G$.

**IP-polynomials in several variables.** As we have already mentioned in the introduction, the inductive procedure that has been so far utilized in all polynomial extensions of Szemerédi's theorem inherently relies on polynomials in several variables. We find it more convenient to define polynomials in $m$ variables not on $F^m$, but rather on the subset $F^m_\prec \subset F^m$ that consists of ordered tuples, that is,

$$F^m_\prec = \{ (\alpha_1, \ldots, \alpha_m) \in F^m : \alpha_1 < \cdots < \alpha_m \}.$$ 

Analogously, $F^m_\prec$ is the set of infinite increasing sequences in $\mathcal{F}$. We will frequently denote elements of $F^m_\prec$ or $F^m_\prec$ by $\bar{a} = (\alpha_1, \alpha_2, \ldots)$.

**Definition 2.19.** Let $G_\ast$ be a prefiltration and $F \leq \text{VIP}(G_\ast)$ a subgroup. We define the set $F^\otimes m$ of polynomial expressions in $m$ variables by induction on $m$ as follows. We set $F^{\otimes 0} = \{ 1_G \}$ and we let $F^{\otimes m+1}$ be the set of functions $g : F^{m+1}_{\prec} \to G_0$ such that

$$g(\alpha_1, \ldots, \alpha_{m+1}) = W^{\alpha_1, \ldots, \alpha_m}(\alpha_{m+1})S(\alpha_1, \ldots, \alpha_m),$$

where $S \in F^{\otimes m}$ and $W^{\alpha_1, \ldots, \alpha_m} \in F$ for every $\alpha_1 < \cdots < \alpha_m$.

Note that $F^{\otimes 1} = F$.

---

1A partial semigroup [BBH94] is a set $\Gamma$ together with a partially defined operation $\ast : \Gamma \times \Gamma \to \Gamma$ that is associative in the sense that $(a \ast b) \ast c = a \ast (b \ast c)$ whenever both sides are defined.
Lemma 2.20. Suppose that $F$ is invariant under conjugation by constant functions. Then, for every $m$, the set $F^\otimes m$ is a group under pointwise operations and admits a canonical prefiltration given by $(F^\otimes m)_i = (F_i)^{\otimes m}$.

If $K \leq F$ is a subgroup that is invariant under conjugation by constant functions then $K^{\otimes m} \leq F^\otimes m$ is also a subgroup.

Proof. We use induction on $m$. For $m = 0$ there is nothing to show. Let
\[ R_j \in (F_j)^{\otimes m+1} : (\alpha_1, \ldots, \alpha_m) \mapsto W_j^{\alpha_1 \cdots \alpha_m}(\alpha_{m+1})S_j(\alpha_1, \ldots, \alpha_m), \quad j = 0, 1 \]
be polynomial expressions in $m + 1$ variables. Suppressing the variables $\alpha_1, \ldots, \alpha_m$ we have
\[ R_0R_1^{-1}(\alpha_{m+1}) = W_0(\alpha_{m+1})\left( S_0S_1^{-1}W_1^{-1}S_1^{-1}\right)(\alpha_{m+1})S_0S_1^{-1}, \]
so that $R_0R_1^{-1} \in F^\otimes m$. Hence $F^\otimes m$ is a group.

In order to show that $(F_j)^{\otimes m+1}$ is indeed a prefiltration we have to verify that
\[ [R_0, R_1] = [W_0S_0, W_1S_1] \in (F_{t_0+t_1})^{\otimes m+1}. \]
This follows from the identity
\[ [xy, uv] = [x, u][x, v][[x, v], [x, u]][[x, u], v] \cdot [[x, v], [x, u]][[x, u], v], \]
\[ [y, v][y, u][[y, u], v]. \]
It is clear that $K^{\otimes m} \leq F^\otimes m$ is a subgroup provided that both sets are groups. \(\square\)

For every $m \in \mathbb{N}$ there is a canonical embedding $F^\otimes m \leq F^\otimes m+1$ that forgets the last variable. Thus we can talk about
\[ F^\otimes \omega := \operatorname{inj lim}_{m \in \mathbb{N}} F^\otimes m = \bigcup_{m \in \mathbb{N}} F^\otimes m, \]
this is a group of maps defined on $\mathcal{F}_\omega$ with prefiltration $(F^\otimes \omega)_i = (F_i)^{\otimes \omega}$.

Polynomial-valued polynomials.

Definition 2.21. Let $G_\bullet$ be a filtration of length $d$. A VIP group is a subgroup $F \leq \operatorname{VIP}(G_\bullet)$ that is closed under conjugation by constant functions and under $\check{D}$ in the sense that for every $g \in F$ and $\alpha \in \mathcal{F}$ the symmetric derivative $\check{D}_\alpha g$ lies in $F_1$ (defined in (2.10)).

In particular, the group $\operatorname{VIP}(G_\bullet)$ itself is VIP.

Proposition 2.22. Let $F \leq \operatorname{VIP}(G_\bullet)$ be a VIP group. Then for every $g \in F^\otimes m$ the substitution map
\[ h : \tilde{\beta} = (\beta_1, \ldots, \beta_m) \in \mathcal{F}_\omega \mapsto (g[\tilde{\beta}] : \tilde{\alpha} \in \mathcal{F}_\omega \mapsto g(\cup_{i \leq \beta_1} \alpha_i, \ldots, \cup_{i \leq \beta_m} \alpha_i)) \]
lies in $\operatorname{VIP}(F^\otimes \omega)^{\otimes m}$.

Proof. We proceed by induction on $m$. In case $m = 0$ there is nothing to show, so suppose that the assertion is known for $m$ and consider $g \in F^\otimes m$. By definition we have
\[ g(\alpha_1, \ldots, \alpha_{m+1}) = W^{\alpha_1 \cdots \alpha_m}(\alpha_{m+1})S(\alpha_1, \ldots, \alpha_m) \]
and
\[ h(\beta_1, \ldots, \beta_{m+1})(\tilde{\alpha}) = W^{\cup_{i \leq \beta_1} \alpha_i, \ldots, \cup_{i \leq \beta_m} \alpha_i}(\beta_{m+1})(\tilde{\alpha})S(\beta_1, \ldots, \beta_m)(\tilde{\alpha}). \]
In view of the induction hypothesis it remains to verify that the map
\[ \check{h} : \beta \mapsto (\tilde{\alpha} \mapsto W^{\cup_{i \leq \beta_1} \alpha_i, \ldots, \cup_{i \leq \beta_m} \alpha_i}(\beta)(\tilde{\alpha})), \quad \beta > \beta_m > \cdots > \beta_1, \]
is in VIP($F^{\otimes \omega}$). The fact that $\tilde{h}(\beta) \in F^{\otimes \omega}$ for all $\beta$ follows by induction on $|\beta|$ using the identity

$$W_{\cup \in \mathcal{F}_1} a_{i_1 \cdots i_n} a_i [\beta \cup \{b\}] (\tilde{a}) = W_{\cup \in \mathcal{F}_1} a_{i_1 \cdots i_n} a_i (\tilde{a}_l) D_{\cup \in \mathcal{F}_1} a_l W_{\cup \in \mathcal{F}_1} a_{i_1 \cdots i_n} a_i (a_b) W_{\cup \in \mathcal{F}_1} a_{i_1 \cdots i_n} a_i [\beta] (\tilde{a})$$

that holds whenever $b > \beta > \beta_m > \cdots > \beta_1$. In order to see that $\tilde{h}$ is polynomial in $\beta$ observe that

$$D_{\gamma} \tilde{h}(\beta) : \tilde{a} \mapsto D_{\cup \in \mathcal{F}_1} a_i W_{\cup \in \mathcal{F}_1} a_{i_1 \cdots i_n} a_i (\cup \in \mathcal{F}_1) \alpha_i, \ \beta > \gamma > \beta_m > \cdots > \beta_1. \ \square$$

**Monomial mappings.** In this section we verify that monomial mappings into nilpotent groups in the sense of Bergelson and Leibman [BLO3, §1.3] are polynomial in the sense of Definition 2.3.

For a sequence of finite sets $R = (R_0, R_1, \ldots)$ only finitely many of which are non-empty and a set $\alpha$ write

$$R[\alpha] := \alpha^0 \times R_0 \cup \alpha^1 \times R_1 \cup \ldots$$

Here the symbol $\cup$ denotes disjoint union and $\alpha^i$ are powers of the set $\alpha$ (note that $\alpha^0$ consists of one element, the empty tuple).

**Proposition 2.24.** Let $G_\ast$ be a prefiltration of length $d$ and $N \subset \mathbb{N}$ any subset. Let $g_\ast : R[N] \to G_\ast, \ x \mapsto g_x$ be a mapping such that $g_\ast (N^i \times R_j) \subset G_i$ for every $i \in \mathbb{N}$ and $\prec$ be any linear ordering on $R[N]$. Then the map

$$g : \mathcal{F}(N) \to G_\ast, \ \alpha \mapsto \prod_{j \in R[\alpha]} g_j$$

is $G_\ast$-polynomial on the partial semigroup $\mathcal{F}(N)$ (here the symbol $\prec$ on top of $\prod$ indicates the order of factors in the product).

**Proof.** We induct on the length of the prefiltration $G_\ast$. If $d = -\infty$ then there is nothing to prove. Otherwise let $\beta \in \mathcal{F}(N)$. We have to show that $D_\beta g$ is $G_{\ast + 1}$-polynomial.

Let $B \subset R[N]$ be a finite set and $A \subset B$. By induction on the length of an initial segment of $A$ (that proceeds by pulling the terms $g_j, j \in A$, out of the double product one by one, leaving commutators behind) we see that

$$\prod_{j \in B}^\prec g_j = \prod_{j \in A} g_j \prod_{j \in B \setminus A} \prod_{k \in A^\leq d}^\prec \prod_{k \in A^\leq d}^\prec g_{j,k}, \quad (2.25)$$

where $A^\leq d$ is the set of all tuples of elements of $A$ with at most $d$ coordinates in $N$ and

$$g_{j,0} = g_j, \quad g_{j,(k_0, \ldots, k_i)} = \begin{cases} [g_{j,(k_0, \ldots, k_{i-1})}, g_{k_i}] & \text{if } j \prec k_0 \prec \cdots \prec k_i, \\ 1 & \text{otherwise.} \end{cases}$$

Let $\alpha \in \mathcal{F}(N)$ be disjoint from $\beta$. Applying (2.25) with $A := R[\alpha]$ and $B := R[\alpha \cup \beta]$ we obtain

$$D_\beta g(\alpha) = \prod_{j \in R[\alpha \cup \beta] \setminus R[\alpha]} \prod_{k \in R[\alpha] \setminus \alpha^d}^\prec \prod_{j \in R[\alpha \cup \beta] \setminus R[\alpha]}^\prec g_{j,k},$$

where $g_{j,(k_0, \ldots, k_i)} \in G_{d_1 + d_2 + \cdots + d_i}$ provided that $j \in \alpha^i \times R_j$ and $k_0 \in \alpha^{d_0} \times R_{d_0}, \ldots, k_i \in \alpha^{d_i} \times R_{d_i}$.

The double product can be rewritten as $\prod_{j \in S[\alpha]} h_j$ for some sequence of finite sets $S$, an ordering $\prec'$ on $S[N']$, where $N' = N \setminus \beta$, and $h_\ast : S[N'] \to G$. The sequence of sets $S$ is obtained by the requirement

$$(R[\alpha \cup \beta] \setminus R[\alpha]) \times R[\alpha] \leq d = S[\alpha]$$
for every $\alpha \subset N'$. The lexicographic ordering on $(R[N] \setminus R[N']) \times R[N']^{\leq d}$ induces an ordering $\prec'$ on $S[N']$. Define $h_z = g_{j,k}$ if $(j,k)$ corresponds to $z \in S[N']$.

By construction we have $h_z((N')^l \times S_i) \subset G_{i+1}$ since each element of $R[N] \setminus R[N']$ has at least one coordinate in $N$ but not $N'$. Thus $D^*_\beta \mu$ is $G_{i+1}$-polynomial by the induction hypothesis.

**Corollary 2.26.** Let $G$ be a nilpotent group with lower central series

$$G = G_0 = G_1 \geq \cdots \geq G_i \geq G_{i+1} = \{1_G\},$$

let $g_* : N^d \rightarrow G$ be an arbitrary mapping and $\prec$ be any linear ordering on $N^d$. Then the map

$$g : \mathcal{F}(N) \rightarrow G, \quad \alpha \mapsto \prod_{j \in \alpha} g_j$$

is polynomial on the partial semigroup $\mathcal{F}(N)$ with respect to the filtration

$$(2.27) \quad G_0 \geq G_1 \geq \cdots \geq G_i \geq \cdots \geq G_s \geq G_{s+1}.$$ 

3. **Topological multiple recurrence**

In this section we refine the nilpotent Hales-Jewett theorem due to Bergelson and Leibman [BL03, Theorem 0.19] using the induction scheme from [BL99, Theorem 3.4]. This allows us to deduce a *multiparameter* nilpotent Hales-Jewett theorem that will be ultimately applied to polynomial-valued polynomials mappings.

**PET induction.** First we describe the PET (polynomial exhaustion technique) induction scheme [Ber87]. For a polynomial $g \in \text{VIP}(G_*)$ define its level $l(g)$ as the greatest integer $l$ such that $g \in \text{VIP}(G_{*+l})$. We define an equivalence relation on the set of non-zero $G_*$-polynomials by $g \sim h$ if and only if $l(g) = l(h) < l(g^{-1}h)$. Transitivity and symmetry of $\sim$ follow from Theorem 2.5.

**Definition 3.1.** A system is a finite subset $A \subset \text{VIP}(G_*)$. The weight vector of a system $A$ is the function

$$l \mapsto \text{the number of equivalence classes modulo } \sim \text{ of level } l \text{ in } A.$$ 

The lexicographic ordering is a well-ordering on the set of weight vectors and the PET induction is induction with respect to this ordering.

**Proposition 3.2.** Let $A$ be a system, $h \in A$ be a mapping of maximal level and $B \subset G_s$, $M \subset \mathcal{F}$ be finite sets. Then the weight vector of the system

$$A'' = \{bg^{-1}gD_{a^b}^gb^{-1} \mid g \in A, \alpha \in M, b \in B\} \setminus \{1_G\}$$

precedes the weight vector of $A$.

**Proof.** We claim first that the weight vector of the system

$$A' = \{h^{-1}gD_{a^h}^g \mid \alpha \in M, g \in A\} \setminus \{1_G\}$$

precedes the weight vector of $A$. Indeed, if $l(g) < l(h)$, then $g \sim h^{-1}gD_{a^h}^g$. If $l(g) = l(g) = l(h)$ and $g \sim h \sim h^{-1}gD_{a^h}^g$. Finally, if $g \sim h$, then $l(h^{-1}gD_{a^h}^g) > l(h)$. Thus the weight vector of $A'$ does not differ from the weight of vector of $A$ before the $l(h)$-th position and is strictly smaller at the $l(h)$-th position, as required.

We now claim that the weight vector of the system

$$A'' = \{bg^b, g \in A', b \in B\}$$

precedes the weight vector of $A$. Indeed, if $l(g) < l(h)$, then $g \sim h^{-1}gD_{a^h}^g$. If $l(g) = l(g) = l(h)$ and $g \sim h \sim h^{-1}gD_{a^h}^g$. Finally, if $g \sim h$, then $l(h^{-1}gD_{a^h}^g) > l(h)$. Thus the weight vector of $A''$ does not differ from the weight of vector of $A$ before the $l(h)$-th position and is strictly smaller at the $l(h)$-th position, as required.
Nilpotent Hales-Jewett. The following refined version of the nilpotent IP polynomial topological multiple recurrence theorem due to Bergelson and Leibman [BL03, Theorem 0.19] does not only guarantee the existence of a “recurrent” point, but also allows one to choose it from a finite subset $S x$ of any given orbit.

Theorem 3.3 (Nilpotent Hales-Jewett). Assume that $G$ acts on the right on a compact metric space $(X, \rho)$ by homeomorphisms. For every system $A$, every $\epsilon > 0$ and every $H \in \mathcal{F}$ there exists $N \in \mathcal{F}$, $N > H$, and a finite set $S \subset G$ such that for every $x \in X$ there exist a non-empty $\alpha \subset N$ and $s \in S$ such that $\rho(x s g(\alpha), xs) < \epsilon$ for every $g \in A$.

Here we follow Bergelson and Leibman and use “Hales-Jewett” as a shorthand for “IP polynomial multiple recurrence”. Indeed, this follows directly from

$$bg b^{-1} = g [g, b^{-1}] \sim g.$$

Proof. We use PET induction on the weight vector $w(A)$. If $w(A)$ vanishes identically then $A$ is the empty system and there is nothing to show. Assume that the conclusion is known for every system whose weight vector precedes $w(A)$. Let $h \in A$ be an element of maximal level, without loss of generality we may assume $h \neq 1_0$. Let $k$ be such that every $k$-tuple of elements of $X$ contains a pair of elements at distance $< \epsilon/2$.

We define finite sets $H_i \in \mathcal{F}$, finite sets $B_i, \tilde{B}_i \subset G$, systems $A_i$ whose weight vector precedes $w(A)$, positive numbers $\epsilon_i$, and finite sets $N_i \in \mathcal{F}$ by induction on $i$ as follows. Begin with $H_0 := H$ and $B_0 = \tilde{B}_0 = \{1_0\}$. The weight vector $w(A_i)$ of the system

$$A_i := \{bh^{-1}g \tilde{D}_m g b^{-1}, \quad g \in A, m \in N_0 \cup \cdots \cup N_{i-1}, b \in B_i\}$$

precedes $w(A)$ by Proposition 3.2. By uniform continuity we can choose $\epsilon_i$ such that

$$\rho(x, y) < \epsilon_i \implies \forall b \in B_i \quad \rho(x b, y b) < \frac{\epsilon}{2k}.$$  

By the induction hypothesis there exists $N_i \in \mathcal{F}$, $N_i > H_i$, and a finite set $S_i \subset G$ such that

$$\forall x \in X \quad \exists n_i \subset N_i, s_i \in S_i \quad \forall g \in A_i \quad \rho(xs_i g(n_i), xs_i) < \epsilon_i.$$

Finally, let $H_{i+1} := H_i \cup N_i$ and

$$B_{i+1} := \{s_i bh(\alpha_i)^{-1}, \quad \alpha_i \subset N_i, s_i \in S_i, b \in B_i\} \subset G,$$

$$\tilde{B}_{i+1} := \{bg(m), \quad g \in A, m \in N_0 \cup \cdots \cup N_i, b \in B_{i+1}\} \subset G.$$

This completes the inductive definition. Now fix $x \in X$. We define a sequence of points $y_i$ by descending induction on $i$. Begin with $y_0 := x$. Assume that $y_i$ has been chosen and choose $n_i \subset N_i$ and $s_i \in S_i$ as in (3.4), then set $y_{i-1} := y_i s_i$.

Finally, let $x_0 := y_0 s_{0} h(n_0)^{-1}$ and $x_{i+1} := x_i h(n_{i+1})^{-1}$. We claim that for every $g \in A$ and any $0 \leq i \leq j \leq k$ we have

$$\rho(x_j g(n_{i+1} \cup \cdots \cup n_j), x_i) < \frac{\epsilon}{2k}(j - i).$$
This can be seen by ascending induction on \( j \). Let \( i \) be fixed, the claim is trivially true for \( j = i \). Assume that the claim holds for \( j - 1 \) and let \( g \) be given. Consider

\[
 b := s_{j-1} \ldots s_0 h(n_0)^{-1} \ldots h(n_{j-1})^{-1} \in B_j \quad \text{and} \quad \bar{b} := bg(n_{i+1} \cup \cdots \cup n_{j-1}) \in \bar{B}_j.
\]

By choice of \( n_j \) and \( s_j \) we have

\[
 \rho(y_j s_j bh(n_j)^{-1} g(n_{i+1} \cup \cdots \cup n_j) g(n_{i+1} \cup \cdots \cup n_{j-1})^{-1} b^{-1}, y_j s_j) < \epsilon_j.
\]

By definition of \( \epsilon_j \) this implies

\[
 \rho(y_j s_j bh(n_j)^{-1} g(n_{i+1} \cup \cdots \cup n_j) g(n_{i+1} \cup \cdots \cup n_{j-1})^{-1} b^{-1}, y_j s_j \bar{b}) < \frac{\epsilon}{2k}.
\]

Plugging in the definitions we obtain

\[
 \rho(x_j g(n_{i+1} \cup \cdots \cup n_j), x_{j-1} g(n_{i+1} \cup \cdots \cup n_{j-1})) < \frac{\epsilon}{2k}.
\]

The induction hypothesis then yields

\[
 \rho(x_j g(n_{i+1} \cup \cdots \cup n_j), x_i)
 \leq \rho(x_j g(n_{i+1} \cup \cdots \cup n_j), x_{j-1} g(n_{i+1} \cup \cdots \cup n_{j-1})) + \rho(x_{j-1} g(n_{i+1} \cup \cdots \cup n_{j-1}), x_i)
 \leq \frac{\epsilon}{2k} + \frac{\epsilon}{2k} (j - 1 - i) = \frac{\epsilon}{2k} (j - i)
\]

as required.

Recall now that by definition of \( k \) there exist \( 0 < i < j \) such that \( \rho(x_i, x_j) < \frac{\epsilon}{2} \). By \( (3.5) \) we have

\[
 \rho(x_j g(n_{i+1} \cup \cdots \cup n_j), x_i) \leq \rho(x_j g(n_{i+1} \cup \cdots \cup n_j), x_i) + \rho(x_i, x_j) < \epsilon
\]

for every \( g \in A \). But \( x_j = xs \) for some

\[
 s \in S := S_k \ldots S_0 h(\mathcal{F}(N_0))^{-1} \ldots h(\mathcal{F}(N_k))^{-1},
\]

and we obtain the conclusion with \( N = N_0 \cup \cdots \cup N_k \) and \( S \) as above. \( \square \)

We remark that [BL99, Theorem 3.4] provides a slightly different set \( S \) that can be recovered substituting \( y_k := xh(N_k) \) and \( y_{i-1} := y_i s h(N_{i-1}) \) in the above proof and making the corresponding adjustments to the choices of \( B_i, b \) and \( S \).

**Multiparameter nilpotent Hales-Jewett.** We will now prove a version of the nilpotent Hales-Jewett theorem in which the polynomial configurations may depend on multiple parameters \( \alpha_1, \ldots, \alpha_m \).

**Theorem 3.6** (Multiparameter nilpotent Hales-Jewett). Assume that \( G \) acts on the right on a compact metric space \( (X, \rho) \) by homeomorphisms and let \( m \in \mathbb{N} \). For every finite set \( A \subset \text{VIP}(G_*)^m \), every \( \epsilon > 0 \) and every \( H \in \mathcal{F} \) there exists a finite set \( N \subset \mathcal{F}, N > H \), and a finite set \( S \subset G \) such that for every \( x \in X \) there exists \( s \in S \) and non-empty subsets \( \alpha_1 < \cdots < \alpha_m \subset N \) such that \( \rho(xs g(\alpha_1, \ldots, \alpha_m), xs) < \epsilon \) for every \( g \in A \).

**Proof.** We use induction on \( m \). The base case \( m = 0 \) is trivial. Assume that the conclusion is known for some \( m \), we prove it for \( m + 1 \).

Let \( A \subset \text{VIP}(G_*)^{m+1} \) and \( H \) be given. For convenience we write \( \bar{\alpha} = (\alpha_1, \ldots, \alpha_m) \) and \( \alpha = \alpha_{m+1} \). By definition each \( g \in A \) can be written in the form

\[
 g(\alpha_1, \ldots, \alpha_{m+1}) = g_{2}(\bar{\alpha}) g_1(\bar{\alpha})
\]

with \( g_1 \in \text{VIP}(G_*)^m \) and \( g_{2}(\bar{\alpha}) \in \text{VIP}(G_*) \).
We apply the induction hypothesis with the system \( \{g_1, g \in A\} \) and \( \epsilon / 2 \), thereby obtaining a finite set \( N \in \mathcal{F} \), \( N > H \), and a finite set \( S \subset G \). We write “\( \overline{a} \subset N \)” instead of “\( \alpha_1 < \cdots < \alpha_m < N \)”.

By uniform continuity there exists \( \epsilon' \) such that
\[
\rho(x, y) < \epsilon' \implies \forall s \in S, \overline{a} \subset N, g \in A \quad \rho(x s g_1(\overline{a}), y s g_1(\overline{a})) < \epsilon / 2.
\]

We invoke Theorem 3.3 with the system \( \{s g s^{-1}, s \in S, \overline{a} \subset N, g \in A\} \) and \( \epsilon' \), this gives us a finite set \( N' \subset \mathcal{F}, N' > N \), and a finite set \( S' \subset G \) with the following property: for every \( x \in X \) there exist \( s' \in S' \) and \( \alpha \subset N' \) such that
\[
\forall s \in S, \overline{a} \subset N, g \in A \quad \rho(x s' \overline{a} g_1(\overline{a}), x s' g_1(\overline{a})) < \epsilon'.
\]

By choice of \( \epsilon' \) this implies
\[
\forall s \in S, \overline{a} \subset N, g \in A \quad \rho(x s' \overline{a} g_1(\overline{a}), x s' g_1(\overline{a})) < \epsilon / 2.
\]

By choice of \( N \) and \( S \), considering the point \( x s' \), we can find \( \overline{a} \subset N \) and \( s \in S \) such that
\[
\forall g \in A \quad \rho(x s' \overline{a} g_1(\overline{a}), x s' g_1(\overline{a})) < \epsilon / 2.
\]

Combining the last two inequalities we obtain
\[
\forall g \in A \quad \rho(x s' g(\overline{a}), x s') < \epsilon.
\]

This yields the conclusion with finite sets \( N \cup N' \) and \( S' \).

The combinatorial version is derived using the product space construction of Furstenberg and Weiss [FW78].

Corollary 3.7. Let \( G_\bullet \) be a filtration on a countable nilpotent group \( G \), \( m \in \mathbb{N} \), \( A \subset \text{VIP}(G_\bullet) \otimes m \) a finite set, and \( l \in \mathbb{N} \). Then there exists \( N \in \mathbb{N} \) and finite sets \( S, T \subset G \) such that for every \( l \)-coloring of \( T \) there exist \( \alpha_1 < \cdots < \alpha_m \subset N \) and \( s \in S \) such that the set \( \{s g(\overline{a}), g \in A\} \) is monochromatic (and in particular contained in \( T \)).

Proof. Let \( X := \text{I}^G \) be the compact metrizable space of all \( l \)-colorings of \( G \) with the right \( G \)-action \( x g(h) = x(gh) \). We apply Theorem 3.6 to this space, the system \( A \), the set \( H = \emptyset \), and an \( \epsilon > 0 \) that is sufficiently small to ensure that \( \rho(x, x') < \epsilon \) implies \( x(e_G) = x'(e_G) \).

This yields certain \( N \in \mathbb{N} \) and \( S \subset G \) that enjoy the following property: for every coloring \( x \in X \) there exist \( \alpha_1 < \cdots < \alpha_m \subset N \) and \( s \in S \) such that \( \{s g(\overline{a}), g \in A\} \) is monochromatic. Observe that this property only involves a finite subset \( T = \bigcup_{g \in A} S g(N_G)^n \subset G \).

In the proof of our measurable recurrence result we will apply this combinatorial result to polynomial-valued polynomial mappings. We encode all the required information in the next corollary.

Corollary 3.8. Let \( m \in \mathbb{N} \), \( K \leq F \leq \text{VIP}(G_\bullet) \) be VIP groups, and \( FE \leq \text{VIP}(G_\bullet) \otimes \omega \) be a countable subgroup that is closed under substitutions \( g \mapsto g[\overline{\beta}] \) (recall (2.23)).

Then for any finite subsets \( (R_i)_{i=0}^w \subset K^{\otimes m} \cap FE \) and \( (W_i)_{i=0}^{w-1} \subset F^{\otimes m} \cap FE \) there exist \( N, \omega \in \mathbb{N} \) and \( (L_i, M_i)_{i=1}^w \subset (K^{\otimes N} \cap FE) \times (F^{\otimes \omega} \cap FE) \) such that for every \( l \)-coloring of the latter set there exists an index \( a \) and sets \( \beta_1 < \cdots < \beta_m \subset N \) such that the set \( \{L_a R_i[\overline{\beta}], M_i W_k[\overline{\beta}] L_{a-i}^{-1}, i, k \} \) is monochromatic (and in particular contained in the set \( (L_i, M_i)_{i=1}^w \)). We may assume \( L_1 = 1_G \).

Proof. By Proposition 2.22 the maps \( \overline{\beta} \mapsto (R_i[\overline{\beta}], W_k[\overline{\beta}] R_i[\overline{\beta}]) \) are polynomial expressions with values in \( K^{\otimes \omega} \times F^{\otimes \omega} \). By the assumption they also take values in \( FE \times FE \). Given an \( l \)-coloring \( \chi \) of \( (K^{\otimes \omega} \cap FE) \times (F^{\otimes \omega} \cap FE) \) we pass to the \( l \)-coloring \( \tilde{\chi}(g, h) = \chi(g, hg^{-1}) \). Corollary 3.7 then provides the desired \( N \) and \( (L_i, M_i)_{i=1}^w = T \cup S \).
In this section we consider a class of IP-polynomials that IP-converge to orthogonal projections.

**Definition 4.1.** An FVIP group is a finitely generated VIP group. An FVIP system is a member of some FVIP group.

The main result about FVIP groups is the following nilpotent version of [BFM96, Theorem 1.8] and [BHLM06, Theorem 1.9] that will be used to construct “primitive extensions” (we will recall the definitions of a primitive extension and an IP-limit in due time).

**Theorem 4.2.** Let $G_{\bullet}$ be a prefiltration of finite length and $F \leq \text{VIP}(G_{\bullet})$ be an FVIP group. Suppose that $G_0$ acts on a Hilbert space $H$ by unitary operators and that for each $(g_a)_a \in F$ the weak limit $P_g = \text{w-IP-lim}_{a \in F} g_a$ exists. Then

1. each $P_g$ is an orthogonal projection and
2. these projections commute pairwise.

The finite generation assumption cannot be omitted in view of a counterexample in [BHLM06]. We begin with some algebraic preliminaries.

**Hirsch length.** We use Hirsch length of a group as a substitute for the concept of the rank of a free $\mathbb{Z}$-module. Recall that a subnormal series in a group is called polycyclic if the quotients of consecutive subgroups in this series are cyclic and a group is called polycyclic if it admits a polycyclic series.

**Definition 4.3.** The Hirsch length $h(G)$ of a polycyclic group $G$ is the number of infinite quotients of consecutive subgroups in a polycyclic series of $G$.

Recall that the Hirsch length is well-defined by the Schreier refinement theorem, see e.g. [Rot95, Theorem 5.11]. For a finitely generated nilpotent group $G$ with a filtration $G_{\bullet}$ one has

$$h(G) = \sum_i \text{rank } G_i / G_{i+1}.$$  

**Lemma 4.4.** Let $G$ be a finitely generated nilpotent group. Then for every subgroup $V \leq G$ we have that $h(V) = h(G)$ if and only if $[G : V] < \infty$.

**Proof.** If $[G : V] < \infty$, then we can find a finite index subgroup $W \leq V$ that is normal in $G$, and the equality $h(G) = h(W) = h(V)$ follows from the Schreier refinement theorem.

Let now $G_i$ be the lower central series of $G$. Let $V \leq G$ be a subgroup with $h(V) = h(G)$ and assume in addition that $G_i \leq V$ for some $i = 1, \ldots, d + 1$. We show that $[G : V] < \infty$ by induction on $i$. For $i = 1$ the claim is trivial and for $i = d + 1$ it provides the desired equivalence.

Assume that the claim holds for some $i$. Let $V_i := V \cap G_i$ be the filtration on $V$ induced by $G_{\bullet}$ and assume $V_{i+1} = G_{i+1}$. By the assumption we have

$$\sum_{j=1}^d \text{rank } G_j / G_{j+1} = h(G) = h(V) = \sum_{j=1}^d \text{rank } V_j / V_{j+1},$$

and, since $V_j / V_{j+1} \cong V_j G_{j+1} / G_{j+1} \leq G_j / G_{j+1}$ for every $j$, this implies that $V_i / G_{i+1} \leq G_i / G_{i+1}$ is a finite index subgroup. Let $K \subset G_i$ be a finite set such that $KV / G_{i+1} = G_i / G_{i+1}$. Then $KV \leq G$ is a subgroup and a finite index extension of $V$. Moreover, we have $KV \simeq G_i$, and by the first part of the lemma we obtain $h(KV) = h(V)$.

By the induction hypothesis $KV$ has finite index in $G$, so the index of $V$ is also finite. 


Lemma 4.5. Let $G$ be a finitely generated nilpotent group with a filtration $G_*$ of length $d$ and let $V \leq G$ a subgroup. Then for every $j = 1, \ldots, d + 1$ and every $g \in G$ there exist at most finitely many finite index extensions of $V$ of the form $\langle V, gc \rangle$ with $c \in G_j$.

Proof. We use descending induction on $j$. The case $j = d + 1$ is clear, so assume that the conclusion is known for $j + 1$ and consider some $g \in G$.

Let $c_j, a = 0, 1$ be elements of $G_j$ such that $\langle V, gc_a \rangle$ are finite index extensions of $V$. Then also $\langle V G_{j+1}, gc_a \rangle / G_{j+1}$ is a finite index extension of $V G_{j+1} / G_{j+1}$, so that there exists an $m > 0$ such that $(gc_a G_{j+1})^m \subseteq V G_{j+1} / G_{j+1}$ for $a = 0, 1$.

Since the elements $c_a G_{j+1}$ are central in $G / G_{j+1}$ this implies $(c_0^{-1} c_1)^m G_{j+1} \subseteq (V G_{j+1} \cap G_j) / G_{j+1}$. But the latter group is a subgroup of the finitely generated abelian group $G_j / G_{j+1}$, so that $c_0^{-1} c_1 \in K(V G_{j+1} \cap G_j)$ for some finite set $K \subseteq G_j$ that does not depend on $c_0, c_1$.

Multiplying $c_1$ with an element of $V$ we may assume that $c_1 = c_0 K G_{j+1}$. By the induction hypothesis for each $g' \in gc_0 K$ there exist at most finitely many finite index extensions of the form $\langle V, g' c' \rangle$ with $c' \in G_{j+1}$, so we have only finitely many extensions of the form $\langle V, gc_1 \rangle$ as required.

□

Corollary 4.6. Let $G$ be a finitely generated nilpotent group and $V$ be a subgroup. Then there exist at most finitely many finite index extensions of $V$ of the form $\langle V, c \rangle$.

Proof. Consider any filtration $G_*$ and apply Lemma 4.5 with $j = 1$ and $g = 1_G$.

□

The following example shows that Corollary 4.6 may fail for virtually nilpotent groups. Consider the semidirect product $G = \mathbb{Z}_2 \ltimes \mathbb{Z}$ that is associated to the inversion action $\pi : \mathbb{Z}_2 \subseteq \mathbb{Z}$ given by $\pi(\bar{a}) (b) = (-1)^a b$. Then $G_2 = [G, G] = 2\mathbb{Z}$ is an abelian subgroup of index 4 and $G_{j+1} = [G, G_j] = 2^j \mathbb{Z}$ for all $j \in \mathbb{N}$, so $G$ is not nilpotent. Let $V = \{0\} \leq G$ be the trivial subgroup. Since we have $\langle 1a \rangle^2 = 0 \in V$ for any $a \in \mathbb{Z}$, each group of the form $\langle V, 1a \rangle$ is an extension of $V$ with index 2. On the other hand, for every value of $a$ we obtain a different extension.

Partition theorems for IP-rings. An IP-ring is a subset of $\mathcal{F}$ that consists of all finite unions of a given strictly increasing chain $\alpha_0 < \alpha_1 < \ldots$ of elements of $\mathcal{F}$ [FK85, Definition 1.1]. In particular, $\mathcal{F}$ is itself an IP-ring (associated to the chain $\{0\} < \{1\} < \ldots$). Polynomials are generally assumed to be defined on $\mathcal{F}$ even if we manipulate them only on some sub-IP-ring of $\mathcal{F}$.

Since we will be dealing a lot with assertions about sub-IP-rings we find it convenient to introduce a shorthand notation. If some statement holds for a certain sub-IP-ring $\mathcal{F}' \subset \mathcal{F}$ then we say that it holds without loss of generality (wlog). In this case we reuse the symbol $\mathcal{F}$ to denote the sub-IP-ring on which the statement holds (in particular this IP-ring may change from use to use). With this convention the basic Ramsey-type theorem about IP-rings reads as follows.

Theorem 4.7 (Hindman [Hin74]). Every finite coloring of $\mathcal{F}$ is wlog monochrome.

This is not the same as the assertion “wlog every finite coloring of $\mathcal{F}$ is monochrome”, since the latter would mean that there exists a sub-IP-ring on which every coloring is monochromatic.

As a consequence of Hindman’s theorem 4.7, a map from $\mathcal{F}$ to a compact metric space for every $\varepsilon > 0$ wlog has values in an $\varepsilon$-ball. As the next lemma shows, for polynomial maps into compact metric groups the ball can actually be chosen to be centered at the identity.

In a metric group we denote the distance to the identity by $\| \cdot \|$.

Lemma 4.8. Let $G_*$ be a prefiltration in the category of compact metric groups and $P \in \text{VIP}(G_*)$. Then for every $\varepsilon > 0$ we have wlog $\| P \| < \varepsilon$. 

Theorem 4.6. A map from $\mathcal{F}$ to a compact metric space for every $\varepsilon > 0$ wlog has values in an $\varepsilon$-ball. As the next lemma shows, for polynomial maps into compact metric groups the ball can actually be chosen to be centered at the identity.
Proof. We use induction on the length of the prefiltration $G_*$. If the prefiltration is trivial, then there is nothing to show, so assume that the conclusion is known for $G_{*,1}$.

Let $\delta, \delta' > 0$ be chosen later. By compactness and Hindman’s theorem 4.7 we may wlog assume that the image $P(\mathcal{F})$ is contained in some ball $B(g, \delta)$ with radius $\delta$ in $G_1$. By uniform continuity of the group operation we have $\tilde{D}_\beta P(\alpha) \in B(g^{-1}, \delta')$ for any $\alpha > \beta \in \mathcal{F}$ provided that $\tilde{\delta}$ is small enough depending on $\delta'$. On the other hand, for a fixed $\beta$, by the induction hypothesis we have wlog $\|\tilde{D}_\beta P\| < \delta'$, so that $\|g^{-1}\| < 2\delta'$. By continuity of inversion this implies $\|g\| < \epsilon/2$ provided that $\delta'$ is small enough. This implies $\|P\| < \epsilon$ provided that $\tilde{\delta}$ is small enough.

**Corollary 4.9 ([BHMK06, Proposition 1.1]).** Let $W_*$ be a prefiltration, $A \subset VIP(W_*)$ be finite, and $V \leq W$ be a finite index subgroup. Then wlog for every $g \in A$ we have $g(\mathcal{F}) \subset V$.

Proof. Let $g \in A$. Passing to a subgroup we may assume that $V$ is normal. Taking the quotient by $V$, we may assume that $W$ is finite and $V = \{1_W\}$. By Lemma 4.8 with an arbitrary discrete metric we may wlog assume that $g \equiv 1_W$. □

In course of proof of Theorem 4.2 it will be more convenient to use a convention for the symmetric derivative that differs from (2.18), namely

$$\hat{D}_\alpha g(\beta) = g(\alpha)^{-1} D_\alpha g(\beta).$$

Clearly a VIP group is also closed under $\hat{D}$.

**Lemma 4.10.** Let $F$ be a VIP group, $W \leq F$ be a subgroup and $V \leq W$ be a finite index subgroup. Suppose that $g \in F$ is such that the symmetric derivative $\hat{D}_\alpha g \in W$ for all $\alpha$. Then wlog for every $\alpha$ the symmetric derivative $\hat{D}_\alpha g$ coincides with an element of $V$ on some sub-IP-ring of the form $\{\beta \in \mathcal{F} : \beta > \beta_0\}$.

Proof. Since $V$ has finite index and by Hindman’s theorem 4.7 we can wlog assume that $\hat{D}_\alpha g \in W^{-1}V$ for some $w \in W$ and all $\alpha$. Assume that $w \not\in V$. Let

$$h(\alpha) := w\hat{D}_\alpha g = \begin{cases} w, & \alpha = \emptyset \\ v_\alpha \in V & \text{otherwise}. \end{cases}$$

Let $\alpha_1, \cdots, \alpha_t$ be non-empty, by induction on $d$ we see that $D_{\alpha_1} \cdots D_{\alpha_t} h(\alpha) \in V$ for all $\alpha \neq \emptyset$ and $D_{\alpha_1} \cdots D_{\alpha_t} h(\emptyset) \in V W^{-d-1} V$.

On the other hand the map $\alpha \mapsto h(\alpha)(\beta)$ is $G_\alpha$-polynomial on $\{\alpha : \alpha \cap \beta = \emptyset\}$ for fixed $\beta$. Therefore $D_{\alpha_1} \cdots D_{\alpha_t} h(\emptyset)$ vanishes at all $\beta > \alpha_d$, that is, $w$ coincides with an element of $V$ on $\{\beta : \beta > \alpha_d\}$. □

It is possible to see Lemma 4.10 (and Lemma 5.16 later on) as a special case of Corollary 4.9 by considering the quotient of VIP($G_\alpha$) by the equivalence relation of equality on IP-rings of the form $\{\alpha : \alpha > \alpha_0\}$, but we prefer not to set up additional machinery.

In order to apply the above results we need a tool that provides us with finite index subgroups. To this end recall the following multiparameter version of Hindman’s theorem 4.7

**Theorem 4.11** (Milliken [Mil75], Taylor [Tay76]). Every finite coloring of $\mathcal{F}_k^\times$ is wlog monochrome.

The next lemma is a substitute for [BFM96, Lemma 1.6] in the non-commutative case. This is the place where the concept of Hirsch length is utilized.

**Lemma 4.12.** Let $G$ be a finitely generated nilpotent group and $g : \mathcal{F} \to G$ be any map. Then wlog there exist a natural number $l > 0$ and a subgroup $W \leq G$ such that for any $\alpha_1 < \cdots < \alpha_t \in \mathcal{F}$ the elements $g_{\alpha_1}, \ldots, g_{\alpha_t}$ generate a finite index subgroup of $W$. 

Proof. By the Milliken-Taylor theorem $^{[4.11]}$ we may wlog assume that for each $l \leq h(G) + 1$ the Hirsch length $h\left(\left\langle g_{a_1}, \ldots, g_{a_l}\right\rangle\right)$ does not depend on $(\alpha_1, \ldots, \alpha_l) \in F^l$. Call this value $h_l$. It is an increasing function of $l$ that is bounded by $h(G)$, hence there exists an $l$ such that $h_l = h_{l+1}$. Fix some $(\alpha_1, \ldots, \alpha_l) \in F^l$ and let $V := \left\langle g_{a_1}, \ldots, g_{a_l}\right\rangle$.

Since $h_{l+1} = h_l$ and by Lemma $^{[4.4]}$, we see that $\langle V, g_{\alpha}\rangle$ is a finite index extension of $V$ for each $\alpha > \alpha_l$. By Corollary $^{[4.6]}$ and Hindman’s Theorem $^{[4.7]}$ we may wlog assume that each $g_{\alpha}$ lies in one such extension $W$. By definition of $h_l$ this implies that wlog for every $(\alpha_1, \ldots, \alpha_l) \in F^l$ the Hirsch length of the group $\left\langle g_{a_1}, \ldots, g_{a_l}\right\rangle \leq W$ is $h(W)$. Hence each $\left\langle g_{a_1}, \ldots, g_{a_l}\right\rangle \leq W$ is a finite index subgroup by Lemma $^{[4.4]}$.

**IP-limits.** Let $X$ be a topological space, $m \in \mathbb{N}$ and $g : F^m \to X$ be a map. We call $x \in X$ an IP-limit of $g$, in symbols $\text{IP}\lim_a g_{\bar{a}} = x$, if for every neighborhood $U$ of $x$ there exists $\alpha_0$ such that for all $\bar{a} \in F^m$, $\bar{a} > \alpha_0$, one has $g_{\bar{a}} \in U$.

By the Milliken-Taylor theorem $^{[4.11]}$ and a diagonal argument, cf. $^{[FK85]}$ Lemma 1.4, we may wlog assume the existence of an IP-limit (even of countably many IP-limits) if $X$ is a compact metric space, see $^{[FK85]}$ Theorem 1.5.

If $X$ is a Hilbert space with the weak topology we write w-IP-lim instead of IP-lim to stress the topology.

Following a tradition, we write arguments of maps defined on $F$ as subscripts in this section. We also use the notation and assumptions of Theorem $^{[4.2]}$.

The next lemma follows from the equivalence of the weak and the strong topology on the unit sphere of $H$ and is stated for convenience.

**Lemma 4.13.** Assume that $f \in \text{fix} P_{g}$, that is, that $w\text{-IP}\lim_{a} g_{\bar{a}} f = f$. Then also $\text{IP}\lim_{a} g_{\bar{a}} f = f$ (in norm).

For any subgroup $V \leq F$ we write $P_{V}$ for the orthogonal projection onto the space $\bigcap_{g \in V} \text{fix} P_{g}$.

**Lemma 4.14.** Assume that $V = \left\langle g_{1}, \ldots, g_{s}\right\rangle$ is a finitely generated group and that $P_{g_{1}}, \ldots, P_{g_{s}}$ are commuting projections. Then $P_{V} = \prod_{i=1}^{s} P_{g_{i}}$.

**Proof.** Clearly we have $P_{V} \leq \prod_{i=1}^{s} P_{g_{i}}$, so we only need to prove that each $f$ that is fixed by $P_{g_{1}}, \ldots, P_{g_{s}}$ is also fixed by $P_{g}$ for any other $g \in V$.

To this end it suffices to show that if $f$ is fixed by $P_{g}$ and $P_{h}$ for some $g, h \in V$, then it is also fixed by $P_{gh^{-1}}$. Lemma $^{[4.13]}$ shows that $\text{IP}\lim_{a} g_{\bar{a}} f = f$ and $\text{IP}\lim_{a} h_{\bar{a}} f = f$. Since each $h_{\alpha}$ is unitary we obtain $\text{IP}\lim_{a} h_{\bar{a}}^{-1} f = f$. Since each $g_{\alpha}$ is isometric, this implies

$$w\text{-IP}\lim_{a} g_{\bar{a}} h_{\bar{a}}^{-1} f = \text{IP}\lim_{a} g_{\bar{a}} h_{\bar{a}}^{-1} f = f$$

as required. □

The next lemma is the main tool to ensure IP-convergence to zero.

**Lemma 4.15** ($^{[BFM95]}$ Lemma 1.7]). Let $(P_{a})_{a \in F}$ be a family of commuting orthogonal projections on a Hilbert space $H$ and $f \in H$. Suppose that, whenever $\alpha_{1} < \cdots < \alpha_{l}$, one has $\prod_{i=1}^{l} P_{a_{i}} f = 0$. Then $\text{IP}\lim_{a} \|P_{a} f\| = 0$.

Finally, we also need a van der Corput-type estimate.

**Lemma 4.16** ($^{[FK85]}$ Lemma 5.3]). Let $(x_{a})_{a \in F}$ be a bounded family in a Hilbert space $H$. Suppose that

$$\text{IP}\lim_{a} \text{IP}\lim_{b} \langle x_{a} x_{a b} \rangle = 0.$$
Then wlog we have
\[ \text{w-IP-lim}_a x_a = 0. \]

**Proof of Theorem 4.2** We proceed by induction on the length of the prefiltration \( G_* \). If \( G_* \) is trivial there is nothing to prove. Assume that the conclusion is known for \( G_{*+1} \).

First, we prove that \( P_g \) is an orthogonal projection for any \( g \in F \) (that we now fix). Since \( P_g \) is clearly contractive it suffices to show that it is a projection.

By Lemma 4.12 we may assume that, for some \( l > 0 \) and any \( \alpha_1 < \cdots < \alpha_l \), the derivatives \( \hat{D}_{\alpha_1}g, \ldots, \hat{D}_{\alpha_l}g \) generate a finite index subgroup of some \( W \leq F_1 \) (recall that \( F_1 = F \cap \text{VIP}(G_{*+1}) \)). We split
\[ (4.17) \quad H = \bigcap_{V \leq W} \ker P_V \oplus \text{lin} \left( \bigcup_{V \leq W} \text{im} P_V \right) =: H_0 \cup H_1, \]
where \( V \) runs over the finite index subgroups of \( W \). It suffices to show \( P_g f = P^2_g f \) for each \( f \) in one of these subspaces.

Case 0. Let \( f \in H_0 \) and \( \alpha_1 < \cdots < \alpha_l \). By choice of \( W \) we know that
\[ V := \langle \hat{D}_{\alpha_1}g, \ldots, \hat{D}_{\alpha_l}g \rangle \leq W \]
is a finite index subgroup. Since the projections \( P_{\hat{D}_{\alpha}}g \) commute by the inductive hypothesis, their product equals \( P_V \) (Lemma 4.14), and we have \( P_V f = 0 \) by the assumption.

By Lemma 4.15 this implies \( \text{IP-lim}_a \| P_{\hat{D}_{\alpha}}g f \| = 0. \) Therefore
\[ \text{IP-lim}_a \| P_{\hat{D}_{\alpha}}g f \| = 0, \]
so that
\[ \text{IP-lim}_a \text{IP-lim}_\beta \langle g_{\alpha \beta}f, g_{\beta}f \rangle = 0. \]

By Definition of IP-convergence and by uniform continuity of \( P_g \) there exists \( \alpha_0 \) such that
\[ \forall \alpha > \alpha_0 \quad \rho(g_{\alpha} f, P_g f) < \epsilon \quad \text{and} \quad \rho(P_g g_{\alpha} f, P^2_g f) < \epsilon. \]

By Lemma 4.10 we can choose \( \alpha > \alpha_0 \) such that \( \hat{D}_{\alpha}g \) coincides with an element of \( V \) on some sub-IP-ring, so that in particular \( P_{\hat{D}_{\alpha}}g f = f \). By Lemma 4.13 there exists \( \beta_0 > \alpha \) such that
\[ \forall \beta > \beta_0 \quad \| (\hat{D}_{\alpha}g)_{\beta} f - f \| < \epsilon. \]
Applying \( g_{\beta}g_{\alpha} \) to the difference on the left-hand side we obtain
\[ \| g_{\alpha \beta}f - g_{\beta}g_{\alpha}f \| < \epsilon, \quad \text{so that} \quad \rho(g_{\alpha \beta}f, g_{\beta}g_{\alpha}f) < \epsilon. \]

Observe that \( \alpha \cup \beta > \alpha_0 \), so that
\[ \rho(P_g f, g_{\alpha} g_{\beta}f) < 2\epsilon. \]

Taking IP-limit along \( \beta \) we obtain
\[ \rho(P_g f, P_g g_{\alpha}f) \leq 2\epsilon. \]

A further application of the triangle inequality gives
\[ \rho(P_g f, P^2_g f) < 3\epsilon, \]
and, since \( \epsilon > 0 \) was arbitrary, we obtain \( P_g f = P^2_g f \).
Commutativity of projections. Let us now prove the second conclusion, namely that \( P_g \) and \( P_{g'} \) commute for any \( g, g' \in F \). Observe that the function \( \alpha \mapsto g_\alpha \) can be seen as a polynomial-valued function in \( \text{VIP}(P(\mathcal{F}_0, G_\alpha)) \) whose values are constant polynomials. Moreover we can consider the constant function in \( P(\mathcal{F}_0, F_\alpha) \) whose value is \( g' \). Taking their commutator we see that
\[
\alpha \mapsto [g_\alpha, g'] \in \text{VIP}(P(\mathcal{F}_0, G_\alpha)),
\]
and, since \( F \) is a VIP group, this map in fact lies in \( \text{VIP}(F_\alpha) \). By \((2.17)\) it takes values in \( F_1 \). By Lemma 3.12 we may assume that for any \( \alpha_1 < \cdots < \alpha_t \) the maps \([g_{\alpha_1}, g'], \ldots, [g_{\alpha_t}, g']\) generate a finite index subgroup of some \( W \leq F_1 \). Interchanging \( g \) and \( g' \) and repeating this argument we may also wlog assume that for any \( \alpha_1, < \cdots < \alpha_t \) the maps \([g', g], \ldots, [g_{\alpha_t}, g]\) generate a finite index subgroup of some \( W' \leq F_1 \). Consider the splitting
\[
H = \left( \bigcap_{V \leq W} \ker P_V \cap \bigcap_{V' \leq W'} \ker P_{V'} \right) \oplus \operatorname{lim} \left( \bigcup_{V \leq W} \text{im} P_V \cup \bigcup_{V' \leq W'} \text{im} P_{V'} \right) =: H_0 \cup H_1.
\]
Case 0. Let \( f \in H_0 \). As above we have \( \text{IP}-\lim_{\alpha} \|P_{[g_\alpha, g']} f \| = 0 \), and in particular
\[
0 = \text{IP}-\lim_{\alpha} \left( \text{w-IP}\text{-}\lim_{\beta} [g_\alpha, g'_\beta] f, g_\alpha^{-1} P_g f \right) = \text{IP}-\lim_{\alpha} \left( g_\alpha g'_\beta f, g'_\beta P_g f \right) = \text{IP}-\lim_{\alpha} \left( g_\alpha P_g f, P_g f \right),
\]
since \( \text{IP-\lim}_{\beta} g'_\beta P_g f = P_g f \) by Lemma 4.13. Hence \( P_g P_g f \perp P_g f \), which implies \( P_g P_g f = 0 \)
since \( P_g \) is an orthogonal projection.

Interchanging the roles of \( g \) and \( g' \), we also obtain \( P_{g'} P_g f = 0 \).

Case 1. Let \( V \leq W \) and \( f = P_V f \). By Corollary 4.9 we may wlog assume that \([g_\alpha, g'] \in V \)
for all \( \alpha \). Let \( \alpha \) be arbitrary, by Lemma 4.13 the limit
\[
\text{IP-\lim}_{\beta} [g_\alpha, g'_\beta] f = f
\]
also exists in norm. Therefore
\[
g_\alpha P_{g'} f = \text{w-IP-\lim}_{\beta} g_\alpha g'_\beta f = \text{w-IP-\lim}_{\beta} g'_\beta g_\alpha [g_\alpha, g'_\beta] f = \text{w-IP-\lim}_{\beta} g'_\beta g_\alpha f = P_{g'} g_\alpha f.
\]
Taking IP-limits on both sides we obtain
\[
P_{g'} P_g f = P_{g'} P_g f.
\]
The case \( V' \leq W' \) and \( f = P_{V'} f \) can be handled in the same way. \( \square \)

If the group \( G \) acts by measure-preserving transformations then the Hilbert space projections identified in Theorem 4.2 are in fact conditional expectations as the following folklore lemma shows.

**Lemma 4.19.** Let \( X \) be a probability space and \((T_\alpha)\) a be a net of operators on \( L^2(X) \) induced by measure-preserving transformations. Assume that \( T_\alpha \to P \) weakly for some projection \( P \). Then \( P \) is a conditional expectation.

**Proof.** Note that \( \text{im} P \cap L^\infty(X) \) is dense in \( \text{im} P \).

Let \( f, g \in \text{im} P \cap L^\infty(X) \). Since the weak and the norm topology coincide on the unit sphere of \( L^2(X) \), we have \( \|T_\alpha f - f\|_2 \to 0 \) and \( \|T_\alpha g - g\|_2 \to 0 \). Therefore
\[
\|P(f g) - f g\|_2 \leq \limsup_{\alpha} \|T_\alpha(f g) - f g\|_2 = \limsup_{\alpha} \|(T_\alpha f - f)T_\alpha g + f(T_\alpha g - g)\|_2 \leq \limsup_{\alpha} \|T_\alpha f - f\|_2 \|T_\alpha g\|_\infty + \|f\|_\infty \|T_\alpha g - g\|_2 = 0.
\]
This shows that \( \text{im} P \cap L^\infty(X) \) is an algebra, and the assertion follows. \( \square \)
**Generalized polynomials and examples of FVIP groups.** In order to obtain some tangible combinatorial applications of our results we will need non-trivial examples of FVIP groups. The first example somewhat parallels Proposition 2.24.

**Lemma 4.20** ([BHMK06]). Let \((n^1_i)_{i \in \mathbb{N}}, \ldots, (n^e_i)_{i \in \mathbb{N}} \subset \mathbb{Z}\) be any sequences, \((G, \cdot)\) be a commutative group, \((y_i)_{i \in \mathbb{N}} \subset G\) be any sequence, and \(d \in \mathbb{N}\). Then the maps of the form

\[
(4.21) \quad v(\alpha) = \sum_{i_1 < \cdots < i_e \in \alpha} n^1_{i_1} \cdots n^{e-1}_{i_{e-1}} y_{i_e}, \quad e \leq d, 1 \leq j_1, \ldots, j_{e-1} \leq a,
\]

generate an FVIP subgroup \(F \leq \text{VIP}(G_\bullet)\), where the prefiltration \(G_\bullet\) is given by \(G_0 = \cdots = G_d = G\), \(G_{d+1} = \{1_G\}\).

Maps of the form \((4.21)\) were originally studied in connection with admissible generalized polynomials (Definition 4.24). We will not return to the maps \((4.21)\) in the sequel and a proof of the above lemma is included for completeness.

**Proof.** The group \(F\) is by definition finitely generated and closed under conjugation by constants since \(G\) is commutative. It remains to check that the maps of the form \((4.21)\) are polynomial and that the group \(F\) is closed under symmetric derivatives.

To this end we use induction on \(d\). The cases \(d = 0, 1\) are clear (in the latter case the maps \((4.21)\) are IP-systems), so let \(d > 1\) and consider a map \(v\) as in \((4.21)\) with \(e = d\). For \(\beta < \alpha\) we have

\[
\tilde{D}_\beta v(\alpha) = -v(\alpha) + v(\beta \cup \alpha) - v(\beta)
\]

\[
= \sum_{k=1}^{d-1} \sum_{i_1 < \cdots < i_k \in \beta \cup \alpha} n^1_{i_1} \cdots n^{k-1}_{i_{k-1}} y_{i_k} = \sum_{k=1}^{d-1} \sum_{i_1 < \cdots < i_k \in \beta} n^1_{i_1} \cdots n^{k-1}_{i_{k-1}} \sum_{i_{k+1} < \cdots < i_{d} \in \alpha} n^k_{i_{k+1}} \cdots n^{d-1}_{i_{d-1}} y_{i_d},
\]

The underlined expression is \(G_{\bullet + 1}\)-polynomial by the induction hypothesis and lies in \(F\) by definition. Since this holds for every \(\beta\), the map \(v\) is \(G_\bullet\)-polynomial. Since the derivatives are in \(F\) for every map \(v\), the group \(F\) is FVIP.

The following basic property of FVIP groups will be used repeatedly.

**Lemma 4.22.** Let \(F, F' \leq \text{VIP}(G_\bullet)\) be FVIP groups. Then the group \(F \lor F'\) is also FVIP.

**Proof.** The group \(F \lor F'\) is clearly finitely generated and invariant under conjugation by constants. Closedness under \(\tilde{D}\) follows from the identity

\[
(4.23) \quad \tilde{D}_m(gh) = h^{-1} \tilde{D}_m g g(m) h \tilde{D}_m h g(m)^{-1}.
\]

We will now elaborate on the example that motivated Bergelson, Håland Knutson and McCutcheon to study FVIP systems in the first place [BHMK06]. They have shown that ranges of generalized polynomials from a certain class necessarily contain FVIP systems.

We begin by recalling the definition of the appropriate class. We denote the integer part function by \([\cdot]\), the nearest integer function by \([\cdot] = \lfloor \cdot + 1/2 \rfloor\) and the distance to nearest integer by \(\|a\| = |a - [a]|\) (this is consistent with the notation for general metric groups applied to \(\mathbb{R}/\mathbb{Z}\)).

**Definition 4.24.** The set \(\mathcal{G}\) of generalized polynomials (in \(l\) variables) is the smallest \(\mathbb{Z}\)-algebra of functions \(\mathbb{Z}^l \to \mathbb{Z}\) that contains \(\mathbb{Z}[x_1, \ldots, x_l]\) such that for every \(p_1, \ldots, p_t \in \mathcal{G}\) and \(c_1, \ldots, c_t \in \mathbb{R}\) the map \(\sum c_i p_i\) is in \(\mathcal{G}\). The notion of degree is extended from \(\mathbb{Z}[x_1, \ldots, x_l]\) to \(\mathcal{G}\) inductively by requiring \(\deg p_0 p_1 \leq \deg p_0 + \deg p_1, \deg(p_0 + p_1) \leq \deg p_0 + \deg p_1\).
max(deg \(p_0\), deg \(p_1\)), and \(\deg(\sum_{i=1}^{t} c_i p_i) \leq \max_i \deg p_i\), the degree of each generalized polynomial being the largest number with these properties.

The set of \(\mathcal{G}_a\) of admissible generalized polynomials is the smallest ideal of \(\mathcal{G}\) that contains the maps \(x_1, \ldots, x_t\) and is such that for every \(p_1, \ldots, p_t \in \mathcal{G}_a\), \(c_1, \ldots, c_t \in \mathbb{R}\), and \(0 < k < 1\) the map \(\sum_{i=1}^{t} c_i p_i + k\) is in \(\mathcal{G}_a\).

The construction of FVIP systems in the range of an admissible generalized polynomial in [BHMK06] proceeds by induction on the polynomial and utilizes Lemma 4.20 at the end. We give a softer argument that gives a weaker result in the sense that it does not necessarily yield an FVIP system of the form (4.21), but requires less computation.

For a ring \(R\) (with not necessarily commutative multiplication, although we will only consider \(R = \mathbb{Z}\) and \(R = \mathbb{R}\) in the sequel) and \(d \in \mathbb{N}\) we denote by \(R^d\) the prefiltration (with respect to the additive group structure) given by \(R_0 = \cdots = R_d = R\) and \(R_{d+1} = \{0\}\).

**Lemma 4.25.** Let \(F_i \leq \text{VIP}(R^d_i), i = 0, 1\), be FVIP groups. Then the pointwise products of maps from \(F_0\) and \(F_1\) generate an FVIP subgroup of \(\text{VIP}(R^{d_0+d_1})\).

**Proof.** This follows by induction on \(d_0 + d_1\) using the identity

\[
\tilde{D}_{\beta} v w = (v + \tilde{D}_{\beta} v + v(\beta))\tilde{D}_{\beta} w + (\tilde{D}_{\beta} v + v(\beta))w + (\tilde{D}_{\beta} v + v)w(\beta)
\]

for the symmetric derivative of a pointwise product.

Applying Lemma 4.8 to \(\mathbb{R}/\mathbb{Z}\) we obtain the following.

**Corollary 4.26.** Let \(P\) be an FVIP system in \(\mathbb{R}\). Then for every \(\epsilon > 0\) wlog \(|P| < \epsilon\).

This allows us to show that we can obtain \(\mathbb{Z}\)-valued FVIP systems from \(\mathbb{R}\)-valued FVIP-systems by rounding.

**Lemma 4.27.** Let \(P \in \text{VIP}(\mathbb{R}^d)\) be an FVIP system. Then wlog \(|P| \in \text{VIP}(\mathbb{Z}^d)\) and \(|P|\) is an FVIP system.

**Proof.** We use induction on \(d\). For \(d = 0\) there is nothing to show, so assume that \(d > 0\). By the assumption every symmetric derivative of \((P_a)\) lies in an FVIP group of polynomials of degree < \(d\) that is generated by \(q_1, \ldots, q_a\), say. By the induction hypothesis we know that wlog each \(|q_i|\) is again an FVIP system and by Lemma 4.22 they lie in some FVIP group \(F\). By Corollary 4.26 we may assume wlog that \(|P| < 1/12\). Let now \(\beta\) be given, by the hypothesis we have

\[
\tilde{D}_{\beta} P(\alpha) = \sum_i c_i q_i(\alpha) \quad \text{for } \alpha > \beta
\]

with some \(c_i \in \mathbb{Z}\). By Corollary 4.26 we may wlog assume that \(|c_i| \cdot |q_i|(\alpha) < 1/(4 \cdot 2^j)\) for all \(\alpha > \beta\). This implies

\[
|\tilde{D}_{\beta} |P|(\alpha) - \sum_i c_i |q_i|(\alpha)| < 1/2 \quad \text{for } \alpha > \beta,
\]

so that

\[
\tilde{D}_{\beta} |P|(\alpha) = \sum_i c_i |q_i|(\alpha) \quad \text{for } \alpha > \beta,
\]

since both sides are integer-values functions. In fact we can do this for all \(\beta\) with fixed \(\max \beta\) simultaneously. By a diagonal argument, cf. [PK85] Lemma 1.4], we may then assume that for every \(\beta\) we have

\[
\tilde{D}_{\beta} |P|(\alpha) = \sum_i c_i |q_i|(\alpha) \quad \text{for } \alpha > \beta
\]

with some \(c_i \in \mathbb{Z}\). Hence \(F \lor \{|P|\} \leq \text{VIP}(\mathbb{Z}^d)\) is an FVIP group. \(\square\)
Recall that an IP-system in $\mathbb{Z}^l$ is a family $(n_a)_{a \in S} \subset \mathbb{Z}^l$ such that $n_{a,\beta} = n_a + n_\beta$ whenever $\alpha, \beta \in \mathcal{F}$ are disjoint.

**Theorem 4.28** ([BHKM06, Theorem 2.8]). For every generalized polynomial $p : \mathbb{Z}^l \to \mathbb{Z}$ and every FVIP system $(n_a)_a$ in $\mathbb{Z}^l$ of degree at most $d$ there exists $n \in \mathbb{Z}$ such that the IP-sequence $(p(n_a) - n)_{a \in S}$ is wlog FVIP of degree at most $d \deg p$. If $p$ is admissible, then we may assume $n = 0$.

**Proof.** We begin with the first part and use induction on $p$. The class of maps that satisfy the conclusion is closed under $\mathbb{Z}$-linear combinations by Lemma 4.22 and under multiplication by Lemma 4.25. This class clearly contains the polynomials $1, x_1, \ldots, x_l$. Thus it remains to show that, whenever $p_1, \ldots, p_l \in \mathcal{G}$ satisfy the conclusion and $c_1, \ldots, c_l \in \mathbb{R}$, the map $[P]$ with $P = \sum_{i=1}^l c_i p_i$ also satisfies the conclusion.

By the assumption we have wlog that $(P(n_a) - C)_a$ is an $\mathbb{R}$-valued FVIP system for some $C \in \mathbb{R}$. By Hindman’s theorem 4.7, we may wlog assume that $[P(n_a)] = [P(n_a) - C] + n$ for some integer $n$ with $|n - C| < 2$ and all $a \in \mathcal{F}$. The conclusion follows from Lemma 4.27.

Now we consider admissible generalized polynomials $p$ and use induction on $p$ again. The conclusion clearly holds for $x_1, \ldots, x_l$, passes to linear combinations and passes to products with arbitrary generalized polynomials by Lemma 4.25 and the first part of the statement. Assume now that $p_1, \ldots, p_l \in \mathcal{G}^l$ satisfy the conclusion and $c_1, \ldots, c_l \in \mathbb{R}$, $0 < k < 1$. Then $(P(n_a))_a$ with $P := \sum_i c_i p_i$ is wlog an $\mathbb{R}$-valued FVIP system, and by Corollary 4.26 we have wlog $\|P\| < \min(k, 1 - k)$. This implies $[P(n_a) + k] = [P(n_a)]$ and this is wlog an FVIP system by Lemma 4.27.

As an aside, consider the set of real-valued generalized polynomials $\mathcal{R} \mathcal{G}$ [BM10, Definition 3.1] that is defined similarly to $\mathcal{G}$, except that it is required to be an $\mathbb{R}$-algebra. Following the proof of Theorem 4.28, we see that for every $p \in \mathcal{R} \mathcal{G}$ and every FVIP system $(n_a)_a \subset \mathbb{Z}^l$ wlog there exists a constant $C \in \mathbb{R}$ such that $(p(n_a) - C)_a$ is an FVIP system. Clearly, if $p$ is of the form $[g]$ then $C \in \mathbb{Z}$ and if $p \in \mathbb{R}[x_1, \ldots, x_l]$ with zero constant term then $C = 0$. This, together with Corollary 4.26, implies (an FVIP* version of) [BL07, Theorem D].

Our main example (that also leads to Theorem 1.1) are maps induced by admissible generalized polynomial sequences in finitely generated nilpotent groups.

**Lemma 4.29**. Let $G$ be a finitely generated nilpotent group with a filtration $G_*$. Let $p : \mathbb{Z}^l \to \mathbb{Z}$ be an admissible generalized polynomial, $(n_a)_a \subset \mathbb{Z}^l$ be an FVIP system of degree at most $d$ and $g \in G_{d \deg p}$. Then wlog $(g^{p(n_a)})_a$ is an element of VIP($G_*$) and an FVIP system.

**Proof.** By Theorem 4.28, we can wlog assume that $(p(n_a))_a$ is a $\mathbb{Z}$-valued FVIP system of degree $d \deg p$. Using the (family of) homomorphism $(s) \cong G_{d \deg p} \to G_*$, $1 \to g$, we see that $(g^{p(n_a)})_a$ is contained in a finitely generated subgroup $F_0 \leq VIP(G_*)$ that is closed under $\tilde{D}$.

Let $A \subset F_0$ and $B \subset G$ be finite generating sets. Then the group generated by $bab^{-1}$, $a \in A$, $b \in B$, is FVIP in view of the identity (4.23).

5. **Measurable multiple recurrence**

Following the general scheme of Furstenberg’s proof, we will obtain our multiple recurrence theorem by (in general transfinite) induction on a suitable chain of factors of the given measure-preserving system. For the whole section we fix a nilpotent group $G$ with a filtration $G_*$ and an FVIP group $F \leq VIP(G_*)$.

In the induction step we pass from a factor to a “primitive extension” that enjoys a dichotomy: each element of $F$ acts on it either relatively compactly or relatively mixingly. Since the reasoning largely parallels the commutative case here, we are able to refer to the
article of Bergelson and McCutcheon [BM00] for many proofs. The parts of the argument that do require substantial changes are given in full detail.

Whenever we talk about measure spaces \((X, \mathcal{A}, \mu), (Y, \mathcal{B}, \nu), \) or \((Z, \mathcal{C}, \gamma)\) we suppose that they are regular and that \(G\) acts on them on the right by measure-preserving transformations. This induces a left action on the corresponding \(L^2\) spaces. Recall that to every factor map \((Z, \mathcal{C}, \gamma) \rightarrow (Y, \mathcal{B}, \nu)\) there is associated an essentially unique measure disintegration

\[
\gamma = \int_{y \in Y} \gamma_y d\nu(y),
\]

see [Fur81, §5.4]. We write \(\|\cdot\|_y\) for the norm on \(L^2(Z, \gamma_y)\). Recall also that the fiber product \(Z \times_Y Z\) is the space \(Z \times Z\) with the measure \(\int_{y \in Y} \gamma_y \otimes \gamma_y d\nu(y)\).

**Compact extensions.** We begin with the appropriate notion of relative compactness. Heuristically, an extension is relatively compact if it is generated by the image of a relatively Hilbert-Schmidt operator.

**Definition 5.1 ([BM00, Definition 3.4]).** Let \(Z \rightarrow Y\) be a factor. A \(Z|Y\)-kernel is a function \(H \in L^\infty(Z \times_Y Z)\) such that

\[
\int H(z_1, z_2) d\gamma_{z_2}(z_1) = 0
\]

for a.e. \(z_2 \in Z\). If \(H\) is a \(Z|Y\)-kernel and \(\phi \in L^2(Z)\) then

\[
H \ast \phi(z_1) := \int H(z_1, z_2) \phi(z_2) d\gamma_{z_1}(z_2).
\]

The map \(\phi \mapsto H \ast \phi\) is a Hilbert-Schmidt operator on almost every fiber over \(Y\) with uniformly bounded Hilbert-Schmidt norm. These operators are self-adjoint provided that \(H(z_1, z_2) = H(z_2, z_1)\) a.e.

**Definition 5.2 ([BM00, Definition 3.6]).** Suppose that \(X \rightarrow Z \rightarrow Y\) is a chain of factors, \(K \leq F\) is a subgroup and \(H\) is a non-trivial self-adjoint \(X|Y\)-kernel that is \(K\)-invariant in the sense that

\[
\text{IP-lim}_{\alpha} g(\alpha) H = H
\]

for every \(g \in K\). The extension \(Z \rightarrow Y\) is called \(K\)-compact if it is generated by functions of the form \(H \ast \phi, \phi \in L^2(X)\).

**Lemma 5.3 ([BM00, Remark 3.7(i)]).** Let \(X \rightarrow Z \rightarrow Y\) be a chain of factors in which \(Z \rightarrow Y\) is a \(K\)-compact extension generated by a \(X|Y\)-kernel \(H\). Then \(H\) is in fact a \(Z|Y\)-kernel and \(Z\) is generated by functions of the form \(H \ast \phi, \phi \in L^2(Z)\).
Proof. Call the projection maps $\pi : X \to Z$, $\theta : X \to Y$. Let $\phi \in L^2(X)$. Since $H \ast \phi$ is $Z$-measurable we have
\[
H \ast \phi(x) = \int H \ast \phi(x_1) d\mu_{\pi(x)}(x_1)
\]
\[
= \int \int H(x_1, x_2) \phi(w_2) d\mu_{\theta(x_2)}(x_2) d\mu_{\pi(x)}(x_1)
\]
\[
= \int \int H(x_1, x_2) \phi(x_2) d\mu_{\theta(x)}(x_2) d\mu_{\pi(x)}(x_1)
\]
\[
= \int \int H(x_1, x_2) d\mu_{\pi(x)}(x_1) \phi(x_2) d\mu_{\theta(x)}(x_2)
\]
\[
= \mathbb{E}(H|Z \times_Y X) \ast \phi(w).
\]
Since this holds for all $\phi$ we obtain $H = \mathbb{E}(H|Z \times_Y X)$. Since $H$ is self-adjoint this implies that $H$ is $Z \times_Y Z$-measurable. This in turn implies that $H \ast \phi = H \ast \mathbb{E}(\phi|Z)$ for all $\phi \in L^2(X)$. □

In view of this lemma the reference to the ambient space $X$ is not necessary in the definition of a $K$-compact extension. Just like in the commutative case, compactness is preserved upon taking fiber products (this is only used in the part of the proof of Theorem 5.18 that we do not write out).

Lemma 5.4 ([BM00] Remark 3.7(ii)). Let $Z \to Y$ be a $K$-compact extension. Then $Z \times_Y Z \to Y$ is also a $K$-compact extension.

Mixing and primitive extensions. Now we define what we mean by relative mixing and the dichotomy between relative compactness and relative mixing.

Definition 5.5 ([BM00] Definition 3.5]). Let $Z \to Y$ be an extension. A map $g \in F$ is called mixing on $Z$ relatively to $Y$ if for every $H \in L^2(Z \times_Y Z)$ with $\mathbb{E}(H|Y) = 0$ one has $\lim_{w \to 0} g(\alpha) H = 0$. An extension $Z \to Y$ is called $K$-primitive if it is $K$-compact and each $g \in F \setminus K$ is mixing on $Z$ relative to $Y$.

The above notion of mixing might be more appropriately called “mild mixing”, but we choose a shorter name since there will be no danger of confusion.

The next lemma is used in the suppressed part of the proof of Theorem 5.18.

Lemma 5.6 ([BM00] Proposition 3.8]). Let $Z \to Y$ be a $K$-primitive extension. Then $Z \times_Y Z \to Y$ is also a $K$-primitive extension.

Like in the commutative setting [McC05], Lemma 2.8] the compact part of a primitive extension is wlog closed under taking derivatives, but there is also a new aspect, namely that it is also closed under conjugation by constants.

Lemma 5.7. Let $Z \to Y$ be a $K$-primitive extension. Then $K$ is closed under conjugation by constant functions. Moreover wlog $K$ is an FVIP group.

Proof. Let $g \in F \setminus K$, $h \in G$ and $H \in L^2(Z \times_Y Z)$ be such that $\mathbb{E}(H|Y) = 0$. Then
\[
\lim_{\alpha \to 0} (h^{-1} g h)(\alpha) H = h^{-1} \lim_{\alpha \to 0} g(\alpha)(hH) = 0,
\]
so that $F \setminus K$ is closed under conjugation by constant functions, so that $K$ is also closed under conjugation by constant functions.

Since $F$ is Noetherian, the subgroup $K$ is finitely generated as a semigroup. Fix a finite set of generators for $K$. By Hindman’s Theorem 4.7 we may wlog assume that for every
generator \( g \) we have either \( \tilde{D}_a g \in K \) for all \( a \in \mathcal{F} \) or \( \tilde{D}_a g \notin K \) for all \( a \in \mathcal{F} \). In the latter case we obtain
\[
0 = \text{w-IP-lim}_{a,\beta} \tilde{D}_\beta g(a)H = \text{IP-lim}_{a,\beta} g(a)^{-1} g(\alpha \cup \beta) g(\beta)^{-1} H = H,
\]
a contradiction. Thus we may assume that all derivatives of the generators lie in \( K \). This extends to the whole group \( K \) by (4.23) and invariance of \( K \) under conjugation by constants. \( \square \)

**Existence of primitive extensions.** Since our proof proceeds by induction over primitive extensions we need to know that such extensions exist. First, we need a tool to locate non-trivial kernels.

**Lemma 5.8 ([BM00, Lemma 3.12]).** Let \( X \to Y \) be an extension. Suppose that \( 0 \neq H \in L^2(X \times_\gamma X) \) satisfies \( E(H|Y) = 0 \) and that there exists \( g \in F \) such that \( \text{IP-lim}_a g(\alpha)H = H \).

Then there exists a non-trivial self-adjoint non-negative definite \( X|Y \)-kernel \( H' \) such that \( \text{IP-lim}_a g(\alpha)H' = H' \).

Second, we have to make sure that we cannot accidentally trivialize them.

**Lemma 5.9 ([BM00, Lemma 3.14]).** Let \( Z \to Y \) be a \( K \)-compact extension. Suppose that for some \( g \in K \) and self-adjoint non-negative definite \( Z|Y \)-kernel \( H \) we have
\[
\text{IP-lim}_a \int (g(\alpha)H(f' \otimes f^\gamma) d\gamma = 0
\]
for all \( f' \in L^\infty(Z) \). Then \( H = 0 \).

The next theorem that provides existence of primitive extensions can be proved in the same way as in the commutative case ([BM00, Theorem 3.15]). The only change is that Theorem 4.2 is used instead of ([BM00, Theorem 2.17] (note that \( F \) is Noetherian, since it is a finitely generated nilpotent group).

**Theorem 5.10.** Let \( X \to Y \) be a proper factor. Then there exists a subgroup \( K \leq F \) and a factor \( X \to Z \to Y \) such that the extension \( Z \to Y \) is proper and wlog \( K \)-primitive.

**Almost periodic functions.** For the rest of Section 5 we fix a good group \( \text{FE} \leq \text{VIP}(G_\epsilon)^{\otimes \omega} \).

We will describe what we mean by “good” in Definition 5.23, for the moment it suffices to say that \( \text{FE} \) is countable.

**Definition 5.11 ([BM00, Definition 3.1]).** Suppose that \( (Z, \mathcal{E}, \gamma) \to (Y, \mathcal{B}, \nu) \) is a factor and \( K \leq F \) a subgroup. A function \( f \in L^2(Z) \) is called \( K \)-almost periodic if for every \( \epsilon > 0 \) there exist \( g_1, \ldots, g_l \in L^2(Z) \) and \( D \in \mathcal{B} \) with \( \nu(D) < \epsilon \) such that for every \( \delta > 0 \) and \( R \in K^{\otimes \omega} \cap \text{FE} \) there exists \( \alpha_0 \) such that for every \( \alpha_0 < \tilde{\alpha} \in \mathcal{F}_\omega \) there exists a set \( E = E(\tilde{\alpha}) \in \mathcal{B} \) with \( \nu(E) < \delta \) such that for all \( y \in Y \setminus (D \cup E) \) there exists \( 1 \leq j \leq l \) such that
\[
||R(\alpha)f - g_j||_Y < \epsilon.
\]
The set of \( K \)-almost periodic functions is denoted by \( \text{AP}(Z, Y, K) \).

The next lemma says that a characteristic function that can be approximated by almost periodic functions can be replaced by an almost periodic function right away.

**Lemma 5.12 ([BM00, Theorem 3.3]).** Let \( A \in \mathcal{E} \) be such that \( 1_\mathcal{A} \in \text{AP}(Z, Y, K) \) and \( \delta > 0 \). Then there exists a set \( A' \subset A \) such that \( \gamma(A \setminus A') < \delta \) and \( 1_{A'} \in \text{AP}(Z, Y, K) \).

In the following lemma we have to restrict ourselves to \( K^{\otimes \omega} \cap \text{FE} \) since \( K^{\otimes \omega} \) need not be countable.
**Lemma 5.13** ([BM00 Proposition 3.9]). Let \( X \to Y \) be an extension, \( K \leq F \) a subgroup and \( H \) a \( X\mid Y \)-kernel that is \( K \)-invariant. Then wlog for all \( R \in K^{\aleph_0} \cap FE \) and \( \epsilon > 0 \) there exists \( \alpha_0 \) such that for all \( \alpha < \alpha_0 \) we have
\[
\|R(\alpha)H - H\| < \epsilon.
\]

With help of the above lemma we can show that in fact wlog every characteristic function can be approximated by almost periodic functions. In view of Lemma 5.12 this allows us to reduce the question of multiple recurrence in a primitive extension to multiple recurrence for (relatively) almost periodic functions.

**Lemma 5.14** ([BM00 Theorem 3.11]). Let \( Z \to Y \) be a \( K \)-compact extension. Then wlog \( \text{AP}(Z, Y, K) \) is dense in \( L^2(Z) \).

**Multiple mixing.** Under sufficiently strong relative mixing assumptions the limit behavior of a multicorrelation sequence \( \prod S_i(\alpha)f_i \) on a primitive extension only depends on the expectations of the functions on the base space. The appropriate conditions on the set \( \{S_i\}_i \) are as follows.

**Definition 5.15.** Let \( K \leq F \) be a subgroup. A subset \( A \subset F \) is called \( K \)-mixing if \( 1_A \in A \) and \( g^{-1}h \in F \setminus K \) whenever \( g \neq h \in A \).

The requirement \( 1_A \in A \) is not essential, but it is convenient in inductive arguments. In order to apply PET induction we will need the next lemma.

We say that a subgroup \( K \leq F \) is invariant under equality of tails if whenever \( S \in K \) and \( T \in F \) are such that there exists \( \beta \in \mathcal{F} \) with \( S_\beta = T_\alpha \) for all \( \alpha > \beta \) we have \( T \in K \). Every group \( K \leq F \) that is the compact part of some primitive extension has this property.

**Lemma 5.16.** Let \( K \leq F \) be a subgroup that is invariant under equality of tails. Let \( S, T \in F \) be such that \( S^{-1}T \notin K \). Then wlog
\[
(S\tilde{D}_\beta S)^{-1}(T\tilde{D}_\beta T) \notin K \quad \text{and} \quad S^{-1}(T\tilde{D}_\beta T) \notin K
\]
for all \( \beta \in \mathcal{F} \).

**Proof.** If the first conclusion fails then by Hindman’s theorem 4.7 wlog
\[
h(\alpha) := (S\tilde{D}_\alpha S)^{-1}(T\tilde{D}_\alpha T) \in K \quad \text{for all} \quad \alpha \in \mathcal{F}
\]
and \( h(0) \notin K \). Proceed as in the proof of Lemma 4.10. Analogously for the second conclusion.

The next lemma is a manifestation of the principle that compact orbits can be thought of as being constant.

**Lemma 5.17** ([BM00 Proposition 4.2]). Let \( Z \to Y \) be a \( K \)-primitive extension, \( R^\beta \in K \) for each \( \beta \in \mathcal{F} \) and \( W \in F \setminus K \). Let also \( f, f' \in L^\infty(Z) \) be such that either \( \mathbb{E}(f \mid Y) = 0 \) or \( \mathbb{E}(f' \mid Y) = 0 \). Then wlog
\[
\text{IP-}\lim_{\beta, \alpha} \|\mathbb{E}(R^\beta(\alpha)f)W(\beta)f'\mid Y)\| = 0.
\]

We come to the central result on multiple mixing.

**Theorem 5.18** (cf. [BM00 Theorem 4.10]). Let \( K \leq F \) be a subgroup. For every \( K \)-mixing set \( \{S_0 \equiv 1_G, S_1, \ldots, S_t\} \subset F \) the following statements hold.

1. For every \( K \)-primitive extension \( Z \to Y \) and any \( f_0, \ldots, f_t \in L^\infty(Z) \) we have wlog
\[
\text{w-IP-}\lim_{\alpha} \prod_{i=1}^t S_i(\alpha)f_i - \prod_{i=1}^t S_i(\alpha)\mathbb{E}(f_i \mid Y) = 0.
\]
(2) For every $K$-primitive extension $Z \to Y$ and any $f_0, \ldots, f_t \in L^\infty(Z)$ we have \(\text{wlog} \): 
\[
\text{IP}\lim_{\alpha} \left\| \mathbb{E} \left( \prod_{i=0}^{t} S_i(\alpha)f_i \middle| Y \right) - \prod_{i=0}^{t} S_i(\alpha) \mathbb{E}(f_i \middle| Y) \right\| = 0.
\]

(3) For every $K$-primitive extension $Z \to Y$, any $U_{i,j} \in K$, and any $f_{i,j} \in L^\infty(Z)$ we have \(\text{wlog} \):
\[
\text{IP}\lim_{\alpha} \left\| \mathbb{E} \left( \prod_{i=0}^{t} S_i(\alpha) \left( \sum_{j=0}^{s} U_{i,j}(\alpha)f_{i,j} \right) \middle| Y \right) - \prod_{i=0}^{t} S_i(\alpha) \mathbb{E}\left( \sum_{j=0}^{s} U_{i,j}(\alpha)f_{i,j} \middle| Y \right) \right\| = 0.
\]

We point out that the main induction loop is on the mixing set. It is essential that, given $K \leq F$, all statements are proved simultaneously for all $K$-compact extensions since the step from weak convergence to strong convergence involves a fiber product via Lemma 5.6.

**Proof.** The proof is by PET-induction on the mixing set. We only prove that the last statement for mixing sets with lower weight vector implies the first, the proofs of other implications are the same as in the commutative case.

By the telescope identity it suffices to consider the case $\mathbb{E}(f_{i_0} \middle| Y) = 0$ for some $i_0$. By the van der Corput Lemma 4.16 it suffices to show that \(\text{wlog} \):
\[
\text{IP}\lim_{\beta,\alpha} \int_{Z} \prod_{i=1}^{t} S_i(\alpha)f_i \prod_{i=1}^{t} S_i(\alpha \cup \beta)\tilde{f}_i = 0.
\]

This limit can be written as
\[
\text{IP}\lim_{\beta,\alpha} \int_{Z} \prod_{i=1}^{t} S_i(\alpha)f_i \prod_{i=1}^{t} S_i(\alpha \cup \beta)\tilde{f}_i = \text{IP}\lim_{\beta,\alpha} \int_{Z} \prod_{i=1}^{t} S_i(\alpha)f_i \prod_{i=1}^{t} S_i(\alpha)\tilde{f}_i = 0.
\]

By Lemma 5.16 we may \(\text{wlog} \) assume that $T_{i,\beta}^{-1}T_{j,\beta}$ and $S_i^{-1}T_{j,\beta}$ are mixing for all $\beta \in F$ provided that $i \neq j$. Re-indexing if necessary and using Hindman’s theorem 4.7 we may \(\text{wlog} \) assume $S_i^{-1}T_{i,\beta} \in K$ for all $\beta \in F$ and $i \leq w$ and $S_i^{-1}T_{i,\beta} \notin K$ for all $\beta \in F$ and $i > w$ for some $w = 0, \ldots, t$. Thus
\[
(5.19) \quad S_{0}, S_{1}, \ldots, S_{t}, T_{w+1,\beta}, \ldots, T_{t,\beta}
\]

is a $K$-mixing set for every $\beta \neq 0$. Moreover it has the same weight vector as $\{S_{1}, \ldots, S_{t}\}$ since $T_{i,\beta} \sim S_{i}$. Assume that $S_{j}, j \neq 0$, has the maximal level in (5.19). We have to show
\[
\text{IP}\lim_{\beta,\alpha} \int_{Z} \prod_{i=1}^{w} S_j^{-1}(\alpha)S_i(\alpha)(f_i \tilde{D}_\beta S_i(\alpha)(S_i(\beta)\tilde{f}_i))
\]
\[
\cdot \prod_{i=w+1}^{t} S_j^{-1}(\alpha)S_i(\alpha)f_i S_j^{-1}(\alpha)T_{i,\beta}(\alpha)(S_i(\beta)\tilde{f}_i) = 0.
\]

For each fixed $\beta \in F$ the limit along $\alpha$ comes from the $K$-mixing set
\[
S_j^{-1}S_{1}, \ldots, S_j^{-1}S_{t}, S_j^{-1}T_{w+1,\beta}, \ldots, S_j^{-1}T_{t,\beta}
\]
that has lower weight vector. Hence we can apply the induction hypothesis, thereby obtaining that the limit equals

\[
\operatorname{IP-lim}_{\beta, \alpha} \left( \prod_{i=1}^{w} S^{-1}_j(\alpha)S_i(\alpha)\mathbb{E}\left(f_i, \mathcal{D}_\beta S_i(\alpha)(\beta)\tilde{f}_i\right|Y) \cdot \prod_{i=w+1}^{t} S^{-1}_j(\alpha)S_i(\alpha)\mathbb{E}\left(f_i|Y\right)S^{-1}_j(\alpha)T_{i, \beta}(\alpha)\mathbb{E}\left(S_i(\beta)\tilde{f}_i\right|Y) \right)
\]

This clearly vanishes if \( i_0 > w \), otherwise use Lemma 5.17.

\[ \square \]

**Multiparameter multiple mixing.** In fact we need some information about relative polynomial mixing in several variables. First we need to say what we understand under a mixing system of polynomial expressions. Recall that by definition each \( S \in F^{\otimes m} \) can be written in the form

\[
S(\alpha_1, \ldots, \alpha_m) = W^{(\alpha_1, \ldots, \alpha_{m-1})}(\alpha_m) \ldots W^{\alpha_1}(\alpha_2)W(\alpha_1).
\]

**Definition 5.21.** Let \( K \leq F \) be a subgroup and \( m \in \mathbb{N} \). A set \( \{S_i\}_{i=0}^{m} \subset F^{\otimes m} \) is called \( K \)-mixing if \( S_0 \equiv 1_G \), the polynomial expressions \( \{S_i\} \) are pairwise distinct, and for all \( r \) and \( i \neq j \) we have either \( \forall \bar{a} \in \mathcal{F}_r W_i^{\bar{a}} = W_j^{\bar{a}} \) or \( \forall \bar{a} \in \mathcal{F}_r W_i^{\bar{a}} \cdot W_j^{\bar{a}} / E \not\in K \).

For \( m = 1 \) this coincides with Definition 5.15. However, in general, this definition requires more than \( \{W_i^{\bar{a}}\}_i \) being (up to multiplicity) a \( K \)-mixing set in the sense of Definition 5.15 for every \( \bar{a} \).

**Theorem 5.22 (cf. [BM00, Theorem 4.12]).** Let \( Z \to Y \) be a \( K \)-primitive extension. Then for every \( m \geq 1 \), every \( K \)-mixing set \( \{S_0, \ldots, S_t\} \subset F^{\otimes m} \) and any \( f_0, \ldots, f_t \in L^\infty(Z) \) we have wlog

\[
\operatorname{IP-lim}_{\alpha_1, \ldots, \alpha_m} \left\| \mathbb{E}\left( \prod_{i=0}^{t} S_i(\alpha_1, \ldots, \alpha_m)f_i\right|Y) - \prod_{i=0}^{t} S_i(\alpha_1, \ldots, \alpha_m)\mathbb{E}(f_i|Y) \right\| = 0.
\]

**Proof.** We use induction on \( m \). The case \( m = 0 \) is trivial since the product then consists only of one term. Assume that the conclusion holds for \( m \) and consider a \( K \)-mixing set of polynomial expressions in \( m + 1 \) variables. For brevity we write \( \vec{a} = (\alpha_1, \ldots, \alpha_m) \) and \( \alpha = \alpha_{m+1} \). We may assume that \( \|f_i\|_\infty \leq 1 \) for all \( i \) and \( \mathbb{E}(f_i|Y) = 0 \) for some \( i_0 \).

By Definition 5.21 and with notation from (5.20), for every \( \vec{a} \) there exists a \( K \)-mixing set \( \{V_\vec{a}\} \subset F \) such that \( W_\vec{a} = V_{\vec{a}} \), where the assignment \( i \to j_i \) does not depend on \( \vec{a} \). Let also

\[
A_j = \{S(\cdot) = S_j(\cdot, \emptyset) : j_i = j\}.
\]

In view of the Milliken-Taylor theorem 4.11 and by a diagonal argument, cf. [FK85, Lemma 1.4], it suffices to show that for every \( \delta > 0 \) there exist \( \vec{a} < \alpha \) such that

\[
\left\| \mathbb{E}\left( \prod_{j} V_j^{\vec{a}}(\alpha)(\prod_{S \in A_j} S(\vec{a})f_{S,j})\right|Y\right\| \leq \delta
\]

provided that \( \mathbb{E}(f_{S_0,j}|Y) = 0 \) for some \( j_0, S_0 \). By the induction hypothesis there exists \( \vec{a} \) such that

\[
\left\| \mathbb{E}\left( \prod_{S \in A_{j_0}} S(\vec{a})f_{S,j_0}\right|Y\right\| < \delta,
\]

since \( A_{j_0} \) is a \( K \)-mixing set. This implies

\[
\left\| \prod_{j} V_j^{\vec{a}}(\alpha)\mathbb{E}\left( \prod_{S \in A_j} S(\vec{a})f_{S,j}\right|Y\right\| < \delta
\]
for all $\alpha > \bar{\alpha}$. Since $\{V^\alpha_j\}_j$ is a K-mixing set, Theorem \[5.18\] implies
\[
\text{IP-}\lim_{\bar{\alpha}} \left\| \mathbb{E}\left( \prod_j V_{j\bar{\alpha}}(\alpha)^{-1} \prod_{S\inA_j} S(\bar{\alpha}f_{S,j}^)|Y) \right) \right\| \leq \delta.
\]

**Lifting multiple recurrence to a primitive extension.** We are nearing our main result, a multiple recurrence theorem for polynomial expressions. In order to guarantee the existence of the limits that we will encounter during its proof we have to restrict ourselves to a certain good subgroup of the group of polynomial expressions. It will be shown later that this restriction can be removed, cf. Corollary \[5.31\].

**Definition 5.23.** We call a group $\text{FE} \leq \text{VIP}(G_\star)^{\otimes \omega}$ good if it has the following properties.

1. (Cardinality) $\text{FE}$ is countable.
2. (Substitution) If $m \in \mathbb{N}$, $g \in \text{VIP}(G_\star)^{\otimes m} \cap \text{FE}$, and $\vec{g} \in \mathcal{F}_m$, then $g[\vec{g}] \in \text{FE}$.
3. (Decomposition) If $\text{K} \leq F$ is a subgroup invariant under conjugation by constants and $\{S_i\}_{i=0}^t \subset F^{\otimes m} \cap \text{FE}$ is a finite set with $S_0 \equiv 1_F$, then we have finite sets $\{T_k\}_{k=0}^{m-1} \subset F^{\otimes m} \cap \text{FE}$ and $\{R_i\}_{i=0}^m \subset K^{\otimes m} \cap \text{FE}$ with $R_0 = T_0 \equiv 1_F$ such that $S_i = T_k R_i$ and for every sub-IP-ring $\mathcal{F}' \subset \mathcal{F}$ the set $\{T_k\}$ is wlog $K$-mixing.

The property of being good is hereditary in the sense that a group that is good with respect to some $\text{IP}$-ring is also good with respect to any sub-$\text{IP}$-ring.

Let $(X, \mathcal{A}, \mu)$ be a regular measure space with a right action of $G$ by measure-preserving transformations. Let also $\text{FE} \leq F^{\otimes \omega}$ be a good group. By Hindman's theorem \[4.7\] we may wlog assume that
\[
\text{w-IP-}\lim_{\bar{\alpha}} g(\alpha) f
\]
eexists for every $g \in F$ and $f \in L^2(X)$. By the Milliken-Taylor theorem \[4.11\] we may wlog assume that the limit
\[
\text{IP-}\lim_{\bar{\alpha} \in \mathcal{F}_m} \mu \left( \bigcap_{i=0}^t A S_i(\bar{\alpha})^{-1} \right)
\]
eexists for every $m \in \mathbb{N}$, every $A \in \mathcal{A}$, and every finite set $\{S_0, \ldots, S_t\} \subset F^{\otimes m} \cap \text{FE}$.

**Definition 5.24.** A factor $(X, \mathcal{A}, \mu) \to (Y, \mathcal{B}, \nu)$ is said to have the SZ (Szemerédi) property if for every $B \in \mathcal{B}$ with $\nu(B) > 0$ and every set of polynomial expressions $\{S_0 \equiv 1_G, S_1, \ldots, S_t\} \subset F^{\otimes m} \cap \text{FE}$ one has
\[
\text{IP-}\lim_{\bar{\alpha} \in \mathcal{F}_m} \mu \left( \bigcap_{i=0}^t B S_i(\bar{\alpha})^{-1} \right) > a(B, m, \{S_i\},) > 0.
\]

Note that unlike in \([BM00]\), Definition 5.1] the constant depends not only on the number of polynomial expressions. We cannot obtain more uniform results due to the lack of control on the number $w$ provided by Corollary \[3.8\].

**Lemma 5.25** ([BM00], Prop. 5.2]. For every separable regular measure-preserving system $X$ there exists a maximal factor that has the SZ property.

**Theorem 5.26.** The identity factor $X \to X$ has the SZ property.

This is the central result of the present article and generalizes \[BM00\], Theorem 1.3].

**Proof.** By Lemma [5.25] there exists a maximal factor $X \to Y$ with the SZ property. Assume that $X \neq Y$, then by Theorem [5.10] wlog there exists a subgroup $K \leq F$ and a factor $X \to Z$ such that $(Z, \mathcal{C}, \lambda) \to (Y, \mathcal{B}, \nu)$ is a proper $K$-primitive extension. We will show that $Z$ also has the SZ property, thereby contradicting maximality of $Y$. 

Let \( A \in \mathcal{B} \) with \( \lambda(A) > 0 \) and \( \{S_i\}_{i=0}^t \subseteq F^{\otimes n} \cap \text{FE} \) be a finite set with \( S_0 \equiv 1_G \). We have to show
\[
(5.27) \quad \text{IP-lim}_{\bar{a} \in \mathcal{F}_n^0} \mu \left( \bigcap_{i=0}^t A S_i(\bar{a})^{-1} \right) > 0.
\]

By Lemma 5.7 we may wlog assume that \( K \) is an FVIP group and by Lemma 5.14 that AP is dense in \( L^2(\mathbb{Z}) \). Note that FE is still good with respect to the new IP-ring implied in the “wlog” notation. Thus wlog we have a \( K \)-mixing set \( \{W_k\}_{k=0}^\infty \) and polynomial expressions \( R_i \in K^{\otimes n} \cap \text{FE} \) with \( R_0 \equiv 1_G \) such that \( S_i = W_i R_i \).

By Lemma 5.12 we may replace \( A \) by a subset that has at least one half of its measure such that \( 1_A \in A \). There exist \( c = c(\lambda(A)) > 0 \) and a set \( B \in \mathcal{B} \) such that \( \nu(B) > c \) and \( \lambda_\nu(A) > c \) for every \( y \in B \). Pick \( 0 < \epsilon < \min(c/2, c^\gamma/(4(t + 1))) \).

By Corollary 3.8 there exist \( N, w \in \mathbb{N} \) and
\[
\{L_i, M_i\}_{i=1}^w \subseteq (K^{\otimes N} \cap \text{FE}) \times (F^{\otimes N} \cap \text{FE})
\]
such that for every \( l \)-coloring of \( \{L_i, M_i\} \) there exists a number \( a \) and sets \( \beta_1 < \cdots < \beta_m \subseteq N \) such that the set \( \{L_a R_i[\bar{\beta}], M_a W_k[\bar{\beta}]L_a^{-1} \}_{0 \leq i \leq t, 0 \leq k < v} \) is monochrome (and in particular contained in the set \( \{L_i, M_i\} \)).

Since \( f = 1_i \in A \) there exist functions \( g_1, \ldots, g_l \in L^2(\mathbb{Z}) \) and a set \( D \in \mathcal{B} \) such that \( \nu(D) < \epsilon \) and for every \( \delta > 0 \) and \( T \in K^{\otimes N} \cap \text{FE} \) there exists \( \alpha_0 \) such that for every \( \alpha_0 < \bar{a} \in \mathcal{F}_n^0 \) there exists a set \( E = E(\bar{a}) \in \mathcal{B} \) with \( \nu(E) < \delta \) such that for every \( y \in (D \cup E)^\delta \) there exists \( j \) such that \( ||T(\bar{a})f - g_j||_y < \epsilon \). Let \( B' = B \cap D^\delta \), so that \( \nu(B') > c/2 \).

Let \( Q = |\mathcal{F}(N)_c^m| \) be the number of possible choices of \( \bar{\beta} \in \mathcal{F}(N)_m \) and
\[
a_1 := a(B', N, \{1_G\} \cup \{M_i W_i[\bar{\beta}] \}_{1 \leq i \leq w, k < v, \bar{\beta} \in \mathcal{F}(N)_m^0}) > 0.
\]

Using this with \( \delta = a_1/2w^2 \) and \( T = L_1, \ldots, L_w \) we obtain wlog for every \( \bar{a} \in \mathcal{F}_n^0 \) a set \( E = E(\bar{a}) \in \mathcal{B} \) with \( \nu(E) < a_1/2w \) such that for every \( y \in (D \cup E)^\delta \) and every \( i = 1, \ldots, w \) there exists \( j = j(y, i) \) such that
\[
(5.28) \quad ||L_i(\bar{a})f - g_j||_y < \epsilon \quad \text{for every} \quad 1 \leq i \leq w.
\]

By Theorem 5.22 we may also wlog assume that for every \( \bar{a} \in \mathcal{F}_n^0 \) we have
\[
(5.29) \quad \left| \left| \mathbb{E}(\prod_{k<v} W_k(\bar{a}) f | Y) - \prod_{k<v} W_k(\bar{a}) \mathbb{E}(f | Y) \right| \right| < c^\gamma \left( A_1/2wQ \right)^{1/2}/4.
\]

Recall that we have to show (5.27). To this end it suffices to find \( a(A, m, \{S_i\}_{0 \leq i \leq t}) \) such that for an arbitrary sub-IP-ring there exists \( \bar{\gamma} \in \mathcal{F}_n^0 \) with
\[
\mu \left( \bigcap_{i=0}^t A S_i(\bar{\gamma})^{-1} \right) > a(A, m, \{S_i\}_{0 \leq i \leq t}) > 0,
\]
so fix a sub-IP-ring \( \mathcal{F} \). By definition of \( a_1 \) there exists a tuple \( \bar{a} \in \mathcal{F}_n^0 \) (that will remain fixed) such that
\[
\nu(C_0) > a_1, \quad \text{where} \quad C_0 := \bigcap_{i=1, \ldots, w, k<v, \bar{\beta}} B W_k[\bar{\beta}](\bar{a})^{-1} M(\bar{a})^{-1}.
\]

Let
\[
C := C_0 \setminus \bigcup_{i=1}^w E(\bar{a}) M(\bar{a})^{-1},
\]
so that \( \nu(C) > a_1/2 \). For every \( y \in C \) consider an \( l \)-coloring of \( \{L_i, M_i\}_i \) given by \( i \in [1, w] \mapsto j(y M_i(\bar{a}), i) \) determined by (5.28). By the assumptions on \( \{L_i, M_i\}_i \) there exist \( j(y), \alpha \in [1, w] \) and \( \beta_1 < \cdots < \beta_m \subseteq N \) such that
\[
||L_a(\bar{a})R_i[\bar{\beta}](\bar{a})f - g_{j(y)}||_{Y M(\bar{a}) W(\bar{a})^\delta} < \epsilon \quad \text{for every} \quad 0 \leq i \leq t, 0 \leq k < v.
\]

This can also be written as
\[
||W_k[\bar{\beta}](\bar{a})R_i[\bar{\beta}](\bar{a})f - W_k[\bar{\beta}](\bar{a})L_a(\bar{a})^{-1} g_{j(y)}||_{Y M(\bar{a})} < \epsilon \quad \text{for every} \quad 0 \leq i \leq t, 0 \leq k < v.
\]
Since this holds for every \( i, k \) and we have \( R_0 \equiv \mathbb{1}_G \), this implies
\[
\| (W_k R_i) [\vec{\beta}] (\vec{d}) f - W_k [\vec{\beta}] (\vec{d}) f \|_{y, M_k(\vec{d})} < 2\epsilon.
\]
Passing to a subset \( C' \subset C \) with measure at least \( a_1 / 2wQ \), we may assume that \( a \) and \( \vec{\beta} \) do not depend on \( y \). Thus we obtain a set \( B'' := C'M_a(\vec{d}) \) of measure at least \( a_1 / 2wQ \) and a tuple \( (\gamma_j = \cup_{i \in \beta} \alpha_j)_{j=1}^m \) such that
\[
\| W_y R_i (\vec{\gamma}) f - W_k (\vec{\gamma}) f \|_y < 2\epsilon
\]
for every \( y \in B'' \), \( i \) and \( k \). Recall that \( f \) is \( \{0,1\} \)-valued, so that
\[
\left\| \prod_{i=0}^t S_i (\vec{\gamma}) f - \prod_{k<y} W_k (\vec{\gamma}) f \right\|_y = \left\| \prod_{i=0}^t W_k R_i (\vec{\gamma}) f - \prod_{i=0}^t W_k (\vec{\gamma}) f \right\|_y < 2(t+1)\epsilon
\]
for all \( y \in B'' \). Moreover, since \( B'' \subset \cap_i B'W_j (\vec{\gamma})^{-1} \), one has
\[
\left| \prod_{k<y} W_k (\vec{\gamma}) E(f | Y)(y) \right| \geq c^y
\]
for every \( y \in B'' \). Therefore and by (5.29) we obtain
\[
\left\| \prod_{i=0}^t S_i (\vec{\gamma}) f \right\| \geq \left\| \prod_{i=0}^t S_i (\vec{\gamma}) f \right\|_{L^2(B''_y)}
\]
\[
> \left\| \prod_{k<y} W_k (\vec{\gamma}) f \right\|_{L^2(B''_y)} - 2(t+1)\epsilon v(B''_y)^{1/2}
\]
\[
\geq \left\| E(\prod_{k<y} W_k (\vec{\gamma}) f | Y) \right\|_{L^2(B''_y)} - 2(t+1)\epsilon v(B''_y)^{1/2}
\]
\[
\geq \left\| \prod_{k<y} W_k (\vec{\gamma}) E(f | Y) \right\|_{L^2(B''_y)} - \left\| E(\prod_{k<y} W_k (\vec{\gamma}) f | Y) - \prod_{k<y} W_k (\vec{\gamma}) E(f | Y) \right\|
\]
\[
- 2(t+1)\epsilon v(B''_y)^{1/2}
\]
\[
> c^y (a_1 / 2wQ)^{1/2} / 4 = : a(A, m, \{S_j\}_1^1)^{1/2}.
\]

**Good groups of polynomial expressions.** As we have already mentioned, good groups are just technical vehicles. The point is that we can perform all operations that we are interested in within a countable set of polynomial expressions, so that we can wlog assume the existence of all IP-limits that we encounter.

The only non-trivial property of good groups is the decomposition property. However, the following lemma essentially shows that it is always satisfied.

**Proposition 5.30.** Let \( K \leq F \) be a subgroup that is invariant under conjugation by constants, \( m \in \mathbb{N} \) and \( \{S_j\}_{j=0}^1 \subset F_{\leq m} \) be any finite set with \( S_0 \equiv \mathbb{1}_G \). Then there exists a set \( \{T_k\}_{k=0}^y \subset F_{\leq m} \) that is wlog K-mixing and decompositions \( S_i = T_k R_i \) such that \( R_i \in K_{\leq m} \).

**Proof.** We argue by induction on \( m \). The claim is trivial for \( m = 0 \). Assume that it holds for \( m \), we show its validity for \( m + 1 \). For brevity we write \( \vec{d} = (\alpha_1, \ldots, \alpha_m) \), \( \alpha = \alpha_{m+1} \).

Consider the maps \( \tilde{S}_i(\vec{d}) = S_i(\vec{d}, \emptyset) \). By the induction hypothesis there exists a set \( \{\tilde{T}_k\}_{k=0}^y \subset F_{\leq m} \) that is wlog K-mixing and decompositions \( \tilde{S}_i = \tilde{T}_k \tilde{R}_i \) such that \( \tilde{R}_i \in K_{\leq m} \). Then \( S_i(\vec{d}, \alpha) = W_i^\alpha(\alpha) \tilde{T}_k(\tilde{d}) \tilde{R}_i(\tilde{d}) \).

Let \( i < j \). By the Milliken-Taylor Theorem\(^{11}\) we may wlog assume that either \( (W_i^\alpha)^{-1} W_j^\alpha \notin K \) for all \( \vec{d} \in \mathcal{F}_m \) (in which case we do nothing) or \( (W_i^\alpha)^{-1} W_j^\alpha \in K \) for all \( \vec{d} \in \mathcal{F}_m \). In the
Moreover, each sub-IP-ring. The required sub-IP-ring is then obtained by a diagonal procedure, cf. Proposition 5.30 to obtain a countable descending chain consisting of all products \( W_i^a T_k^a(\alpha) \) that occur above. □

**Corollary 5.31.** Every finite subset of \( F^{\otimes \omega} \) is wlog contained in a good subgroup of \( \text{VIP}(G_\bullet)^{\otimes \omega} \).

**Proof.** Since \( F \) is a countable Noetherian group, it has at most countably many subgroups. Moreover, each \( \mathcal{F}_m \) is countable, and there are only countably many finite tuples in any countable set. Hence we can use Proposition 5.30 to obtain a countable descending chain of sub-IP-rings such that the decomposition property holds for each tuple for one of these sub-IP-rings. The required sub-IP-ring is then obtained by a diagonal procedure, cf. [FK85] Lemma 1.4. □

Thus the good group is not really relevant for our multiple recurrence theorem, which we can now formulate as follows.

**Theorem 5.32.** Let \( G \) be a nilpotent group and \( F \leq \text{VIP}(G_\bullet) \) an VIP group. Consider a right measure-preserving action of \( G \) on an arbitrary (not necessarily regular) probability space \((X, \mathcal{A}, \mu)\). Let \( S_0, \ldots, S_t \in F^{\otimes m} \) be arbitrary polynomial expressions and \( A \in \mathcal{A} \) with \( \mu(A) > 0 \). Then there exists a sub-IP-ring \( \mathcal{F} \subset \mathcal{F}_m \) such that

\[
\text{IP-lim}_{\mathcal{F}_m} \mu \left( \bigcap_{i=0}^t AS_i(\alpha)^{-1} \right) > 0.
\]

**Proof.** We can assume \( S_0 \equiv 1_G \). By Corollary 5.31 we may assume that \( S_0, \ldots, S_t \in FE \) for some good subgroup \( FE \leq \text{VIP}(G_\bullet)^{\otimes m} \). Then we can replace \( G \) by a countable group that is generated by the union of ranges of elements of \( FE \). Next, we can replace \( \mathcal{A} \) by a separable \( G \)-invariant \( \sigma \)-algebra generated by \( A \). Finally, we can assume that \( X \) is regular and apply Theorem 5.26. □

Theorem 1.1 follows from Theorem 5.32 and Lemma 4.29 with the filtration (2.27), \( d \) being the maximal degree of the generalized polynomials \( p_{i,j} \). By the Furstenberg correspondence principle we obtain the following combinatorial corollary.

**Corollary 5.33.** Let \( G \) be a finitely generated nilpotent group, \( T_1, \ldots, T_t \in G \) and \( p_{i,j} : \mathbb{Z}^m \to \mathbb{Z}_\ell \) \( i = 1, \ldots, t, j = 1, \ldots, s \) be admissible generalized polynomials. Then for every subset \( E \subset G \) with positive upper Banach density the set

\[
\left\{ \vec{n} \in \mathbb{Z}^m : \exists g \in G : \prod_{i=1}^t T_{i}^{p_{i,j}(\vec{n})} \in E, j = 1, \ldots, s \right\}
\]

is \( \text{VIP}^* \) in \( \mathbb{Z}^m \).

Observe that in Theorem 4.28 for (not necessarily admissible) generalized polynomials we can choose \( n \) from a finite set that only depends on the generalized polynomial. In view of this fact we have the following variant of Corollary 5.33 for generalized polynomials.

**Corollary 5.34.** Let \( G \) be a finitely generated nilpotent group, \( T_1, \ldots, T_t \in G \) and \( p_{i,j} : \mathbb{Z}^m \to \mathbb{Z}_\ell \) \( i = 1, \ldots, t, j = 1, \ldots, s \) be generalized polynomials. Then there exists finite sets \( S_j, j = 1, \ldots, s \), such that for every subset \( E \subset G \) with positive upper Banach density the set

\[
\left\{ \vec{n} \in \mathbb{Z}^m : \exists g \in G, \exists s_j \in S_j : gs_j \prod_{i=1}^t T_i^{p_{i,j}(\vec{n})} \in E, j = 1, \ldots, s \right\}
\]

is \( \text{VIP}^* \) in \( \mathbb{Z}^m \).
Since every member set of an idempotent ultrafilter contains an IP set this implies a multidimensional version of \cite{BM10} Theorem 1.23 that holds for every idempotent ultrafilter, see \cite{BM10, Remark 3.42}.

**References**

\cite{BBH94} V. Bergelson, A. Blass, and N. Hindman, *Partition theorems for spaces of variable words*, Proc. London Math. Soc. (3) **68** (1994), no. 3, 449–476. MR1262304 (95i:05107) 15

\cite{Ber87} V. Bergelson, *Weakly mixing PET*, Ergodic Theory Dynam. Systems **7** (1987), no. 3, 337–349. MR0912373 (89g:28022) 18

\cite{BFM96} V. Bergelson, H. Furstenberg, and R. McCutcheon, *IP-sets and polynomial recurrence*, Ergodic Theory Dynam. Systems **16** (1996), no. 5, 963–974. MR1417769 (98c:28024) 11, 12, 14, 15

\cite{BHKM06} V. Bergelson, I. J. Håland Knutson, and R. McCutcheon, *IP-systems, generalized polynomials and recurrence*, Ergodic Theory Dynam. Systems **26** (2006), no. 4, 999–1019. MR2245589 (2008d:37001) 11, 12, 14, 18, 19, 20

\cite{BL03} V Bergelson and A. Leibman, *Topological multiple recurrence for polynomial configurations in nilpotent groups*, Adv. Math. **175** (2003), no. 2, 271–296. MR1972634 (2004d:54036) 17, 8, 9

\cite{BL07} V Bergelson and A. Leibman, *Distribution of values of bounded generalized polynomials*, Acta Math. **198** (2007), no. 2, 155–230. MR2318563 (2008m:11149) 20

\cite{BL96} V Bergelson and A. Leibman, *Polynomial extensions of van der Waerden’s and Szemerédi’s theorems*, J. Anal. Math. **96** (1996), no. 3, 725–753. MR1325795 (96j:11013) 1

\cite{BL99} _____, *Set-polynomials and polynomial extension of the Hales-Jewett theorem*, Ann. of Math. (2) **150** (1999), no. 1, 33–75, available at math/9907201. MR1715320 (2000h:05026) 18, 9, 10

\cite{BM00} V Bergelson and R. McCutcheon, *An ergodic IP polynomial Szemerédi theorem*, Mem. Amer. Math. Soc. **146** (2000), no. 695, viii+106. MR1692634 (2000m:28018) 11, 21, 22, 23, 24, 26, 27

\cite{BM10} V Bergelson and R. McCutcheon, *Idempotent ultrafilters, multiple weak mixing and Szemerédi’s theorem for generalized polynomials*, J. Anal. Math. **111** (2010), 77–130. MR2747662 (2012b:11117) 20, 31

\cite{FK85} H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for IP-systems and combinatorial theory*, J. Analyse Math. **45** (1985), 117–168. MR833409 (87m:05007) 11, 13, 15, 19, 26, 30

\cite{FK91} _____, *A density version of the Hales-Jewett theorem*, J. Anal. Math. **57** (1991), 64–119. MR1191743 (94f:28020) 11

\cite{Pur77} H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256. MR0498471 (58 #16583) 11

\cite{Pur81} H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures. MR603625 (82f:28010) 12, 21

\cite{FW78} H. Furstenberg and B. Weiss, *Topological dynamics and combinatorial number theory*, J. Analyse Math. **34** (1978), 61–85 (1979). MR531271 (80g:05009) 11

\cite{GT12} B. Green and T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), no. 2, 465–540, available at 0709.3562. MR2877055 12

\cite{GTZ12} B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers \(U^k\)-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372, available at math/1006.1372. MR2950773 12

\cite{Hin74} N. Hindman, *Finite sums from sequences within cells of a partition of \(N\)*, J. Combinatorial Theory Ser. A **17** (1974), 1–11. MR0349574 (50 #2067) 113

\cite{Lei02} A. Leibman, *Polynomial mappings of groups*, Israel J. Math. **129** (2002), 29–60. with erratum. MR1910931 (2003g:20050) 11, 2, 4

\cite{Lei98} _____, *Multiple recurrence theorem for measure preserving actions of a nilpotent group*, Geom. Funct. Anal. **8** (1998), no. 5, 853–931. MR1650102 (99k:28019) 11

\cite{McC05} R. McCutcheon, *FVIP systems and multiple recurrence*, Israel J. Math. **146** (2005), 157–188. MR2151599 (2006b:37015) 11, 22

\cite{Mil75} K. R. Milliken, *Ramsey's theorem with sums or unions*, J. Combinatorial Theory Ser. A **18** (1975), 276–290. MR0379906 (51 #10106) 14

\cite{MKS66} W. Magnus, A. Karrass, and D. Solitar, *Combinatorial group theory: Presentations of groups in terms of generators and relations*, Interscience Publishers [John Wiley & Sons, Inc.], New York-London-Sydney, 1966. MR0207802 (34 #7617) 13

\cite{Rot95} J. J. Rotman, *An introduction to the theory of groups*, Fourth, Graduate Texts in Mathematics, vol. 148, Springer-Verlag, New York, 1995. MR1307623 (95m:20001) 12
[Sze75] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 199–245. Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR0369312 (51 #5547)↑1

[Tay76] A. D. Taylor, *A canonical partition relation for finite subsets of ω*, J. Combinatorial Theory Ser. A 21 (1976), no. 2, 137–146. MR0424571 (54 #12530)↑14

EINSTEIN INSTITUTE OF MATHEMATICS, EDMOND J. SAFRA CAMPUS, GIVAT RAM, THE HEBREW UNIVERSITY OF JERUSALEM, JERUSALEM, 91904, ISRAEL

E-mail address: pzorin@math.huji.ac.il

URL: http://math.huji.ac.il/~pzorin/