Polynomial Szemerédi theorems for countable modules over integral domains and finite fields

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

A. Leibman
Dept. of Mathematics
The Ohio State University
Columbus, OH 43210, USA
leibman@math.ohio-state.edu

R. McCutcheon
Dept. of Mathematical Sciences
University of Memphis
Memphis, TN 38152
rmcctchn@memphis.edu

Abstract

Given a pair of vector spaces V and W over a countable field F and a probability space X, one defines a *polynomial measure preserving action* of V on X to be a composition $T \circ \varphi$, where $\varphi\colon V \longrightarrow W$ is a polynomial mapping and T is a measure preserving action of W on X. We show that the known structure theory of measure preserving group actions extends to polynomial actions and establish a Furstenberg-style multiple recurrence theorem for such actions. Among the combinatorial corollaries of this result are a polynomial Szemerédi theorem for sets of positive density in finite dimensional modules over integral domains as well as the following fact: *Let $\mathcal{P}$ be a finite family of polynomials with integer coefficients and zero constant term. For any $\alpha > 0$ there exists $N \in \mathbb{N}$ such that whenever F is a field with $|F| \geq N$ and $E \subseteq F$ with $|E|/|F| \geq \alpha$, there exist $u \in F$, $u \neq 0$, and $w \in E$ such that $w + \varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

0. Introduction

Many familiar theorems of combinatorics and number theory establish combinatorial and/or arithmetic richness of large sets in groups or rings. For example, Szemerédi's theorem ([Sz]) states that any set E of natural numbers having positive upper density $\bar{d}(E) = \limsup_{N\to\infty} \frac{|E\cap\{1,\ldots,N\}|}{N} > 0$ contains arbitrarily long arithmetic progressions. An equivalent formulation, more geometric in spirit, says that if a set $E \subseteq \mathbb{N}$ satisifies $\bar{d}(E) > 0$ then for any finite set $S \subset \mathbb{Z}$ there exist $x \in E$ and $n \in \mathbb{N}$ such that $x + nS = \{x + ns : s \in S\} \subset E$. In other words, sets of positive upper density in $\mathbb{N}$ contain homothetic images of every finite set of integers. An ergodic-theoretic proof of Szemerédi's theorem given by Furstenberg in [F1] has connected density combinatorics with the phenomenon of multiple recurrence in ergodic theory and has led to powerful new results in this vein for which no non-ergodic proofs have been offered. (The original proof of Szemerédi's theorem, by contrast, is purely combinatorial, and Gowers in [G] provides yet another non-ergodic proof, with very good bounds, proceeding via harmonic analysis.) In order to formulate some results of this type, we introduce the following more general notion of upper density: if G is a countable abelian group and $E \subset G$, the *upper Banach density* of E is given by $d^*(E) = \sup_{\{\Phi_n\}} \lim_{n\to\infty} \frac{|E\cap\Phi_n|}{|\Phi_n|}$, where the supremum here is taken over all

The first two authors are supported by NSF, grant DMS-0070566. The second author is supported by the A. Sloan Foundation. The third author is supported by NSF, grant DMS-0200700.

Følner sequences for G. Now, our first example of a result having as yet no non-ergodic proof is a multidimensional version of Szemerédi's theorem established by Furstenberg and Katznelson in [FK1], which states that if $d \in \mathbb{N}$ and $E \subseteq \mathbb{Z}^d$ with $d^*(E) > 0$ then E contains homothetic images of every finite set $S \subset \mathbb{Z}^d$. Our second such result is an extension of the first: a polynomial Szemerédi theorem ([BL1]) which says, roughly, that for any polynomial mapping $P: \mathbb{Z}^d \longrightarrow \mathbb{Z}^c$, $d, c \in \mathbb{N}$, it is the case that every subset $E \subseteq \mathbb{Z}^c$ with $d^*(E) > 0$ contains "homothetic affine P-images" of all finite subsets of $\mathbb{Z}^d$:

Theorem PSZ. *Let $P: \mathbb{Z}^d \longrightarrow \mathbb{Z}^c$ be a polynomial mapping with $P(0) = 0$ and suppose $E \subset \mathbb{Z}^c$ with $d^*(E) > 0$. For any finite set $S \in \mathbb{Z}^d$ there exist $u \in \mathbb{Z}^c$ and $n \in \mathbb{Z}$, $n \neq 0$, such that $u + P(nS) = \{u + P(ns) : s \in S\} \subset E$.*

From Theorem PSZ one can derive the following *finitary* version (which is easily shown to be equivalent to Theorem PSZ):

Theorem PSZf. *Let P be a polynomial mapping $\mathbb{Z}^d \longrightarrow \mathbb{Z}^c$ with $P(0) = 0$ and suppose S is a finite subset of $\mathbb{Z}^d$. For any $\alpha > 0$ there exists N such that whenever $m > N$ and E is a subset of $\{1, \ldots, m\}^c$ with $|E|/m^c > \alpha$, there exist $u \in \mathbb{Z}^c$ and $n \in \mathbb{Z}$, $n \neq 0$, such that $u + P(nS) \subset E$.*

It is natural to inquire whether Theorem PSZ is a manifestation of a phenomenon pertaining to polynomial mappings of more general structures. In particular, one would like to know whether statements analogous to Theorems PSZ and PSZf hold for modules over arbitrary rings. Another natural question is whether there exists a result about polynomial mappings of finite fields which would be analogous to Theorem PSZf. A strong indication that one might expect affirmative answers to these questions is provided by the validity of a polynomial Hales-Jewett theorem ([BL2]), from which one can derive *partition* results of this kind.

The goal of this paper is to show that the answers to the questions raised above are positive in certain cases. For example the following two theorems appear below as Theorems 5.12 and 5.18.

Theorem. *Let K be a countable integral domain, let M be a finitely generated K-module and let φ be a polynomial $K^d \longrightarrow M$ with $\varphi(0) = 0$. For any finite set $S \subset K^d$ and any $E \subseteq M$ with $d^*(E) > 0$ there exist $u \in K$, $u \neq 0$, and $w \in E$ such that $w + \varphi(uS) \subset E$.*

Theorem. *Let $\mathcal{P}$ be a finite family of integer polynomials with zero constant term. There exists $N \in \mathbb{N}$ such that whenever F is a finite field with $|F| \geq N$ and $E \subseteq F$ with $|E| \geq \alpha|F|$, there exist $u \in F$, $u \neq 0$, and $w \in E$ such that $w + \varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

To achieve our goal we use the ergodic method pioneered in Furstenberg's seminal paper [F1], which allows one to reduce the original combinatorial problem to that of demonstrating a type of multiple recurrence in measure preserving systems. However, in order to obtain the results in their proper generality one has to circumvent some obstacles which were not present in the more special situations treated in [FK1] and [BL1]. The first major problem is related to the fact that the structures we deal with (integral domains, e.g. $\mathbb{Z}[x]$, countable fields, etc.) may be infinitely generated as abelian groups. While this would not a priori be an issue if we considered only linear expressions, it becomes relevant when polynomial expressions φ are considered, due to the fact that we are forced to work with "derivatives" of φ having the form $\varphi_v(u) = \varphi(u+v) - \varphi(u) - \varphi(v)$, where v can be any element of the infinitely generated group under consideration. (Note these are identically zero when φ is linear.) The issue, then, is not so much infinite generation of the v's as infinite generation of the φ_v's that come into play, despite the fact that we start with only finitely many φ's. Such complications do not arise when the underlying group is finitely generated, nor do they arise when the expressions involved are linear. The difficulties are in fact confined to a single portion of the proof, and this prtion is unique to the ergodic setup; namely the existence of so-called *primitive extensions*.

The difficulties arising when polynomials are mixed with infinite generation are not new. We deal here with the most common mode of convergence in ergodic theory, namely that based on Cesáro averaging, but there is another type of convergence, so-called IP-convergence, that may be employed as a potential alternative strategy (see [FK2], [FK3], [BFM], and [BM]). Ergodic theorems phrased in the language of IP-convergence have a familiar look: certain sequences of unitary operators have limits that turn out to be orthogonal projections. These ergodic theorems form the core of an IP-structure theory on which are based the proofs of IP Szemerédi-type results. As of now, the existing lines of argument seem not to work when the families of "derivatives" of the IP expressions dealt with are infinitely generated. Indeed, a counter-example in [BFM] shows that, even in the most basic case (pertaining to single recurrence), known polynomial IP ergodic theorems do not carry over to the infinitely generated situation. Limits of certain IP-like expressions having a polynomial nature need not necessarily be orthogonal projections, and it is therefore unclear how to establish even single recurrence for the corresponding systems of measure preserving transformations. One would not therefore automatically expect it to be easy to overcome infinite generation in our current setup either, but as it turns out, certain algebraic properties of fields carry the day.

Let us elaborate. Suppose $\mathcal{P}$ is a finite set of polynomial mappings from a finite dimensional vector space V to a vector space W over an infinite field F, and let U be a measure preserving action of W on a probability space X. In an approach analogous to that of [FK1], we reduce the general problem of multiple recurrence for $\mathcal{P}$ in X to the situation where, for each $\varphi \in \mathcal{P}$, the polynomial action $U(\varphi)$ of V on X is either (relatively) compact or (relatively) weakly mixing. To do so, we need to choose a maximal subgroup in W which acts in a compact way on a nontrivial subspace $L^2(\hat{X})$ of $L^2(X)$, and then replace X by $\hat{X}$. However, if W is not finitely generated, such a subgroup does not have to exist. We note however that the group $\langle\varphi(V)\rangle$ generated in W by the elements $\varphi(v)$, $v \in V$, is not arbitrary: if $\operatorname{char} F = 0$, $\langle\varphi(V)\rangle$ is a vector subspace of W (Theorem 1.13); if $\operatorname{char} F = p > 0$, then $\langle\varphi(V)\rangle$ is, in our terminology, a *p-subspace of W* (Theorem 1.21 and section 1.25). We show that, like conventional subspaces of a finite dimensional vector space, p-subspaces satisfy the ascending chain condition (Lemma 1.28), which solves the problem.

A second issue that does not arise in [FK1] or [BL1] is that our multiple recurrence theorems are required to be *uniform* in the sense that they must hold along arbitrary Følner sequences. A familiar instance of this phenomenon is the classical von Neumann ergodic theorem, which can be given a "uniform" formulation: for any Følner sequence in

$\mathbb{Z}$, in particular, for any sequence of intervals $\Phi_k = [N_k, M_k]$ with $M_k - N_k \longrightarrow \infty$, and for any unitary operator U on a Hilbert space $\mathcal{H}$, $\lim\limits_{k\to\infty} \frac{1}{|\Phi_k|}\sum_{n=N_k}^{M_k} U^n f$ exists in the strong topology for any $f \in \mathcal{H}$, and equals the orthogonal projection of f onto the subspace of U-invariant vectors. The combinatorial significance of the fact that our multiple recurrence theorems hold along arbitrary Følner sequences is that it insures that the "good" parameters of the sought-after configurations are more plentiful than would otherwise be known. For example, one would like to know not only that for any set $E \subseteq \mathbb{Q}^d$ with $d^*(E) > 0$, any polynomial mapping $P: \mathbb{Q}^d \longrightarrow \mathbb{Q}^c$ satisfying $P(0) = 0$ and any finite set $S \subset \mathbb{Q}^d$ one can find $w \in \mathbb{Q}^c$ and $r \in \mathbb{Q}$, $r \neq 0$, such that $S_{w,r} = w + P(rS) = \{w + P(rx),\ x \in S\} \subset E$ (this follows from Theorem PSZ), but that there are many such images. In particular, one would like to know that $\{r : \text{there exists } w \text{ such that } S_{w,r} \subset E\}$ is a syndetic (or relatively dense) set – a fact which follows from the special case $K = \mathbb{Q}$ of Theorem 5.2 below. (A subset S of a countable abelian group G is syndetic if finitely many shifts of it cover G. See section 2.6 below.)

Uniformity is achieved by using (in place of a polynomial van der Waerden result, as in [BL1]) a polynomial extension of the Hales-Jewett coloring theorem obtained in [BL2]. This method, which was originally inspired by a judicious and ingenious use of the (linear) Hales-Jewett theorem in [FK2], was previously employed in both [BM] and [Le]. That we are able to use the technique again here is a consequence of the fact that it is impervious to infinite generation of the underlying group.

Finally, one more difficulty which demanded the introduction of new and the sharpening of old techniques is that our goal is to establish a general result holding for fields of both finite and infinite characteristic. In an attempt to streamline our proofs, we have attempted, where possible, to unify both cases in a single line of argument. See section 3.8 for an example of this unification effort.

In most of the paper we deal with polynomial actions of finite dimensional vector spaces over countable fields. In particular, we prove our polynomial multiple recurrence theorem in this setup. We use then a combinatorial argument (see Chapter 5) to derive from it a more general recurrence theorem pertaining to finitely generated modules over countable integral domains. An advantage of this approach is that, unlike in the case of general polynomial actions of modules over integral domains (unlike the most general IP polynomial case as well), for polynomial actions of fields one has a polynomial ergodic theorem (due to P. Larick in [La]) showing that the limits of ergodic averages in polynomial von Neumann-type theorems do in fact turn out to be orthogonal projections. This makes the "polynomial" situation more closely resemble the "linear" one and facilitates handling of some delicate points in the proof. A modest extension (Theorem 3.10) of Larick's theorem is employed in the development of the structure theory we utilize in our proof of Theorem 4.14, which is the main ergodic-theoretic result of this paper and which we now state.

Theorem. *Let V, W be finite dimensional vector spaces over F, let U be a measure preserving action of W on a probability measure space $(X, \mathcal{B}, \mu)$ and let $\mathcal{P}$ be a finite family of polynomials $V \longrightarrow W$ with zero constant term. Then for any $B \in \mathcal{B}$ with $\mu(B) > 0$ there exists $c > 0$ such that the set $\{u \in V : \mu\big(\bigcap_{\varphi\in\mathcal{P}} U(\varphi(u))B\big) > c\}$ is syndetic in V.*

The structure of the paper is as follows. In Chapter 1 we investigate algebraic properties of polynomial mappings of finite dimensional vector spaces over infinite fields. In particular we obtain results, crucial to our work in Chapter 3, pertaining to spans of the images of such polynomial mappings and their derivatives. Chapter 2 is devoted to various properties of densities in countable abelian groups, as well as to establishing a general version of van der Corput's difference lemma – a major tool which allows one to handle the weakly mixing portion of the proof of our main theorem by making it possible to inductively reduce the complexity of the families of polynomial expressions involved. In Chapter 3 the algebraic apparatus developed in Chapter 1 and the van der Corput lemma are used to establish general ergodic theorems for unitary actions of finite dimensional vector space on Hilbert spaces and Hilbert bundles. (Actually, instead of dealing with Hilbert bundles we choose an equivalent approach and work with Hilbert-like spaces whose inner products take values in the space of integrable functions on a probability space.) In Chapter 4 we use the previously-mentioned fact that our p-subspaces satisfy the ascending chain condition to establish existence of primitive extensions. The polynomial Hales-Jewett theorem is then combined with the knowledge gained in Chapter 3 (in the special case of polynomial unitary actions originating in measure preserving actions of vector spaces), to establish our main theorem. Finally in Chapter 5 we derive combinatorial corollaries of Theorem 4.14, including the polynomial Szemerédi theorem for finite fields cited above.

1. Polynomial mappings of vector spaces

Let F be an infinite field.

1.1. A *monomial* (over F, in d variables) is a mapping $\psi\colon F^d \longrightarrow F$, $d \in \mathbb{N}$, of the form $\psi(x_1,\ldots,x_d) = x_1^{n_1}\ldots x_d^{n_d}$ with $n_1,\ldots,n_d \in \mathbb{Z}_+$ (where $\mathbb{Z}_+ = \{0,1,2,\ldots\}$). The *formal degree* of ψ is defined by $\deg\psi = n_1+\ldots+n_d$. A mapping $\varphi\colon F^d \longrightarrow W$, where $d \in \mathbb{N}$ and W is a vector space over F, will be called *a polynomial* if it is representable as a linear combination of monomials with vector coefficients: for $\mathbf{x} = (x_1,\ldots,x_d) \in F^d$, $\varphi(\mathbf{x}) = \psi_1(\mathbf{x})a_1 + \ldots + \psi_l(\mathbf{x})a_l$, where $\psi_1,\ldots,\psi_l$ are monomials $F^d \longrightarrow F$ and $a_1,\ldots,a_l \in W$. Under the assumption that φ is in reduced form, i.e. the ψ_j are distinct and the coefficients a_j are nonzero, the formal degree of φ is defined by $\deg\varphi = \max\{\deg\psi_1,\ldots,\deg\psi_l\}$.

1.2. Given a finite dimensional vector space V over F, we say that a mapping $\varphi\colon V \longrightarrow W$ is a polynomial if it becomes a polynomial after introducing coordinates in V. Clearly, the choice of coordinate system affects neither the polynomiality of φ nor its formal degree.

We are interested in the subgroup of W generated by the range $\varphi(V)$ of a polynomial $\varphi\colon V \longrightarrow W$. We first discuss the properties of general "polynomial mappings" of abelian groups.

1.3. Let G and G' be abelian groups, written additively. Given a mapping $\varphi\colon G \longrightarrow G'$, and $v \in G$, we define *the derivative of φ with step v*, $D_v\varphi\colon G \longrightarrow G'$, by $D_v\varphi(u) = \varphi(u+v) - \varphi(u) - \varphi(v) + \varphi(0)$. (The "non-standard" terms $-\varphi(v)+\varphi(0)$ are present in order to ensure that $D_v\varphi(0) = 0$, which simplifies our further computations.) φ is called *a polynomial mapping* if there exists $n \in \mathbb{N}$ such that $D_{v_n}\ldots D_{v_1}\varphi$ is constant for any

$v_1, \ldots, v_n \in G$. The minimal n with this property is called the *degree* of φ; in order to avoid confusion with the formal degree defined above, we will denote it by $\operatorname{Deg}\varphi$. The reader will notice that we have reserved the term "degree" for a notion more customarily applied to what we refer to as the "formal degree". We do this because for our purposes Deg is a more relevant notion than deg.

For a constant φ we put $\operatorname{Deg}\varphi = 0$. If $\operatorname{Deg}\varphi \le 1$ we say that φ is *linear*; it is clear that φ is linear if and only if it is a sum of a homomorphism and a constant mapping.

1.4. The proof of the following lemma is straightforward.

Lemma. (i) *If* $\varphi_1, \varphi_2 \colon G \longrightarrow G'$ *are polynomial mappings then* $\varphi_1 + \varphi_2$ *is a polynomial mapping with* $\operatorname{Deg}(\varphi_1 + \varphi_2) \le \max\{\operatorname{Deg}\varphi_1, \operatorname{Deg}\varphi_2\}$.
(ii) *If* $\varphi \colon G \longrightarrow G'$ *is a polynomial mapping and* $\eta \colon G' \longrightarrow G''$ *is a homomorphism, then* $\eta \circ \varphi$ *is a polynomial mapping with* $\operatorname{Deg}(\eta \circ \varphi) \le \operatorname{Deg}\varphi$.
(iii) *If* $\varphi \colon G \longrightarrow G'$ *is a polynomial mapping and* $\eta \colon G'' \longrightarrow G$ *is a homomorphism, then* $\varphi \circ \eta$ *is a polynomial mapping with* $\operatorname{Deg}(\varphi \circ \eta) \le \operatorname{Deg}\varphi$.
(iv) *If* $\varphi_j \colon G \longrightarrow G_j$, $j = 1, \ldots, k$, *are polynomial mappings and* $\eta \colon G_1 \times \ldots \times G_k \longrightarrow G'$ *is a* k*-linear mapping, then* $\eta \circ (\varphi_1, \ldots, \varphi_k)$ *is a polynomial mapping with* $\operatorname{Deg}\big(\eta \circ (\varphi_1, \ldots, \varphi_k)\big) \le \operatorname{Deg}\varphi_1 + \ldots + \operatorname{Deg}\varphi_k$.

(All these are special cases of the following fact: the composition of polynomial mappings is polynomial and $\operatorname{Deg}(\varphi_1 \circ \varphi_2) \le (\operatorname{Deg}\varphi_1)(\operatorname{Deg}\varphi_2)$.)

1.5. We now return to conventional polynomials $F^d \longrightarrow W$. A monomial $\psi(\mathbf{x}) = x_1^{n_1} \ldots x_d^{n_d}$, $\mathbf{x} = (x_1, \ldots, x_d) \in F^d$, is a product of $n_1 + \cdots + n_d$ linear mappings of the form $\mathbf{x} \mapsto x_i$. Hence, by Lemma 1.4(iv), ψ is a polynomial mapping with $\operatorname{Deg}\psi \le n_1 + \ldots + n_d$.

By Lemma 1.4(i), any polynomial $\varphi = \psi_1 a_1 + \ldots + \psi_l a_l$ is a polynomial mapping with $\operatorname{Deg}\varphi \le \max\{\operatorname{Deg}\psi_1, \ldots, \operatorname{Deg}\psi_l\}$. We will see that when F has finite characteristic $\operatorname{Deg}(\varphi)$ may not coincide with $\deg(\varphi)$.

1.6. Let us undertake exact computations. For $\psi(\mathbf{x}) = x_1^{n_1} \ldots x_d^{n_d}$ and $\mathbf{y} = (y_1, \ldots, y_d) \in F^d$ we have

$$D_{\mathbf{y}}\psi(\mathbf{x}) = \sum_{\substack{(m_1,\ldots,m_d) \le (n_1,\ldots,n_d) \\ (m_1,\ldots,m_d) \ne (0,\ldots,0) \\ (m_1,\ldots,m_d) \ne (n_1,\ldots,n_d)}} \binom{n_1}{m_1} \ldots \binom{n_d}{m_d} x_1^{m_1} y_1^{n_1 - m_1} \ldots x_d^{m_d} y_d^{n_d - m_d}, \tag{1.1}$$

where we write $(m_1, \ldots, m_d) \le (n_1, \ldots, n_d)$ if $m_i \le n_i$ for all $i = 1, \ldots, d$. If F has zero characteristic, the coefficients of all the monomials in (1.1) are distinct monomials in $\mathbf{y}$ with nonzero numerical coefficients. For any two distinct monomials ψ_1 and ψ_2, any two equal monomials in $D_{\mathbf{y}}\psi_1$ and $D_{\mathbf{y}}\psi_2$ have, as their coefficients, distinct monomials in $\mathbf{y}$. Therefore, for any polynomial $\varphi = \psi_1 a_1 + \ldots + \psi_l a_l$, where $\psi_1, \ldots, \psi_l$ are distinct monomials, the polynomial $D_{\mathbf{y}}\varphi$ contains all the monomials occurring in the derivatives $D_{\mathbf{y}}\psi_1$, ..., $D_{\mathbf{y}}\psi_s$ with coefficients which are nonconstant polynomials in $\mathbf{y}$. By induction we may conclude that $\operatorname{Deg}\varphi = \max\{\operatorname{Deg}\psi_1, \ldots, \operatorname{Deg}\psi_s\}$ and that $\operatorname{Deg}(x_1^{n_1} \ldots x_d^{n_d}) = n_1 + \ldots + n_d$, that is, $\operatorname{Deg}\varphi = \deg\varphi$.

1.7. If char $F = p < \infty$, the situation is more complicated. The mappings $x \mapsto x^{p^k}$, $k \in \mathbb{Z}_+$, are homomorphisms and hence linear polynomials. Let us consider a monomial $\psi\colon F \longrightarrow F$, $\psi(x) = x^n$. Let $n = n_0 + n_1 p + \ldots + n_l p^l$ with $n_0, n_1, \ldots, n_l \in \{0, 1, \ldots, p-1\}$, that is, let $(n_l, \ldots, n_1, n_0)$ be the base p expansion of n. Then $x^n = x^{n_0}(x^p)^{n_1} \ldots (x^{p^l})^{n_l}$. For any $k = 0, \ldots, l$, $\mathrm{Deg}(x^{n_k}) \le n_k$ and, since $x \mapsto x^{p^k}$ is a homomorphism, by Lemma 1.4(iii) we have $\mathrm{Deg}\big((x^{p^k})^{n_k}\big) \le n_k$. Let us denote $n_0+n_1+\ldots+n_l$ by $\mathcal{N}_p(n)$. Then by Lemma 1.4(iv), $\mathrm{Deg}(x^n) \le \mathcal{N}_p(n)$.

In order to show that actually $\mathrm{Deg}(x^n) = \mathcal{N}_p(n)$, we use induction on $\mathcal{N}_p(n)$. Let us write

$$\begin{aligned}
&D_y(x^n) = (x+y)^n - x^n - y^n \\
&= (x+y)^{n_0}(x^p+y^p)^{n_1} \ldots (x^{p^l}+y^{p^l})^{n_l} - x^{n_0}(x^p)^{n_1} \ldots (x^{p^l})^{n_l} - y^{n_0}(y^p)^{n_1} \ldots (y^{p^l})^{n_l} \\
&= \Big(\sum_{m_0=0}^{n_0} \tbinom{n_0}{m_0} x^{m_0} y^{n_0-m_0}\Big)\Big(\sum_{m_1=0}^{n_1} \tbinom{n_1}{m_1} x^{m_1 p} y^{(n_1-m_1)p}\Big) \ldots \Big(\sum_{m_l=0}^{n_l} \tbinom{n_l}{m_l} x^{m_l p^l} y^{(n_l-m_l)p}\Big) \\
&\qquad\qquad - x^{n_0}(x^p)^{n_1} \ldots (x^{p^l})^{n_l} - y^{n_0}(y^p)^{n_1} \ldots (y^{p^l})^{n_l} \\
&= \sum_{\substack{(m_0,m_1,\ldots,m_l) \le (n_0,n_1,\ldots,n_l) \\ (m_0,m_1,\ldots,m_l) \ne (0,\ldots,0) \\ (m_0,m_1,\ldots,m_l) \ne (n_0,n_1,\ldots,n_l)}} \tbinom{n_0}{m_0}\tbinom{n_1}{m_1} \ldots \tbinom{n_l}{m_l} x^{m_0+m_1 p+\ldots+m_l p^l} y^{(n_0-m_0)+(n_1-m_1)p+\ldots+(n_l-m_l)p^l}.
\end{aligned}$$

On the other hand, $D_y(x^n) = \sum_{m=1}^{n-1} \binom{n}{m} x^m y^{n-m}$, so we may write

$$D_y(x^n) = \sum_{\substack{m \preceq n \\ m \ne 0 \\ m \ne n}} \tbinom{n}{m} x^m y^{n-m},$$

where all the coefficients are nonzero, and where we write $m \preceq n$ if all the digits in the p-adic expansion of m are less or equal to the corresponding digits of n. In particular, $D_y(x^n)$ has a summand of the form cx^m, where $\mathcal{N}_p(m) = \mathcal{N}_p(n) - 1$. Hence $\mathrm{Deg}(x^n) \ge \mathrm{Deg}(x^m) + 1 = \mathcal{N}_p(m) + 1 = \mathcal{N}_p(n)$.

For a general monomial $\psi(\mathbf{x}) = x_1^{n_1} \ldots x_d^{n_d}$ we have

$$D_{\mathbf{y}}\psi(\mathbf{x}) = \sum_{\substack{(m_1,\ldots,m_d) \preceq (n_1,\ldots,n_d) \\ (m_1,\ldots,m_d) \ne (0,\ldots,0) \\ (m_1,\ldots,m_d) \ne (n_1,\ldots,n_d)}} \tbinom{n_1}{m_1} \ldots \tbinom{n_d}{m_d} x_1^{m_1} y_1^{n_1-m_1} \ldots x_d^{m_d} y_d^{n_d-m_d}, \tag{1.2}$$

where all the coefficients are nonzero, and where we write $(m_1, \ldots, m_d) \preceq (n_1, \ldots, n_d)$ if $m_i \preceq n_i$ for all $i = 1, \ldots, d$. Arguing as before, we get $\mathrm{Deg}(x_1^{n_1} \ldots x_d^{n_d}) = \mathcal{N}_p(n_1) + \ldots + \mathcal{N}_p(n_d)$. As a corollary we obtain the fact that the only nontrivial linear monomials are those of the form $x_i^{p^k}$, and that a polynomial is linear if and only if it is a linear combination of linear monomials.

1.8. Let $\varphi: V \longrightarrow W$ be a polynomial, $\varphi = \psi_1 a_1 + \ldots + \psi_l a_l$, where $\psi_1, \ldots, \psi_t$ are linear monomials and $\psi_{t+1}, \ldots, \psi_l$ are non-linear. We will call the polynomial $\lambda = \psi_1 a_1 + \ldots + \psi_t a_t$ the *linear part* of φ. (One may check that λ does not depend on the choice of coordinates in V.) We say that φ has *trivial linear part* if $\lambda = 0$. If λ is the linear part of φ then $D_v\lambda = 0$ and $D_v(\varphi - \lambda) = D_v\varphi$ for any $v \in V$.

In the case $\varphi(0) = 0$ we will also say that φ has *zero constant term.*

1.9. Let us say that a monomial $\psi: F^d \longrightarrow W$ is *separable* if ψ is not representable as the p-power of another monomial: $\psi \neq \psi_0^p$. In other words, a monomial ψ, $\psi(\mathbf{x}) = x_1^{n_1} \ldots x_d^{n_d}$, is separable if and only if not all of $n_1, \ldots, n_d$ are multiples of p. It is clear that any monomial ψ is uniquely representable in the form $\psi = \psi_0^{p^k}$ where $k \in \mathbb{Z}_+$ and ψ_0 is a separable monomial. It is also clear that any polynomial $\varphi: F^d \longrightarrow W$ is uniquely representable in the form $\varphi = \eta_1 \circ \psi_1 + \ldots + \eta_l \circ \psi_l$, where $\psi_1, \ldots, \psi_l$ are distinct separable monomials and $\eta_1, \ldots, \eta_l$ are homomorphisms $F \longrightarrow W$: $\eta_j(x) = x a_{j,0} + x^p a_{j,1} + \ldots + x^{p^{k_j}} a_{j,k_j}$. We adopt this representation as canonical.

1.10. When W is a vector space over F and A is a subset of W, we will denote by $\mathrm{Span}(A)$ the subgroup of W generated by A: $\mathrm{Span}(A) = \bigl\{\pm a_1 \pm \ldots \pm a_k : k \in \mathbb{N},\ a_1, \ldots, a_k \in A\bigr\}$. In contrast, we will denote the F-subspace of W generated by A, namely $\mathrm{Span}\bigl(\{xa : x \in F,\ a \in A\}\bigr)$, by $\mathrm{Span}_F(A)$. We will now investigate the group generated by the range of a polynomial φ, $\mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr)$.

1.11. The following lemma is trivial.

Lemma. *For any mapping* $\varphi: V \longrightarrow W$ *and any* $v \in V$ *one has* $\mathrm{Span}\bigl(\mathrm{Ran}(D_v\varphi)\bigr) \subseteq \mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr)$.

1.12. Starting from this point we treat the cases $\mathrm{char}\, F = 0$ and $\mathrm{char}\, F \neq 0$ separately. We begin with the case $\mathrm{char}\, F = 0$.

Theorem. (i) *Let* $\mathrm{char}\, F = 0$, *let* $\psi_1, \ldots, \psi_c$ *be nontrivial distinct monomials* $F^d \longrightarrow F$ *and let* $\varphi = (\psi_1, \ldots, \psi_c)$. *Then* $\mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr) = F^c$.

(ii) *If, in addition, none of* $\psi_j, \ldots, \psi_c$, *is linear, then there exist* $r \in \mathbb{N}$ *and a nonzero polynomial* $\Pi: (F^d)^r \longrightarrow F$ *such that if* $\mathbf{y}_1, \ldots, \mathbf{y}_r \in F^d$ *satisfy* $\Pi(\mathbf{y}_1, \ldots, \mathbf{y}_r) \neq 0$, *then* $\mathrm{Span}\bigl(\mathrm{Ran}(D_{\mathbf{y}_1}\varphi) \cup \ldots \cup \mathrm{Ran}(D_{\mathbf{y}_r}\varphi)\bigr) = F^c$.

The case $d = 1$ of this theorem was obtained in [La].

The proof of Theorem 1.12 will be given in section 1.17; we first formulate some corollaries of this theorem.

1.13. Theorem. *Suppose* $\mathrm{char}\, F = 0$. *Let* φ *be a polynomial* $F^d \longrightarrow W$ *with zero constant term. Write* $\varphi = \psi_1 a_1 + \ldots + \psi_l a_l$, *where* $\psi_1, \ldots, \psi_l$ *are distinct monomials and* $a_1, \ldots, a_l \in W$. *Then* $\mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr) = \mathrm{Span}_F\{a_1, \ldots, a_l\}$.

1.14. In particular,

Corollary. *If* λ *is the linear part of* φ, *then* $\mathrm{Ran}(\lambda) = \mathrm{Span}\bigl(\mathrm{Ran}(\lambda)\bigr) \subseteq \mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr)$.

1.15. Theorem. *For any $d, n \in \mathbb{N}$ there exist $r \in \mathbb{N}$ and a nonzero polynomial $\Pi\colon (F^d)^r \longrightarrow F$ such that for any $\mathbf{y}_1, \dots, \mathbf{y}_r \in F^d$ with $\Pi(\mathbf{y}_1, \dots, \mathbf{y}_r) \neq 0$ and any polynomial $\varphi\colon F^d \longrightarrow W$ of degree $\leq n$ one has $\operatorname{Ran}(\lambda) + \sum_{s=1}^r \operatorname{Span}\bigl(\operatorname{Ran}(D_{\mathbf{y}_r}\varphi)\bigr) = \operatorname{Span}\bigl(\operatorname{Ran}(\varphi)\bigr)$, where λ is the linear part of φ.*

Indeed, for any $v \in V$, $\operatorname{Span}\bigl(\operatorname{Ran}(D_v\varphi)\bigr) \subseteq \operatorname{Span}\bigl(\operatorname{Ran}(\varphi)\bigr)$, and $\operatorname{Ran}(\lambda) \subseteq \operatorname{Span}\bigl(\operatorname{Ran}(\varphi)\bigr)$ by Corollary 1.14. This implies the inclusion "$\subseteq$". To get the opposite inclusion, we apply Theorem 1.12(ii) to the (finite) set of all monomials of degree $\leq n$ in d variables. We omit the details.

1.16. Example. Let $\varphi\colon F^2 \longrightarrow F^3$, $\varphi(x_1, x_2) = (x_1^2, x_1x_2, x_2^2)$. Then, for $\mathbf{y} = (y_1, y_2)$, $\mathbf{x} = (x_1, x_2)$, $D_{\mathbf{y}}\varphi(\mathbf{x}) = (2y_1x_1, y_2x_1 + y_1x_2, 2y_2x_2)$. Clearly this homomorphism $F^2 \longrightarrow F^3$ cannot be surjective for any $\mathbf{y} \in F^2$. Take any $\mathbf{z} = (z_1, z_2, z_3) \in F^3$ and consider the equation $D_{\mathbf{y}_1}\varphi(\mathbf{x}_1) + D_{\mathbf{y}_2}\varphi(\mathbf{x}_2) = \mathbf{z}$. In the coordinates $\mathbf{y}_1 = (y_{1,1}, y_{2,1})$, $\mathbf{y}_2 = (y_{1,2}, y_{2,2})$, $\mathbf{x}_1 = (x_{1,1}, x_{2,1})$, $\mathbf{x}_2 = (x_{1,2}, x_{2,2})$ this is

$$\begin{cases} 2y_{1,1}x_{1,1} + 2y_{1,2}x_{1,2} = z_1 \\ y_{2,1}x_{1,1} + y_{1,2}x_{2,1} + y_{2,2}x_{1,2} + y_{1,2}x_{2,2} = z_2 \\ 2y_{2,1}x_{2,1} + 2y_{2,2}x_{2,2} = z_3. \end{cases} \tag{1.3}$$

The system (1.3) is a corollary of

$$\begin{cases} 2y_{1,1}x_{1,1} + 2y_{1,2}x_{1,2} = z_1 \\ y_{2,1}x_{1,1} + y_{2,2}x_{1,2} = z_2 \\ y_{1,2}x_{2,1} + y_{1,2}x_{2,2} = 0 \\ 2y_{2,1}x_{2,1} + 2y_{2,2}x_{2,2} = z_3, \end{cases}$$

that is, of

$$\begin{pmatrix} 2y_{1,1} & 2y_{1,2} & & \\ y_{2,1} & y_{2,2} & \multicolumn{2}{c}{\smash{\raisebox{.5em}{O}}} \\ \multicolumn{2}{c}{\mathrm{O}} & y_{1,1} & y_{1,2} \\ & & 2y_{2,1} & 2y_{2,2} \end{pmatrix} \begin{pmatrix} x_{1,1} \\ x_{1,2} \\ x_{2,1} \\ x_{2,2} \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ 0 \\ z_3 \end{pmatrix}.$$

So, if the determinant of this system $\Pi(\mathbf{y}_1, \mathbf{y}_2) = 4(y_{1,1}y_{2,2} - y_{2,1}y_{1,2})^2 \neq 0$, the system (1.3) is solvable for any $\mathbf{z}$, so $\operatorname{Ran}(D_{\mathbf{y}_1}\varphi) + \operatorname{Ran}(D_{\mathbf{y}_2}\varphi) = F^3$. Since F is assumed to be infinite, $\Pi(\mathbf{y}_1, \mathbf{y}_2) \neq 0$ for some $\mathbf{y}_1, \mathbf{y}_2 \in F^2$. Since $\operatorname{Ran}(D_{\mathbf{y}}\varphi) \subseteq \operatorname{Span}\bigl(\operatorname{Ran}(\varphi)\bigr)$ for any $\mathbf{y} \in F^2$, we obtain $\operatorname{Span}\bigl(\operatorname{Ran}(\varphi)\bigr) = F^3$.

1.17. Proof of Theorem 1.12. We prove Theorem 1.12 by induction on the maximum degree of the monomials $\psi_1, \dots, \psi_c$. In the case where all the ψ_j are linear the statement is trivial, which gives the base of the induction. We start with the proof of (ii); assume that none of the ψ_j are linear. Let $\lambda_{\mathbf{y}}$, $\mathbf{y} \in F^d$, be the linear part of $D_{\mathbf{y}}\varphi = \bigl(D_{\mathbf{y}}\psi_1, \dots, D_{\mathbf{y}}\psi_c\bigr)$: $\lambda_{\mathbf{y}}(\mathbf{x}) = \Bigl(\sum_{i=1}^d u_{i,1}(\mathbf{y})x_i, \dots, \sum_{i=1}^d u_{i,c}(\mathbf{y})x_i\Bigr)$, where the $u_{i,j}$ are monomials in $\mathbf{y}$, and $D_{\mathbf{y}}\varphi - \lambda_{\mathbf{y}}$ contains no linear monomials in $\mathbf{x}$.

By the induction hypothesis (applied to Corollary 1.14) $\operatorname{Ran}(\lambda_{\mathbf{y}}) \subseteq \operatorname{Span}\bigl(\operatorname{Ran}(D_{\mathbf{y}}\varphi)\bigr)$ for any $\mathbf{y} \in F^d$. Therefore it suffices to find $r \in \mathbb{N}$ and a polynomial $\Pi\colon (F^d)^r \longrightarrow F$ such that $\Pi(\mathbf{y}_1, \dots, \mathbf{y}_r) \neq 0$ implies $\operatorname{Ran}(\lambda_{\mathbf{y}_1}) + \dots + \operatorname{Ran}(\lambda_{\mathbf{y}_r}) = F^c$.

For each $i = 1,\dots,d$ let r_i be the number of nonzero monomials among the $u_{i,j}$, $j = 1,\dots,c$, and let $r = \max\{r_1,\dots,r_d\}$. For each $j = 1,\dots,c$ choose $i = i_j$ such that $u_{i,j} \neq 0$.

Let $(z_1,\dots,z_c) \in F^c$; it suffices to attain the consistency of the following system of $r_1 + \dots + r_d$ linear equations in $r_1 + \dots + r_d$ variables $x_{i,s}$, $i = 1,\dots,d$, $s = 1,\dots,r_i$:

$$\sum_{s=1}^{r_i} u_{i,j}(\mathbf{y}_s)x_{i,s} = z_{i,j},\ i = 1,\dots,d,\ j = 1,\dots,c,\ \text{s.t. } u_{i,j} \neq 0, \qquad \text{where } z_{i,j} = z_j \text{ if } i = i_j \text{ and } z_{i,j} = 0 \text{ otherwise.} \tag{1.4}$$

For, summing from 1 to d, then letting $\mathbf{x}_s = (x_{1,s},\dots,x_{d,s})$ (where we use 0 for $x_{i,s}$ in the event $s > r_i$), one obtains $\sum_{s=1}^{r} \lambda_{\mathbf{y}_s}(\mathbf{x}_s) = (z_1,\dots,z_c)$.

The matrix of the system (1.4) is

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \mathrm{O} \\ & & \ddots & \\ \mathrm{O} & & & A_d \end{pmatrix}, \quad A_i = \begin{pmatrix} u_{i,j_1,k_1}(\mathbf{y}_1) & \dots & u_{i,j_1,k_1}(\mathbf{y}_{r_i}) \\ \vdots & & \vdots \\ u_{i,j_{r_i},k_{r_i}}(\mathbf{y}_1) & \dots & u_{i,j_{r_i},k_{r_i}}(\mathbf{y}_{r_i}) \end{pmatrix}, \ i = 1,\dots,d, \tag{1.5}$$

where $u_{i,j_1,k_1},\dots,u_{i,j_{r_i},k_{r_i}}$ is the list of all nonzero monomials among the $u_{i,j}$ for $j = 1,\dots,c$. Since $\psi_1,\dots,\psi_c$ are all distinct, for any $i = 1,\dots,d$ the nonzero monomials $u_{i,j}$, $j = 1,\dots,c$, are pairwise distinct (recall, $u_{i,j}$ is the coefficient of x_i, in $\mathbf{y}$, of $D_{\mathbf{y}}\psi_j$).

As there can therefore be no cancellation of terms in its computation, the determinant Π of the matrix (1.5) is a nonzero polynomial in the variables $\mathbf{y}_s$, $s = 1,\dots,r$. Thus if $\Pi(\mathbf{y}_1,\dots,\mathbf{y}_r) \neq 0$, then for any $\mathbf{z} = (z_1,\dots,z_c) \in F^c$ the system (1.4) has a solution $(x_{i,s},\ i = 1,\dots,d,\ s = 1,\dots,r_i)$.

We now deduce part (i) from part (ii). Let $\varphi = (\psi_1,\dots,\psi_c)$, where $\psi_1,\dots,\psi_l$ are distinct linear monomials and $\psi_{l+1},\dots,\psi_c$ are distinct non-linear monomials. By (ii), $\sum_{\mathbf{y}\in F^d} \mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}}\varphi)\big) = \{0\}^l \times F^{c-l}$. Since $\mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}}\varphi)\big) \subseteq \mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$ for any $\mathbf{y} \in F^d$, we have $\{0\}^l \times F^{c-l} \subseteq \mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$. Since, plainly, $\mathrm{Ran}(\psi_1,\dots,\psi_l) = F^l$, we are done. ∎

1.18. We now turn to the case $\mathrm{char}\, F = p < \infty$. For $\mathbf{x} = (x_1,\dots,x_d) \in F^d$, let $\mathbf{x}^p = (x_1^p,\dots,x_d^p)$. If V is a vector space over F, the operation $v \mapsto v^p$ is not well defined on V. Hence when we write v^p for $v \in V$ we assume that some coordinate system in V is fixed.

1.19. The proof of the following theorem will be given in section 1.24.

Theorem. (i) *Suppose* $\mathrm{char}\, F = p < \infty$. *Let* $\psi_1,\dots,\psi_c$ *be pairwise distinct nontrivial separable monomials* $F^d \longrightarrow F$ *and let* $\varphi\colon F^d \longrightarrow F^c$, $\varphi(\mathbf{x}) = (\psi_1(\mathbf{x}),\dots,\psi_c(\mathbf{x}))$. *Then* $\mathrm{Span}\big(\mathrm{Ran}(\varphi)\big) = F^c$.

(ii) *Let, in addition, none of* $\psi_1,\dots,\psi_c$ *be linear, let* $n = \max\{\deg\psi_1,\dots,\deg\psi_c\}$ *and let* $m \geq [\log_p n]$. *Then there exist* $r \in \mathbb{N}$ *and a nonzero polynomial* $\Pi\colon (F^d)^r \longrightarrow F$ *such that if* $\mathbf{y}_1,\dots,\mathbf{y}_r \in F^d$ *satisfy* $\Pi(\mathbf{y}_1,\dots,\mathbf{y}_r) \neq 0$, *then* $\mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}_1^{p^m}}\varphi) \cup \dots \cup \mathrm{Ran}(D_{\mathbf{y}_r^{p^m}}\varphi)\big) = F^c$.

The case $d = 1$ of this theorem was obtained in [La].

1.20. Example. Let $\varphi\colon F \longrightarrow F$, $\varphi(x) = x^{p+1}$. Then for $y \in F$, $D_y\varphi(x) = yx^p + y^p x$, which is a linear mapping that is not surjective in general. Take any $z \in F$ and consider the equation $D_{y_1^p}\varphi(x_1) + D_{y_2^p}\varphi(x_2) = z$, that is,

$$y_1^p x_1^p + y_1^{p^2} x_1 + y_2^p x_2^p + y_2^{p^2} x_2 = z. \tag{1.6}$$

Equation (1.6) is a corollary of the system of equations

$$\begin{cases} y_1^p x_1^p + y_2^p x_2^p = 0 \\ y_1^{p^2} x_1 + y_2^{p^2} x_2 = z, \end{cases}$$

which is equivalent to

$$\begin{cases} y_1 x_1 + y_2 x_2 = 0 \\ y_1^{p^2} x_1 + y_2^{p^2} x_2 = z. \end{cases}$$

The determinant of this system of linear equations in x_1, x_2 is $\Pi(y_1, y_2) = y_1 y_2^{p^2} - y_1^{p^2} y_2$. Hence, for y_1, y_2 satisfying $\Pi(y_1, y_2) \neq 0$, the equation (1.6) is solvable for any $z \in F$, so $\mathrm{Ran}(D_{y_1^p}\varphi) + \mathrm{Ran}(D_{y_2^p}\varphi) = F$. Since F is assumed to be infinite, $\Pi(y_1, y_2) \neq 0$ for some y_1, $y_2 \in F$, which implies $\mathrm{Span}\bigl(\{x^{p+1} : x \in F\}\bigr) = F$.

1.21. Let φ be a polynomial $F^d \longrightarrow W$ with zero constant term. Represent φ in the form $\varphi = \eta_1 \circ \psi_1 + \ldots + \eta_l \circ \psi_l$ where $\psi_1, \ldots, \psi_l$ are separable monomials and $\eta_1, \ldots, \eta_l$ are homomorphisms $F \longrightarrow W$. Define a polynomial $\psi\colon F^d \longrightarrow F^l$ by $\psi = (\psi_1, \ldots, \psi_l)$ and a homomorphism $\pi\colon F^l \longrightarrow W$ by $\pi(\mathbf{z}_1, \ldots, \mathbf{z}_l) = \eta_1(\mathbf{z}_1) + \ldots + \eta_l(\mathbf{z}_l)$. Then $\varphi = \pi \circ \psi$, and Theorem 1.19(i) implies $\mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr) = \pi\bigl(\mathrm{Span}\bigl(\mathrm{Ran}(\psi)\bigr)\bigr) = \pi(F^l) = \mathrm{Ran}(\pi)$. We therefore have

Theorem. $\mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr) = \sum_{j=1}^{l} \mathrm{Ran}(\eta_j)$.

Let us remark that, unlike in the case $\mathrm{char}\, F = 0$, $\mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr)$ need not be an F-subspace of W, since homomorphisms $\eta\colon F \longrightarrow F$, $\eta(x) = a_0 x + a_1 x^p + \ldots + a_m x^{p^m}$ need not be surjective. If F is an algebraic field (that is, an algebraic extension of $\mathbb{Z}_p$), the range of such η is a subgroup of finite index of the additive group of F; if F is transcendental over $\mathbb{Z}_p$ then the range of η may have infinite index in F.

1.22. Suppose further that $\psi_1, \ldots, \psi_t$ are linear monomials, while $\psi_{t+1}, \ldots, \psi_l$ are non-linear. Then $\lambda = \sum_{i=j}^{t} \eta_j \circ \psi_j$ is the linear part of φ, and $\mathrm{Ran}(\lambda) = \mathrm{Span}\bigl(\mathrm{Ran}(\lambda)\bigr) = \sum_{j=1}^{t} \mathrm{Ran}(\eta_j)$. In particular, we have

Corollary. $\mathrm{Ran}(\lambda) \subseteq \mathrm{Span}\bigl(\mathrm{Ran}(\varphi)\bigr)$.

1.23. Under the assumptions of 1.21 and 1.22, let $\deg\varphi = n$, $m \geq [\log_p n]$ and let Π be the polynomial arising in Theorem 1.19 applied to the set of all separable monomials of formal degree $\leq n$. Let a polynomial $\psi\colon F^d \longrightarrow F^{l-t}$ be defined by $\psi = (\psi_{t+1}, \ldots, \psi_r)$, and define a homomorphism $\pi\colon F^{l-t} \longrightarrow W$ by $\pi(\mathbf{z}_{t+1}, \ldots, \mathbf{z}_l) = \eta_{t+1}(\mathbf{z}_1) + \ldots + \eta_l(\mathbf{z}_l)$. Then for any $\mathbf{y}_1, \ldots, \mathbf{y}_r \in F^d$ with $\Pi(\mathbf{y}_1, \ldots, \mathbf{y}_r) \neq 0$, Theorem 1.19(ii) gives

$$\sum_{s=1}^{r} \mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}_s^{p^m}}\varphi)\big) = \sum_{s=1}^{r} \mathrm{Span}\Big(\mathrm{Ran}\big(\sum_{j=t+1}^{l} D_{\mathbf{y}_s^{p^m}}(\eta_j \circ \psi_j)\big)\Big)$$
$$= \sum_{s=1}^{r} \mathrm{Span}\Big(\mathrm{Ran}\big(\sum_{j=t+1}^{l} \eta_j \circ D_{\mathbf{y}_s^{p^m}}(\psi_j)\big)\Big) = \sum_{s=1}^{r} \mathrm{Span}\big(\mathrm{Ran}(\pi \circ D_{\mathbf{y}_s^{p^m}}\psi)\big)$$
$$= \pi\Big(\sum_{s=1}^{r} \mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}_s^{p^m}}\psi)\big)\Big) = \pi(F^{l-t}) = \sum_{j=t+1}^{l} \mathrm{Ran}(\eta_j).$$

Combining this with the results of 1.21 and 1.22, we obtain

Theorem. *For any $d, m \in \mathbb{N}$ there exist $r \in \mathbb{N}$ and a nonzero polynomial $\Pi\colon (F^d)^r \longrightarrow F$ such that for any $\mathbf{y}_1, \ldots, \mathbf{y}_r \in F^d$ satisfying $\Pi(\mathbf{y}_1, \ldots, \mathbf{y}_r) \neq 0$ and any polynomial $\varphi\colon F^d \longrightarrow W$ of formal degree $< p^{m+1}$ one has $\mathrm{Ran}(\lambda) + \sum_{s=1}^{r} \mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}_s^{p^m}}\varphi)\big) = \mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$, where λ is the linear part of φ.*

1.24. Proof of Theorem 1.19. We prove Theorem 1.19 by induction on the maximum degree of the monomials $\psi_1, \ldots, \psi_c$. In the case where all the ψ_j are linear the statement is trivial. We start with the proof of (ii); assume that no ψ_j is linear. Let $\lambda_{\mathbf{y}}$, $\mathbf{y} \in F^d$, be the linear part of $D_{\mathbf{y}}\varphi = (D_{\mathbf{y}}\psi_1, \ldots, D_{\mathbf{y}}\psi_c)$: $\lambda_{\mathbf{y}}(\mathbf{x}) = \Big(\sum_{i=1}^{d}\sum_{k=0}^{m} u_{i,1,k}(\mathbf{y})x_i^{p^k}, \ldots, \sum_{i=1}^{d}\sum_{k=0}^{m} u_{i,c,k}(\mathbf{y})x_i^{p^k}\Big)$, where $m \geq [\log_p n]$, $u_{i,j,k}$ are monomials in $\mathbf{y}$, and $D_{\mathbf{y}}\varphi - \lambda_{\mathbf{y}}$ contains no linear monomials in $\mathbf{x}$. Since by the induction hypothesis (applied to Corollary 1.22) $\mathrm{Ran}(\lambda_{\mathbf{y}}) \subseteq \mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}}\varphi)\big)$, $\mathbf{y} \in F^d$, it suffices to show that there are $r \in \mathbb{N}$ and a polynomial $\Pi\colon (F^d)^r \longrightarrow F$ such that $\Pi(\mathbf{y}_1, \ldots, \mathbf{y}_r) \neq 0$ implies $\mathrm{Ran}(\lambda_{\mathbf{y}_1^{p^m}}) + \ldots + \mathrm{Ran}(\lambda_{\mathbf{y}_r^{p^m}}) = F^c$.

Since $\psi_1, \ldots, \psi_c$ are distinct, for any $i = 1, \ldots, d$ the nonzero monomials $u_{i,j,k}$, $j = 1, \ldots, c$, $k = 0, \ldots, m$, are pairwise distinct. For each $i = 1, \ldots, d$ let r_i be the number of nonzero monomials among $u_{i,j,k}$, $j = 1, \ldots, c$, $k = 0, \ldots, m$, and let $r = \max\{r_1, \ldots, r_d\}$. For each $j = 1, \ldots, c$, let $\psi_j = \prod_{i=1}^{d} x_i^{n_{i,j}}$. Since ψ_j is separable, there is $i = i_j$ such that $n_{i,j}$ is not divisible by p; we have $u_{i,j,k} \neq 0$ for this i. Let $(z_1, \ldots, z_c) \in F^c$; consider the system of equations

$$\sum_{s=1}^{r_i} u_{i,j,k}(\mathbf{y}_s^{p^m})x_{i,s}^{p^k} = z_{i,j,k},\ i = 1, \ldots, d,\ j = 1, \ldots, c,\ k = 0, \ldots, m,\ \text{s.t. } u_{i,j,k} \neq 0,$$
$$\text{where } z_{i,j,k} = z_j \text{ if } k = 0 \text{ and } i = i_j, \text{ and } z_{i,j,k} = 0 \text{ otherwise.}$$

Extracting the p^k-th root from the equations corresponding to $k > 0$ (keeping in mind that $z_{i,j,k} = 0$ in this case), we get

$$\sum_{s=1}^{r_i} u_{i,j,k}(\mathbf{y}_s^{p^{m-k}})x_{i,s} = z_{i,j,k},\ i = 1,\dots,d,\ j = 1,\dots,c,\ k = 0,\dots,m,\ u_{i,j,k} \neq 0. \quad (1.7)$$

The matrix of this system has form

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \mathrm{O} \\ \mathrm{O} & & \ddots & \\ & & & A_d \end{pmatrix}, \quad A_i = \begin{pmatrix} u_{i,j_1,k_1}(\mathbf{y}_1^{p^{m-k_1}}) & \dots & u_{i,j_1,k_1}(\mathbf{y}_{r_i}^{p^{m-k_1}}) \\ \vdots & & \vdots \\ u_{i,j_{r_i},k_{r_i}}(\mathbf{y}_1^{p^{m-k_{r_i}}}) & \dots & u_{i,j_{r_i},k_{r_i}}(\mathbf{y}_{r_i}^{p^{m-k_{r_i}}}) \end{pmatrix},\ i = 1,\dots,d, \quad (1.8)$$

where $u_{i,j_1,k_1},\dots,u_{i,j_{r_i},k_{r_i}}$ is the list of all nonzero monomials among $u_{i,j,k}$, $j = 1,\dots,c$, $k = 0,\dots,m$. For any i, $(j_a,k_a) \neq (j_b,k_b)$ implies $u_{i,j_a,k_a}(\mathbf{y}^{p^{m-k_a}}) \neq u_{i,j_b,k_b}(\mathbf{y}^{p^{m-k_b}})$, except for the case when both are zero. Indeed, when both are non-zero, direct computation yields $u_{i,j_a,k_b}(\mathbf{y}^{p^{m-k_a}}) = y_i^{(n_{i,j_a}-p^{k_a})p^{m-k_a}} \prod_{i\neq t=1}^d y_t^{n_{t,j_a}p^{m-k_a}}$ and $u_{i,j_b,k_b}(\mathbf{y}^{p^{m-k_b}}) = y_i^{(n_{i,j_b}-p^{k_b})p^{m-k_b}} \prod_{i\neq t=1}^d y_t^{n_{t,j_b}p^{m-k_b}}$. If these are equal then, equating the exponents of each y_t, $1 \le t \le d$, yields $p^{k_b}n_{t,j_a} = p^{k_a}n_{t,j_b}$ for all t, $1 \le t \le d$ (for $t \neq i$ this is immediate, while for $t = i$ some minor computation is necessary). Since ψ_{j_a} and ψ_{j_b} are distinct monomials, we cannot have $k_a = k_b$. But if, without loss of generality, $k_a > k_b$, then p divides n_{t,j_a}, $1 \le t \le d$, a contradiction.

This implies that the determinant Π of the matrix (1.8) is a nonzero polynomial in $\mathbf{y}_s$, $s = 1,\dots,r$. Thus if $\Pi(\mathbf{y}_1,\dots,\mathbf{y}_r) \neq 0$, then for any $\mathbf{z} = (z_1,\dots,z_c) \in F^c$ the equation (1.7) has a solution $(x_{i,s},\ i = 1,\dots,d,\ s = 1,\dots,r_i)$. For $i = 1,\dots,d$ put $x_{i,s} = 0$ if $r_i < s \le r$ and let $\mathbf{x}_s = (x_{1,s},\dots,x_{d,s})$, $s = 1,\dots,r$. Then $\sum_{s=1}^r \lambda_{\mathbf{y}_s^{p^m}}(\mathbf{x}_s) = \mathbf{z}$ and hence $\sum_{s=1}^r \mathrm{Ran}(\lambda_{\mathbf{y}_s^{p^m}}) = F^c$.

Now we may deduce part (i) of Theorem 1.19 from part (ii). Let $\varphi = (\psi_1,\dots,\psi_c)$, where $\psi_1,\dots,\psi_c$ are distinct nontrivial separable monomials. Assume that $\psi_1,\dots,\psi_l$ are linear while $\psi_{l+1},\dots,\psi_c$ are nonlinear. By (ii), $\sum_{\mathbf{y}\in F^d} \mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}}\varphi)\big) = \{0\}^l \times F^{c-l}$. Since $\mathrm{Span}\big(\mathrm{Ran}(D_{\mathbf{y}}\varphi)\big) \subseteq \mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$ for any $\mathbf{y} \in F^d$, we have $\{0\}^l \times F^{c-l} \subseteq \mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$. Since $\mathrm{Ran}(\psi_1,\dots,\psi_l) = F^l$, we are done. ■

1.25. We will continue to assume that $\mathrm{char}\, F = p < \infty$ for the remainder of this chapter. Theorem 1.21 leads us to introduce the following notion. Let us denote by F_k, $k = 0,1,\dots$, the subfield $\{x^{p^k} : x \in F\}$ of F. Let W be a vector space over F, let A be a subset of W and let $m,l \in \mathbb{N}$. We say that a subgroup L of W is *a (p,A,m)-space of dimension $\le l$* if L is representable in the form

$$L = \Big\{\sum_{i=1}^{l}\sum_{k=0}^{m} x_i^{p^k} a_{i,k} : x_1,\dots,x_l \in F\Big\}$$

with $a_{i,k} \in \mathrm{Span}_{F_k}(A)$, $i = 1,\dots,l$, $k = 0,\dots,m$.

1.26. Let $\varphi: V \longrightarrow W$ be a polynomial with zero constant term. Write $\varphi = \eta_1 \circ \psi_1 + \ldots + \eta_l \circ \psi_l$, where ψ_i are separable monomials and η_i are homomorphisms $F \longrightarrow W$, $\eta_i(x) = a_{i,0}x + x^p a_{i,1} + \ldots + x^{p^{m_i}} a_{i,m_i}$, $i = 1, \ldots, l$. Assume that for every $i = 1, \ldots, l$ and $k = 0, \ldots, m_i$ one has $a_{i,k} \in \mathrm{Span}_{F_k}(A)$. Then by Theorem 1.21,

$$\mathrm{Span}\big(\mathrm{Ran}(\varphi)\big) = \sum_{i=1}^{l} \mathrm{Ran}(\eta_i) = \sum_{i=1}^{l} \Big\{ \sum_{k=0}^{m_i} x^{p^k} a_{i,k} : x \in F \Big\}$$
$$= \Big\{ \sum_{i=1}^{l} \sum_{k=0}^{m_i} x_i^{p^k} a_{i,k} : x_1, \ldots, x_l \in F \Big\}.$$

That is, $\mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$ is a (p, A, m)-space for any $m \geq \max\{m_1, \ldots, m_l\}$.

Remark. The "monomials" and the "coefficients" of a polynomial $V \longrightarrow W$ are only defined modulo a coordinate system in V. When we discuss them, we assume that some coordinate system in V has been chosen.

1.27. Let $\Lambda = \Lambda(A, m)$ be the set of all (p, A, m)-spaces. We may interpret Λ in the following way. Let $V_k = \mathrm{Span}_{F_k}(A)$, $k = 0, \ldots, m$, and let $\mathcal{V} = V_0 \oplus V_1 \oplus \ldots \oplus V_m$. $\mathcal{V}$ is an F-vector space with scalar multiplication $x(b_0, b_1, \ldots, b_m) = (xb_0, x^p b_1, \ldots, x^{p^m} b_m)$, $x \in F$, $(b_0, b_1, \ldots, b_m) \in \mathcal{V}$. Define $\sigma: \mathcal{V} \longrightarrow W$ by $\sigma(b_0, \ldots, b_m) = b_0 + \ldots + b_m$. Then $\Lambda = \big\{\sigma(\mathcal{L}) : \mathcal{L} \text{ is a subspace of } \mathcal{V}\big\}$.

1.28. If A is a finite set, then $\mathcal{V}$ is a finite dimensional space. This implies that

Lemma. *For a finite set $A \subset W$ and $m \in \mathbb{N}$, $\Lambda = \Lambda(A, m)$ satisfies the ascending chain condition.*

Proof. Let $L_1 \subseteq L_2 \subseteq \ldots$ be an ascending chain of (p, A, m) spaces. For $n \in \mathbb{N}$, let $\mathcal{L}_n$ be a subspace of $\mathcal{V}$ such that $\sigma(\mathcal{L}_n) = L_n$ and let $\mathcal{M}_n = \mathcal{L}_1 + \ldots + \mathcal{L}_n$. Then $\mathcal{M}_1 \subseteq \mathcal{M}_2 \subseteq \ldots$ is an ascending chain of subspaces of a finite dimensional space $\mathcal{V}$ and hence, it stabilizes: $\mathcal{M}_m = \mathcal{M}_{m+1} = \ldots$ for some m. Since $\sigma(\mathcal{M}_n) = L_n$ for all n, we have $L_m = L_{m+1} = \ldots$. ∎

1.29. We will also need the following proposition.

Proposition. *Let φ be a polynomial $V \longrightarrow W$ of formal degree $< p^{m+1}$ with trivial linear part and let L be an additive subgroup of W. Assume that there exists $u \in V$ such that $\mathrm{Ran}(D_{u^{p^m}} \varphi) \not\subseteq L$. Then there exists a nonzero polynomial $R: V \longrightarrow F$ such that for any $v \in V$ satisfying $R(v) \neq 0$ one has $\mathrm{Ran}(D_{v^{p^m}} \varphi) \not\subseteq L$.*

Informally, this means that if φ has no linear part and $\mathrm{Ran}(D_{u^{p^m}}) \not\subseteq L$, then $\mathrm{Ran}(D_{v^{p^m}} \varphi) \not\subseteq L$ for almost all $v \in V$.

Proof. We identify V with F^d for some d and let $\Pi: V^r \longrightarrow F$ be the polynomial corresponding to d and m in Theorem 1.23. Let R_1 be a nonzero polynomial $V \longrightarrow F$ such that for any w with $R_1(w) \neq 0$, $\Pi(w, v_2, \ldots, v_r)$ is a nonzero polynomial in the variables $v_2, \ldots, v_r$. If $\mathrm{Ran}(D_{w^{p^m}} \varphi) \not\subseteq L$ for all w with $R_1(w) \neq 0$, we are done. Otherwise, there is

$w_1 \in V$ such that $\mathrm{Ran}(D_{w_1^{p^m}}\varphi) \subseteq L$ and $\Pi(w_1, v_2, \ldots, v_r)$ is a nonzero polynomial. Then let R_2 be a polynomial $V \longrightarrow F$ such that for any w with $R_2(w) \neq 0$, $\Pi(w_1, w, v_3, \ldots, v_r)$ is a nonzero polynomial in the variables $v_3, \ldots, v_r$. If $\mathrm{Ran}(D_{w^{p^m}}\varphi) \nsubseteq L$ for all w with $R_2(w) \neq 0$, we are done. Otherwise, find $w_2 \in V$ such that $\mathrm{Ran}(D_{w_2^{p^m}}\varphi) \subseteq L$ and $\Pi(w_1, w_2, v_3, \ldots, v_r)$ is a nonzero polynomial. And so on. If this process terminates at a step $k \leq r$, then we are done. Otherwise we have obtained $w_1, \ldots, w_r \in V$ such that $\Pi(w_1, \ldots, w_r) \neq 0$ and $\mathrm{Ran}(D_{w_1^{p^m}}\varphi), \ldots, \mathrm{Ran}(D_{w_r^{p^m}}\varphi) \subseteq L$. By Theorem 1.23, this implies that $\mathrm{Ran}\,\varphi \subseteq L$, which contradicts our assumption. ■

2. Densities and a generalized van der Corput lemma

In this chapter we introduce some technical notions and notation and obtain some facts that will be needed in the sequel.

Throughout this chapter G is a countable abelian group, written additively.

2.1. If $\{x_u\}_{u\in G}$ is a set of elements of a topological group indexed by the elements of G, and $\{\Phi_n\}_{n=1}^{\infty}$ is a Følner sequence in G, then we write $\underset{u\in G}{\text{C-lim}}\, x_u = x$ for $\lim_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{u\in\Phi_n} x_u = x$. If the x_u are real numbers, then we write $\underset{u\in G}{\text{C-limsup}}\, x_u = x$ for $\limsup_{n\to\infty} \frac{1}{|\Phi_n|} \sum_{u\in\Phi_n} x_u = x$. Because this notion depends on the averaging Følner sequence, we will implicitly assume such a sequence to be fixed throughout any given discussion, and assertions made will be assumed to be valid for any such Følner sequence. In particular, multiple C-lim expressions in the same discussion are assumed, unless otherwise stated, to be taken along the same averaging Følner sequence. A similar convention will be held for the D-lim operator discussed below.

2.2. For $S \subseteq G$, we write $\overline{d}(S) = \underset{u\in G}{\text{C-limsup}}\, 1_S(u)$ when a given Følner sequence is understood, and $d^*(S) = \sup\{\overline{d}(S)\}$, where the supremum is taken over *all* Følner sequences in G. $\overline{d}(S)$ is called *the upper density* of S with respect to the given Følner sequence, $d^*(S)$ is called *the upper Banach density* of S. Analogously, *the lower density* $\underline{d}(S)$ of S is defined as $\underset{u\in G}{\text{C-liminf}}\, 1_S(u)$ and *the lower Banach density* $d_*(S)$ of S is $\inf\{\underline{d}(S)\}$. If $\overline{d}(S) = \underline{d}(S) = d$ we say that S has *density* d (with respect to the chosen Følner sequence); if $d^*(S) = d_*(S) = d$ we say that S has *Banach density* d.

2.3. We write $\underset{u\in G}{\text{D-lim}}\, x_u = x$ if for any neighborhood K of x the set $\{u \in G : x_u \in K\}$ has density 1. If x_u are reals we write $\underset{u\in G}{\text{D-limsup}}\, x_u$ for the infimum of the set of $x \in \mathbb{R}$ for which the set $\{u : x_u < x\}$ has density 1.

2.4. If some property $P(u)$ holds for all $u \in S$, where S is a subset of Banach density 1 in G, we will say that P holds for *almost all* elements of G. We leave the proof of the following lemma to the reader.

Lemma. *If V is a vector space and φ is a nonzero polynomial on V, then $\varphi(u) \neq 0$ for almost all $u \in V$.*

2.5. Remark. Suppose that G^* is a subgroup of G and let $u \mapsto u^*$ be a homomorphism of G onto G^*. Then sets having Banach density 1 in G map onto sets having Banach density 1 in G^*. Hence, the statement "$P(u^*)$ holds for almost all $u \in G$" implies the statement "$P(u^*)$ holds for almost all $u^* \in G^*$." We will repeatedly use this fact without mentioning it.

2.6. A subset S of G is *syndetic* if finitely many translates of S cover G; that is, if there exist $u_1, \dots, u_k \in G$ such that $\bigcup_{i=1}^k (u_i + S) = G$. A set $T \subseteq G$ is *thick* if $T \cap S \neq \emptyset$ for every syndetic S.

2.7. The following lemma is routine.

Lemma. *Let $S, T \subseteq G$.*
(i) *$\overline{d}$, $\underline{d}$, d^* and d_* are shift invariant: for any $u \in G$, $\overline{d}(S+u) = \overline{d}(S)$, etc.*
(ii) *$\overline{d}(S \cap T) \geq \overline{d}(S) + \underline{d}(T) - 1$ and $d^*(S \cap T) \geq d^*(S) + d_*(T) - 1$;*
(iii) *If $u_i \in G$, $1 \leq i \leq k$, and $\overline{d}\big((u_i + S) \cap (u_j + S)\big) = 0$ for $1 \leq i \neq j \leq k$, then $\overline{d}\big(\bigcup_{i=1}^k (u_i + S)\big) = k\overline{d}(S)$, and similarly for d^*.*
(iv) *S is syndetic if and only if $G \setminus S$ fails to be thick if and only if $d_*(S) > 0$.*
(v) *T is thick if and only if $G \setminus T$ fails to be syndetic if and only if $d^*(T) = 1$.*
(vi) *If T is thick and S has Banach density 1, then $T \cap S$ is thick.*
(vii) *If T is thick then for any $u_1, \dots, u_k \in G$ the intersection $\bigcap_{i=1}^k (u_i + T)$ is also thick.*
(viii) *If T is thick then T contains a Δ-set. That is, there exists a sequence $\{u_i\}_{i=1}^\infty$ in G such that $\{u_j - u_i : 1 \leq i < j\} \subseteq T$.*
(ix) *If $\overline{d}(S) > 0$ and G^* is a subgroup of G, then $S^* = \{u \in G^* : \overline{d}\big((S+u) \cap S\big) > 0\}$ is syndetic in G^*.*

Proof. We prove only (ix). Suppose not. Then by (iv), $G^* \setminus S^* = \{u \in G^* : \overline{d}\big((S+u) \cap S\big) = 0\}$ is thick and hence by (viii) there exists a sequence $\{u_i\}_{i=1}^\infty$ in G^* such that $u_j - u_i \in G^* \setminus S^*$, $1 \leq i < j$. It follows from (i) that for $i \neq j$, $\overline{d}\big((u_i + S) \cap (u_j + S)\big) = 0$, so for all k we have by (iii) that $\overline{d}\big(\bigcup_{i=1}^k (u_i + S)\big) = k\overline{d}(S)$. Taking $k > 1/\overline{d}(S)$ yields a contradiction. ■

2.8. For $M \subseteq G$ we will denote by $\mathrm{FS}(M)$ the set of finite sums of distinct elements of M, that is, $\mathrm{FS}(M) = \{\sum_{v \in \alpha} v : \alpha \subseteq M,\ 0 < |\alpha| < \infty\}$.

Lemma. *Suppose $T \subseteq G$ is thick. Let $d \in \mathbb{N}$ and suppose S is a syndetic subset of G^d. Then there exists $\mathbf{v} = (v_1, \dots, v_d) \in S$ such that $\mathrm{FS}\big(\{v_1, \dots, v_d\}\big) \subset T$.*

Proof. Initially we shall omit the requirement $(v_1, \dots, v_d) \in S$. We use induction on d. Suppose $v_1, \dots, v_{d-1} \in V$ have been chosen so that $\mathrm{FS}\big(\{v_1, \dots, v_{d-1}\}\big) \subset T$. Let $T' = T \cap \bigcap_{v \in \mathrm{FS}(\{v_1, \dots, v_{d-1}\})} (T - v)$. Then, by Lemma 2.7(vii), T' is nonempty, and letting $v_d \in T'$ one gets $\mathrm{FS}\big(\{v_1, \dots, v_d\}\big) \subset T$.

Now we show that in fact we may require $(v_1, \dots, v_d) \in S$. Since S is syndetic, for some $(u_{1,1}, \dots, u_{1,d}), \dots, (u_{k,1}, \dots, u_{k,d}) \in V^d$ one has $\bigcup_{i=1}^k \big((u_{i,1}, \dots, u_{i,d}) + S\big) = V^d$.

Put $B = \bigcup_{i=1}^{k} \mathrm{FS}(\{u_{i,1}, \dots, u_{i,d}\})$. The set $T'' = \bigcap_{u \in B}(u + T)$ is thick, so by our prior argument, there exist $w_1, \dots, w_d \in V$ with $\mathrm{FS}(\{w_1, \dots, w_d\}) \subset T''$.

Choose i with $(v_1, \dots, v_d) = (w_1, \dots, w_d) - (u_{i,1}, \dots, u_{i,d}) \in S$. Then for any nonempty $\alpha \subseteq \{1, \dots, d\}$ one has

$$\sum_{j \in \alpha} v_j = \sum_{j \in \alpha} w_j - \sum_{j \in \alpha} u_{i,j} = w - u,$$

where $w \in T''$ and $u \in B$. Hence, $\sum_{j \in \alpha} v_j \in T$. ■

2.9. We will also need a van der Corput-type lemma:

Lemma. *Let $\{f_u\}_{u \in G}$ be a bounded set of vectors in a Hilbert space indexed by the elements of G. Fix a Følner sequence $\{\Phi_n\}_{n \in \mathbb{N}}$ and assume that there exists an infinite set $H \subseteq G$ such that for all nonzero $v \in H - H$, $\underset{u \in G}{\text{C-limsup}} \langle f_{u+v}, f_u \rangle = 0$. Then $\underset{u \in G}{\text{C-lim}}\, f_u = 0$ in the strong topology. In particular, if G^* is an infinite subgroup of G and $\underset{u \in G}{\text{C-lim}} \langle f_{u+v}, f_u \rangle = 0$ for almost all $v \in G^*$, then $\underset{u \in G}{\text{C-lim}}\, f_u = 0$.*

Proof. Without loss of generality, we assume that $\|f_u\| < 1$ for all $u \in G$. Let $\varepsilon > 0$ be arbitrary. Pick a finite set $K \subset H$ with $|K| > \frac{1}{\varepsilon}$. For $n \in \mathbb{N}$,

$$\frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} f_u = \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \left(\frac{1}{|H|} \sum_{v \in H} f_{u+v} \right) + \delta_n = \sigma_n + \delta_n,$$

where $\|\delta_n\| \to 0$ as $n \to \infty$. Now

$$\|\sigma_n\|^2 \leq \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \left\| \frac{1}{|H|} \sum_{v \in H} f_{u+v} \right\|^2 = \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \frac{1}{|H|^2} \sum_{v_1, v_2 \in H} \langle f_{u+v_1}, f_{u+v_2} \rangle.$$

But for fixed $v_1, v_2 \in H$, with $v_1 \neq v_2$,

$$\limsup_{n \to \infty} \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \langle f_{u+v_1}, f_{u+v_2} \rangle = \limsup_{n \to \infty} \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \langle f_u, f_{u+v} \rangle = 0,$$

where we have substituted $v = v_2 - v_1 \in H - H$.

It follows that for n large enough

$$\|\sigma_n\|^2 \leq \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \frac{1}{|H|^2} \sum_{\substack{v_1, v_2 \in H \\ v_1 \neq v_2}} \langle f_{u+v_1}, f_{u+v_2} \rangle + \frac{1}{|\Phi_n|} \sum_{u \in \Phi_n} \frac{1}{|H|^2} \sum_{v \in H} \|f_{u+v}\|^2$$

$$< \varepsilon + \frac{1}{|H|} < 2\varepsilon.$$

The last assertion follows from Lemma 2.7 (viii). ■

2.10. The following is a version of Lemma 2.9 for D-limits:

Lemma. *Let $\{f_u\}_{u\in G}$ be a bounded set of vectors of a Hilbert space $\mathcal{H}$ indexed by the elements of G. Assume that there exists an infinite subgroup G^* of G such that* $\underset{v\in G^*}{\text{D-lim}}\,\underset{u\in G}{\text{D-limsup}}\langle f_{u+v}, f_u\rangle = 0$. *Then* $\underset{u\in G}{\text{D-lim}}\langle f_u, g\rangle = 0$ *for all $g \in \mathcal{H}$. In particular, if* $\underset{u\in G}{\text{D-lim}}\langle f_{u+v}, f_u\rangle = 0$ *for almost all $v \in G^*$, then* $\underset{u\in G}{\text{D-lim}}\langle f_u, g\rangle = 0$ *for all $g \in \mathcal{H}$.*

Note that there are several D-lim expressions in this lemma. Most are taken with respect to the same Følner sequence in G, while the first one is with respect to a Følner sequence in G^*. This should cause minimal confusion, as one can see from the indexing set whether one is dealing with G or G^*.

Proof. This proof is essentially due to Furstenberg (cf. [F2], Lemma 4.9). Suppose not. Then there exist $\varepsilon > 0$ and $g \in \mathcal{H}$ such that $S = \{u \in G : |\langle f_u, g\rangle| > \varepsilon\}$ satisfies $\overline{d}(S) > 0$. Without loss of generality we may assume that $\|g\| = 1$ and that $\|f_u\| \le 1$, $u \in G$. Take $\delta < \varepsilon^2$ and put $Q = \{v \in G^* : \underset{u\in G}{\text{D-limsup}}\langle f_{u+v}, f_u\rangle < \delta\}$; Q has density 1 in G^*.

We claim that there exist a sequence $v_1, v_2, \ldots \in G^*$ and a nested sequence of sets $S \supseteq S_1 \supseteq S_2 \supseteq \ldots$ such that

(a) $\overline{d}(S_k) > 0$ for all $k \in \mathbb{N}$;

(b) $v_j - v_i \in Q$ for all $1 \le i < j$;

(c) for any $k \in \mathbb{N}$, $1 \le i < j \le k$ and $u \in S_k$, $\langle f_{u+v_i}, f_{u+v_j}\rangle < \delta$;

(d) for any $k \in \mathbb{N}$, $u \in S_k$ and $1 \le i \le k$, $u + v_i \in S$.

To prove the claim, we start by choosing an arbitrary v_1 from $S_1^* = \{v \in G^* : \overline{d}\big((S_1 + v) \cap S_1\big) > 0\}$ (this set is nonempty by Lemma 2.7(ix)) and put $S_1 = (S - v_1) \cap S$. Clearly, (a) – (d) are satisfied. Having chosen $v_1, \ldots, v_k$ and S_k, define $S_k^* = \{v \in G^* : \overline{d}\big((S_k + v) \cap S_k\big) > 0\}$ and choose v_{k+1} to be an element of $S_k^* \cap \bigcap_{i=1}^k (Q + v_i)$. (This intersection has positive upper density by Lemma 2.7(i) and (ii) because Q has density 1 and S_k^* has positive lower density in G^* by (ix) and (iv).) We then have $v_{k+1} - v_i \in Q$ for $1 \le i \le k$. Next let

$$S_{k+1} = (S_k - v_{k+1}) \cap S_k \cap \Big(\bigcap_{i=1}^{k} \{t \in G : \langle f_{t+v_{k+1}-v_i}, f_t\rangle < \delta\} - v_i\Big).$$

This set has positive upper density because $v_{k+1} \in S_k^*$ and all of the sets in the final intersection have upper density 1. One now sees that for $u \in S_{k+1}$ and $1 \le i < k+1$, $\langle f_{u+v_i}, f_{u+v_{k+1}}\rangle < \delta$. Moreover, for every $u \in S_{k+1}$ one has $u + v_{k+1} \in S_k \subseteq S$. This establishes the claim.

Let k be large enough that $k^2\varepsilon^2 > k(k-1)\delta + k$. Let $u \in S_k$ and put $u_i = u + v_i$, $i = 1, \ldots, k$. Then $\langle f_{u_i}, g\rangle > \varepsilon$, $i = 1, \ldots, k$, and $\langle f_{u_i}, f_{u_j}\rangle < \delta$, $i, j = 1, \ldots, k$, $i \ne j$. For $i = 1, \ldots, k$ write $f_{u_i} = c_i g + f_i$, where $\langle f_i, g\rangle = 0$ and $c_i \ge \varepsilon$. Then

$$\Big\|\sum_{i=1}^{k} f_{u_i}\Big\|^2 = \Big\|\sum_{i=1}^{k} c_i g\Big\|^2 + \Big\|\sum_{i=1}^{k} f_i\Big\|^2 \ge k^2\varepsilon^2.$$

On the other hand,

$$\Big\|\sum_{i=1}^{k} f_{u_i}\Big\|^2 = 2\sum_{1\leq i<j\leq k} \langle f_{u_i}, f_{u_j}\rangle + \sum_{i=1}^{k} \|f_{u_i}\|^2 \leq k(k-1)\delta + k.$$

This is a contradiction. ■

3. Unitary actions

In this chapter, F is a countable field, while V and W are finite dimensional vector spaces over F.

3.1. Given a unitary action of W on a Hilbert space $\mathcal{H}$ and a polynomial $\varphi: V \longrightarrow W$, we call the mapping $U(\varphi) = U \circ \varphi$ a *polynomial unitary action* of V on $\mathcal{H}$.

3.2. Remark. If we are given several commuting actions $U_1, \ldots, U_k$ of vector spaces $W_1, \ldots, W_k$ on a Hilbert space $\mathcal{H}$, and polynomials $\varphi_i: V \longrightarrow W_i$, $i = 1, \ldots, k$, then the polynomial actions $U_1(\varphi_1), \ldots, U_k(\varphi_k)$ can also be obtained as a composition of the single action $U = U_1 \times \ldots \times U_k$ of $W_1 \times \ldots \times W_k$ and the polynomials $(\varphi_1, 0, \ldots, 0), \ldots, (0, 0, \ldots, \varphi_k): V \longrightarrow W_1 \times \ldots \times W_k$. Therefore, when dealing with several commuting polynomial actions of V, we may and will always assume that they are induced by several polynomials $V \longrightarrow W$ and a single unitary action of W.

3.3. Given a unitary action U of an abelian group G on a Hilbert space $\mathcal{H}$, we say that a vector $f \in \mathcal{H}$ is *U-invariant* if $U(v)f = f$ for all $v \in G$, and that f is *U-compact*, or that *U is compact on f*, if the set $U(G)f = \big\{U(v)f : v \in G\big\}$ is precompact (with respect to the strong topology). We say that U is *ergodic* on $\mathcal{H}$ if the set of U-invariant vectors in $\mathcal{H}$ is empty, and that U is *compact* on $\mathcal{H}$ if U is compact on all $f \in \mathcal{H}$. We say that U is *weakly mixing* on $\mathcal{H}$ if the induced action of W on $\mathcal{H} \otimes \overline{\mathcal{H}}$, $U(v)(f \otimes g) = \big(U(v)f\big) \otimes \big(U(v)g\big)$, is ergodic.

3.4. Let U be a unitary action of a (countable) abelian group G on a Hilbert space $\mathcal{H}$. The following three facts are well known.

Theorem. *For any $f \in \mathcal{H}$ one has (along any Følner sequence) $\operatorname*{C-lim}_{u\in G} U(u)f = Pf$, where P is the orthogonal projection of $\mathcal{H}$ onto the space of U-invariant vectors.*

3.5. Theorem. *Let $\mathcal{K}$ be the space consisting of U-invariant vectors in $\mathcal{H} \otimes \overline{\mathcal{H}}$ and let $f \in \mathcal{H}$. Then $f \otimes \overline{f} \perp \mathcal{K}$ iff $f \otimes \overline{g} \perp \mathcal{K}$ for all $g \in \mathcal{H}$ iff $\operatorname*{D-lim}_{u\in G}\big\langle U(u)f, f\big\rangle = 0$ iff $\operatorname*{D-lim}_{u\in G}\big\langle U(u)f, g\big\rangle = 0$ for all $g \in \mathcal{H}$. In particular, U is weakly mixing on $\mathcal{H}$ iff $\operatorname*{D-lim}_{u\in G}\big\langle U(u)f, g\big\rangle = 0$ for all $f, g \in \mathcal{H}$.*

3.6. Define

$$\mathcal{H}^{\mathrm{c}}(U)=\big\{f\in\mathcal{H}:\ U \text{ is compact on } f\big\}\quad\text{and}$$
$$\mathcal{H}^{\mathrm{wm}}(U)=\big\{f\in\mathcal{H}:\ \operatorname*{D-lim}_{u\in G}\langle U(u)f,g\rangle=0 \text{ for all } g\in\mathcal{H}\big\}.$$

Theorem. $\mathcal{H}=\mathcal{H}^{\mathrm{c}}(U)\oplus\mathcal{H}^{\mathrm{wm}}(U)$.

3.7. Our next goal is to show that the theory of compactness/weak mixing for "linear" actions may be transferred without appreciable changes to the case of polynomial actions of finite dimensional vector spaces.

Given a unitary action U of W on a Hilbert space $\mathcal{H}$ and a mapping $\varphi\colon V\longrightarrow W$, we say that a vector $f\in\mathcal{H}$ is $U(\varphi)$-*invariant* if $U(\varphi(u))f=f$ for all $u\in V$, and that f is $U(\varphi)$-*compact*, or that $U(\varphi)$ *is compact on* f, if the set $U(\varphi(V))f=\big\{U(\varphi(v))f: v\in V\big\}$ is precompact. We say that $U(\varphi)$ is *ergodic* on $\mathcal{H}$ if there are no $U(\varphi)$-invariant vectors in $\mathcal{H}$, and that $U(\varphi)$ is *compact* on $\mathcal{H}$ if $U(\varphi)$ is compact on all $f\in\mathcal{H}$. We say that $U(\varphi)$ is *weakly mixing* on $\mathcal{H}$ if the induced action of V on $\mathcal{H}\otimes\overline{\mathcal{H}}$, $U(\varphi(v))(f\otimes g)=\big(U(\varphi(v))f\big)\otimes\big(U(\varphi(v))g\big)$, $v\in V$, is ergodic. We also introduce the following notation:

$$\mathcal{H}^{\mathrm{inv}}(U(\varphi))=\big\{f\in\mathcal{H}: f \text{ is } U(\varphi)\text{-invariant}\big\},$$
$$\mathcal{H}^{\mathrm{c}}(U(\varphi))=\big\{f\in\mathcal{H}: f \text{ is } U(\varphi)\text{-compact}\big\}.$$

It is clear that $\mathcal{H}^{\mathrm{inv}}(U(\varphi))$ and $\mathcal{H}^{\mathrm{c}}(U(\varphi))$ are U-invariant subspaces of $\mathcal{H}$.

3.8. In order to unite the cases $\operatorname{char}F=0$ and $\operatorname{char}F=p$, we introduce the following notation. Let A be a finite subset of W and let $n\in\mathbb{N}$. If $\operatorname{char}F=0$, we define Λ to be the set of all subspaces of $\operatorname{Span}_F(A)$, Ω to be the set of all polynomials $V\longrightarrow W$ with coefficients in $\operatorname{Span}_F(A)$, write $v^*=v$ for $v\in V$ and set $V^*=V$. If $\operatorname{char}F=p<\infty$, we take $m=[\log_p n]$, define Λ to be the set of all (p,A,m)-spaces (see 1.25), Ω to be the set of all polynomials $V\longrightarrow W$ of formal degree $\leq n$ with coefficients in $\operatorname{Span}_{F_m}(A)$, write $v^*=v^{p^m}$ for $v\in V$ and set $V^*=\big\{v^*: v\in V\big\}$. Then, independently of $\operatorname{char}F$, we have the following:

(i) Λ satisfies the ascending chain condition. (If $\operatorname{char}F=0$, this is clear; for $\operatorname{char}F<\infty$ see Lemma 1.28.)

(ii) For any polynomial $\varphi\in\Omega$ one has $\operatorname{Span}\big(\operatorname{Ran}(\varphi-\varphi(0))\big)\in\Lambda$. (If $\operatorname{char}F=0$, this follows from Theorem 1.13; for $\operatorname{char}F<\infty$ see 1.26.)

(iii) Ω is closed under addition and under the taking of derivatives with step $v^*\in V^*$: if $\varphi,\varphi'\in\Omega$ then $\varphi+\varphi'\in\Omega$ and $D_{v^*}\varphi\in\Omega$ for any $v^*\in V^*$. (The latter follows from the linearity of D_v and formula (1.1).)

(iv) If λ is the linear part of a polynomial $\varphi\in\Omega$ then $\operatorname{Span}\big(\operatorname{Ran}(\varphi-\lambda)\big)\not\subseteq L\in\Lambda$ implies $\operatorname{Span}\big(\operatorname{Ran}(D_{v^*}\varphi)\big)=\operatorname{Span}\big(\operatorname{Ran}(D_{v^*}(\varphi-\lambda))\big)\not\subseteq L$ for almost all $v\in V$. (For $\operatorname{char}F<\infty$ this follows from Proposition 1.29; the case $\operatorname{char}F=0$ is left to the reader.)

3.9. Let U be a unitary action of W on $\mathcal{H}$, let $\varphi\colon V \longrightarrow W$ be a polynomial, and let W' be the subgroup of W generated by $\varphi(V)$, i.e. $W' = \mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$. Clearly, the space $\mathcal{H}^{\mathrm{inv}}(U(\varphi))$ coincides with the space of $U_{|W'}$-invariant vectors in $\mathcal{H}$, so $U(\varphi)$ is ergodic if and only if $U_{|W'}$ is ergodic. An analogous statement is true for compact/weakly mixing actions, but this is less obvious; see Theorem 3.17 below.

3.10. The mean ergodic theorem for polynomial actions reads as follows:

Theorem. *Let $U(\varphi)$ be a polynomial unitary action of V on a Hilbert space $\mathcal{H}$. For any $f \in \mathcal{H}$ one has (along any Følner sequence) $\operatorname*{C-lim}_{u\in V} U(\varphi(u))f = Pf$, where P is the orthogonal projection onto $\mathcal{H}^{\mathrm{inv}}(U(\varphi))$. In particular, if $U(\varphi)$ is ergodic then $\operatorname*{C-lim}_{u\in V} U(\varphi(u))f = 0$ for all $f \in \mathcal{H}$.*

The case $V = F$ of this theorem was obtained in [La].

Proof. We use induction on $\mathrm{Deg}\,\varphi$; the degree 1 case is given by Theorem 3.4. Suppose the result is valid for polynomials of degree less than that of φ. It suffices to show that if $U(\varphi)$ is ergodic then $\operatorname*{C-lim}_{u\in V} U(\varphi(u))f = 0$ for all $f \in \mathcal{H}$. Moreover, by a routine application of Zorn's Lemma it is enough to find a nontrivial $U(\varphi)$-invariant subspace $\mathcal{L}$ of $\mathcal{H}$ such that $\operatorname*{C-lim}_{u\in V} U(\varphi(u))f = 0$ for all $f \in \mathcal{L}$. This is what we shall do.

We may assume that $\varphi(0) = 0$. Let $\mathrm{Deg}\,\varphi = n$ and let $A \subset W$ be the set of coefficients of φ. Introduce Λ, Ω and $*$ as in 3.8. Utilizing 3.8(i), choose a maximal element L of Λ (possibly, $L = \{0\}$) with the property that the space $\mathcal{L} = \mathcal{H}^{\mathrm{inv}}(U_{|L})$ is nontrivial. Since $U(\varphi)$ has no invariant vectors, $\mathrm{Ran}(\varphi) \not\subseteq L$. Let λ be the linear part of φ. We consider two cases:

Case 1: $\mathrm{Ran}(\varphi - \lambda) \not\subseteq L$.
By 3.8(ii) and (iii), $D_{v^*}\varphi \in \Omega$, so that $\mathrm{Ran}(D_{v^*}\varphi) \in \Lambda$ for all $v \in V$. By 3.8(iv), $\mathrm{Ran}(D_{v^*}\varphi) \in \Lambda$ for all $v \in V$ and $\mathrm{Ran}(D_{v^*}\varphi) \not\subseteq L$ for almost all $v \in V$. For $v \in V$, if $\mathrm{Ran}(D_{v^*}\varphi) \not\subseteq L$ then $\mathrm{Ran}(D_{v^*}\varphi)$ has no invariant vectors in $\mathcal{L}$, since otherwise we could increase L to $L + \mathrm{Span}\big(\mathrm{Ran}(D_{v^*}\varphi)\big)$. This means that for almost all $v \in V$ the polynomial action $U(D_{v^*}\varphi)$ is ergodic on $\mathcal{L}$. By the induction hypothesis, for almost all $v \in V$ we have $\operatorname*{C-lim}_{u\in V} U(D_{v^*}\varphi(u))f = 0$ for all $f \in \mathcal{L}$, which implies that

$$\begin{aligned}\operatorname*{C-lim}_{u\in V}\big\langle U\big(\varphi(u+v^*)\big)f, U(\varphi(u))f\big\rangle &= \operatorname*{C-lim}_{u\in V}\big\langle U\big(\varphi(u+v^*) - \varphi(u) - \varphi(v^*)\big)f, U(\varphi(v^*))f\big\rangle \\ &= \operatorname*{C-lim}_{u\in V}\big\langle U(D_{v^*}\varphi(u))f, U\big(\varphi(v^*)\big)f\big\rangle = \big\langle \operatorname*{C-lim}_{u\in V} U(D_{v^*}\varphi(u))f, U(\varphi(v^*))f\big\rangle = 0.\end{aligned}$$

Since $V^* = \big\{v^* : v \in V\big\}$ is an infinite subgroup of V, the conclusion of Lemma 2.9 gives $\operatorname*{C-lim}_{u\in V} U(\varphi(u))f = 0$ for all $f \in \mathcal{L}$.

Case 2: $\mathrm{Ran}(\varphi - \lambda) \subseteq L$.
By one of Theorems 1.15 or 1.23, depending on whether $\mathrm{char}\,F = 0$ or $\mathrm{char}\,F = p$, $\mathrm{Ran}(\varphi - \lambda) \subseteq \sum_{v\in V} \mathrm{Span}\big(\mathrm{Ran}(D_{v^*}\varphi)\big)$. It follows that $U(\varphi)_{|\mathcal{L}} = U(\lambda)_{|\mathcal{L}}$. Thus, Theorem 3.4 applies, as $\operatorname*{C-lim}_{u\in V} U(\varphi(u))f = \operatorname*{C-lim}_{u\in V} U(\lambda(u))f = 0$ for all $f \in \mathcal{L}$. ■

3.11. We now turn to compactness and weak mixing.

Lemma. *Let C_1, C_2 be two sets of unitary operators on a Hilbert space $\mathcal{H}$ with $TS = ST$ for all $T \in C_1$, $S \in C_2$, and let $f \in \mathcal{H}$ be such that the sets $C_1 f$, $C_2 f$ are precompact. Then $C_1 C_2 f$ is precompact.*

Proof. For any $\varepsilon > 0$ we can choose $T_1, \dots, T_k \in C_1$ and $S_1, \dots, S_l \in C_2$ such that $\{T_i f\}_{i=1}^k$ is an ε-net in $C_1 f$ and $\{S_j f\}_{j=1}^l$ is an ε-net in $C_2 f$. One may routinely check that $\{T_i S_j f\}_{i,j=1}^{k,l}$ is a 2ε-net for $C_1 C_2 f$. ■

3.12. Corollary. *Let U be a unitary action of W on $\mathcal{H}$ and let $\varphi_1, \varphi_2 \colon V \longrightarrow W$. If both $U(\varphi_1)$ and $U(\varphi_2)$ are compact on $\mathcal{H}$ then $U(\varphi_1 + \varphi_2)$ is compact on $\mathcal{H}$.*

3.13. Theorem. *Fix a Følner sequence for V. Let $U(\varphi)$ be a polynomial action of V on a Hilbert space $\mathcal{H}$, let $\mathcal{K}$ be the space of $U(\varphi)$-invariant vectors in $\mathcal{H} \otimes \overline{\mathcal{H}}$ and let $f \in \mathcal{H}$. Then $f \otimes \overline{f} \perp \mathcal{K}$ iff $f \otimes \overline{g} \perp \mathcal{K}$ for all $g \in \mathcal{H}$ iff $\operatorname*{D-lim}_{u \in V} \langle U(\varphi(u))f, f \rangle = 0$ iff $\operatorname*{D-lim}_{u \in V} \langle U(\varphi(u))f, g \rangle = 0$ for all $g \in \mathcal{H}$. In particular, $U(\varphi)$ is weakly mixing on $\mathcal{H}$ if and only if $\operatorname*{D-lim}_{u \in V} \langle U(\varphi(u))f, g \rangle = 0$ for all $f, g \in \mathcal{H}$.*

Proof. We use the fact that for a bounded set $\{a_u\}_{u \in V}$ of real numbers, $\operatorname*{D-lim}_{u \in V} a_u = 0$ if and only if $\operatorname*{C-lim}_{u \in V} |a_u| = 0$ if and only if $\operatorname*{C-lim}_{u \in V} |a_u|^2 = 0$. Suppose $\operatorname*{D-lim}_{u \in V} \langle U(\varphi(u))f, f \rangle = 0$. Letting P be the projection onto the space of $U(\varphi)$-invariant vectors for the induced action on $\mathcal{H} \otimes \overline{\mathcal{H}}$, we have for arbitrary $g \in \mathcal{H}$,

$$\begin{aligned}\|P(f \otimes \overline{g})\|^2 = \operatorname*{C-lim}_{u \in V} \langle U(\varphi(u))(f \otimes \overline{g}), f \otimes \overline{g} \rangle &= \operatorname*{C-lim}_{u \in V} \Big(\langle U(\varphi(u))f, f \rangle \cdot \overline{\langle U(\varphi(u))g, g \rangle} \Big) \\ &\leq \operatorname*{C-lim}_{u \in V} \|g\|^2 \Big(\operatorname*{C-lim}_{u \in V} \big| \langle U(\varphi(u))f, f \rangle \big| \Big) = 0.\end{aligned}$$

Suppose next that $f \otimes \overline{f} \perp \mathcal{K}$. Then

$$\begin{aligned}\operatorname*{C-lim}_{u \in V} \big| \langle U(\varphi(u))f, g \rangle \big|^2 &= \operatorname*{C-lim}_{u \in V} \Big(\langle U(\varphi(u))f, g \rangle \cdot \overline{\langle U(\varphi(u))f, g \rangle} \Big) \\ &= \operatorname*{C-lim}_{u \in V} \langle U(\varphi(u))(f \otimes \overline{f}), g \otimes \overline{g} \rangle = \langle P(f \otimes \overline{f}), g \otimes \overline{g} \rangle = 0.\end{aligned}$$

■

3.14. Lemma. *Let U be a unitary action of W on $\mathcal{H}$ and let φ_1, φ_2 be polynomials $V \longrightarrow W$. If $U(\varphi_1)$ is compact on $\mathcal{H}$ and $U(\varphi_2)$ is weakly mixing on $\mathcal{H}$, then $U(\varphi_1 + \varphi_2)$ is weakly mixing on $\mathcal{H}$.*

Proof. Let $f, g \in \mathcal{H}$. Let $\{f_1, \dots, f_k\}$ be an ε-net for $\{U(\varphi_1(u))f : u \in V\}$. Then for all $u \in V$ we have

$$\big| \langle U\big((\varphi_1 + \varphi_2)(u)\big)f, g \rangle \big| \leq \varepsilon \|g\| + \sum_{i=1}^{k} \big| \langle U(\varphi_2(u))f_i, g \rangle \big|.$$

Since ε is arbitrary and $\operatorname*{D-lim}_{u \in V} \big| \langle U(\varphi_2(u))f_i, g \rangle \big| = 0$, $1 \leq i \leq k$, we are done. ■

3.15. In light of Theorem 3.13 we introduce the following terminology. Given a unitary action U of W on $\mathcal{H}$ and a mapping $\varphi: V \longrightarrow W$, we say that $U(\varphi)$ is *weakly mixing on* $f \in \mathcal{H}$ if $\operatorname*{D\text{-lim}}_{u\in V}\langle U(\varphi(u))f, g\rangle = 0$ for all $g \in \mathcal{H}$. (Note that this property is independent of the Følner sequence chosen in Theorem 3.13.) We say that $U(\varphi)$ is *weakly mixing on* $\mathcal{L} \subseteq \mathcal{H}$ if $U(\varphi)$ is weakly mixing on all $f \in \mathcal{L}$. Finally we define

$$\mathcal{H}^{\mathrm{wm}}(U(\varphi)) = \big\{f \in \mathcal{H} : U(\varphi) \text{ is weakly mixing on } f\big\}.$$

We remark that $\mathcal{H}^{\mathrm{wm}}(U(\varphi))$ is a U-invariant subspace of $\mathcal{H}$.

3.16. Lemma. *Let U be a unitary action of W on $\mathcal{H}$. For any mapping $\varphi: V \longrightarrow W$, $\mathcal{H}^{\mathrm{c}}(U(\varphi)) \perp \mathcal{H}^{\mathrm{wm}}(U(\varphi))$.*

Proof. Let $f \in \mathcal{H}^{\mathrm{c}}(U(\varphi))$ and $g \in \mathcal{H}^{\mathrm{wm}}(U(\varphi))$. Let $\varepsilon > 0$ be arbitrary and let $f_1, \ldots, f_k$ be an ε-net for $\big\{U(\varphi(u))f : u \in V\big\}$. Fix a Følner sequence and choose i, $1 \le i \le k$, such that $E_i = \big\{u \in V : \|U(\varphi(u))f - f_i\| < \varepsilon\big\}$ satisfies $\overline{d}(E_i) > 0$.

Let $F = \big\{u \in V : |\langle U(\varphi(u))g, f_i\rangle| > \varepsilon\big\}$. Then $\overline{d}(F) = 0$. Let $u \in E_i \setminus F$. We have

$$\begin{aligned}\langle f, g\rangle = \big\langle U(\varphi(u))f, U(\varphi(u))g\big\rangle &\le \big|\big\langle U(\varphi(u))f - f_i, U(\varphi(u))g\big\rangle\big| + \big|\big\langle f_i, U(\varphi(u))g\big\rangle\big| \\ &\le \|U(\varphi(u))f - f_i\| \cdot \|g\| + \varepsilon \le \varepsilon(\|g\| + 1).\end{aligned}$$

As ε is arbitrary, $\langle f, g\rangle = 0$. ■

3.17. Theorem. *Let U be a unitary action of W on a Hilbert space $\mathcal{H}$, let $\varphi: V \longrightarrow W$ be a polynomial with zero constant term and let $W' = \operatorname{Span}\big(\operatorname{Ran}(\varphi)\big)$. Then*

(i) $\mathcal{H}^{\mathrm{c}}(U(\varphi)) = \mathcal{H}^{\mathrm{c}}(U_{|W'})$;

(ii) $\mathcal{H}^{\mathrm{wm}}(U(\varphi)) = \mathcal{H}^{\mathrm{wm}}(U_{|W'})$;

(iii) $\mathcal{H} = \mathcal{H}^{\mathrm{wm}}(U(\varphi)) \oplus \mathcal{H}^{\mathrm{c}}(U(\varphi))$.

Proof. In fact, the weak mixing/compactness dichotomy is a corollary of the mean ergodic theorem, applied to the induced diagonal action on $\mathcal{H} \otimes \overline{\mathcal{H}}$, however we give a direct proof, analogous to that of Theorem 3.10. (i) is a direct consequence of Lemma 3.11. (ii) is corollary of (i), (iii) and Theorem 3.6. To prove (iii), we use induction on $\operatorname{Deg}\varphi$. The degree 1 case is Theorem 3.6. The orthogonality of $\mathcal{H}^{\mathrm{wm}}(U(\varphi))$ and $\mathcal{H}^{\mathrm{c}}(U(\varphi))$ was established in Lemma 3.16. Hence it suffices to show that $\mathcal{H}^{\mathrm{c}}(U(\varphi))^{\perp} \subseteq \mathcal{H}^{\mathrm{wm}}(U(\varphi))$, that is, that $U(\varphi)$ is weakly mixing on $\mathcal{H}^{\mathrm{c}}(U(\varphi))^{\perp}$. Denote $\mathcal{H}^{\mathrm{c}}(U(\varphi))^{\perp}$ by $\mathcal{K}$. We must show that $U(\varphi)$ is weakly mixing on $\mathcal{K}$ under the assumption that $U(\varphi)$ has no compact vectors in $\mathcal{K}$. Again, by a routine application of Zorn's lemma, it is enough to find a nontrivial $U(\varphi)$-invariant subspace $\mathcal{L}$ of $\mathcal{K}$ such that $U(\varphi)$ is weakly mixing on $\mathcal{L}$.

Let $\operatorname{Deg}\varphi = n$ and let $A \subset W$ be the set of coefficients of φ. Let Λ, Ω and $*$ be as in 3.8. Utilizing 3.8(i), choose a maximal element L of Λ (possibly, $L = \{0\}$) with the property that the space $\mathcal{L} = \mathcal{K}^{\mathrm{c}}(U_{|L})$ is nontrivial. Since $U(\varphi)$ has no compact vectors in $\mathcal{K}$, $\operatorname{Ran}(\varphi) \not\subseteq L$. Let λ be the linear part of φ. We have two cases:

Case 1: $\mathrm{Ran}(\varphi-\lambda)\not\subseteq L$.
By 3.8(iv), $\mathrm{Ran}(D_{v^*}\varphi)\not\subseteq L$ for almost all $v\in V$. For $v\in V$ let $W'_v=\mathrm{Span}\big(\mathrm{Ran}(D_{v^*}\varphi)\big)$. By 3.8(ii) and (iii), $D_{v^*}\varphi\in\Omega$, so that $W'_v\in\Lambda$, $v\in V$. Thus, if $\mathrm{Ran}(D_{v^*}\varphi)\not\subseteq L$, then $D_{v^*}\varphi$ cannot have compact vectors in $\mathcal{L}$, as otherwise we could increase L to $L+W'_v$. We have, therefore, that for almost all $v\in V$, $U(D_{v^*}\varphi)$ has no compact vectors in $\mathcal{L}$. By the induction hypothesis, $U(D_{v^*}\varphi)$ is weakly mixing on $\mathcal{L}$ for almost all $v\in V$, and by Theorem 3.13, for these v we have $\operatorname*{D-lim}_{u\in V}\big\langle U(D_{v^*}\varphi(u))f,g\big\rangle=0$ for all $f,g\in\mathcal{L}$. Hence

$$\begin{aligned}\operatorname*{D-lim}_{u\in V}\big\langle U\big(\varphi(u+v^*)\big)f,U(\varphi(u))f\big\rangle&=\operatorname*{D-lim}_{u\in V}\big\langle U\big(\varphi(u+v^*)-\varphi(u)-\varphi(v^*)\big)f,U(\varphi(v^*))f\big\rangle\\&=\operatorname*{D-lim}_{u\in V}\big\langle U(D_{v^*}\varphi(u))f,U(\varphi(v^*))f\big\rangle=0\end{aligned}$$

for almost all $v\in V$. The conclusion of Lemma 2.10 therefore holds, namely that $\operatorname*{D-lim}_{u\in V}\big\langle U(\varphi(u))f,g\big\rangle=0$ for $f,g\in\mathcal{L}$.
Case 2: $\mathrm{Ran}(\varphi-\lambda)\subseteq L$.
Since $\mathrm{Span}\big(\mathrm{Ran}(D_{v^*}\varphi)\big)=\mathrm{Span}\big(\mathrm{Ran}(D_{v^*}(\varphi-\lambda))\big)\subseteq\mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$, $U(D_{v^*}\varphi)$ are compact on $\mathcal{L}$ for all $v\in V$. By (i), $U_{|W'_v}$ is compact on $\mathcal{L}$ for all $v\in V$. By either Theorem 1.15 or Theorem 1.23, applied to the polynomial $\varphi-\lambda$, there exist $v_1,\ldots,v_r\in V$ such that $\mathrm{Span}\big(\mathrm{Ran}(\varphi-\lambda)\big)=\sum_{s=1}^r W'_{v_s}$. So, by Lemma 3.11, $U(\varphi-\lambda)$ is compact on $\mathcal{L}$. If $U(\lambda)$ had a compact vector in $\mathcal{L}$, by Lemma 3.11 it would also be compact for $U_{|W'}$, a contradiction. Hence, by Theorem 3.6, $U(\lambda)$ is weakly mixing on $\mathcal{L}$, and therefore by Lemma 3.14, $U(\varphi)$ is weakly mixing on $\mathcal{L}$. ■

3.18. We will need a "relative" version of Theorem 3.6, where one deals with a Hilbert space "over" a measure space Y instead of a conventional Hilbert space. We will exposit the corresponding theory in the language of *Y-Hilbert spaces* (see [Le]). One could use the language of *Hilbert bundles* instead (see, for example, [Z1], [Z2]). Yet another possibility is to confine ourselves to $L^\infty(Y)$-submodules of the Hilbert space $L^2(X)$, where X is an extension of Y (see also 4.2 below); in this case we address the reader to [F1], [FK3] or [BMZ]. The proofs of the propositions below are "relativizations" of the proofs of the corresponding propositions 3.5 – 3.17 and we omit them.

3.19. Let $\mathbf{Y}=(Y,\mathcal{D},\nu)$ be a probability measure space. When it does not lead to confusion, we will identify $\mathbf{Y}$ and Y. A Y-pre-Hilbert space $\mathcal{H}$ is a $L^\infty(Y)$-module with a sesquilinear mapping $\langle\,,\rangle:\mathcal{H}\times\mathcal{H}\longrightarrow L^1(Y)$, the value of $\langle f,g\rangle$, $f,g\in\mathcal{H}$, at $y\in Y$ being $\langle f,g\rangle_y$, which satisfies $\langle f,f\rangle_y\geq 0$ for almost all $y\in Y$ and all $f\in\mathcal{H}$. We put $\|f\|_y=\sqrt{\langle f,f\rangle_y}$. A Y-pre-Hilbert space inherits a pre-Hilbert space structure by $\langle f,g\rangle_Y=\int\langle f,g\rangle_y\,d\nu$; we say that $\mathcal{H}$ is *a Y-Hilbert space* if $\mathcal{H}$ is a Hilbert space with respect to $\langle\,,\rangle_Y$.

3.20. Algebraic operations on Y-Hilbert spaces are naturally defined. In particular, the tensor product $\mathcal{H}_1\otimes_Y\mathcal{H}_2$ of two Y-Hilbert spaces is the Y-Hilbert space which is obtained as the completion of the algebraic tensor product $\mathcal{H}_1\otimes_{L^\infty(Y)}\mathcal{H}_2$. *The complex conjugate* $\overline{\mathcal{H}}$ of a Y-Hilbert space $\mathcal{H}$ is defined by $\overline{\mathcal{H}}=\big\{\bar f: f\in\mathcal{H}\big\}$, $h\bar f=\overline{hf}$ for $h\in L^\infty(Y)$, and $\langle\bar f,\bar g\rangle=\overline{\langle f,g\rangle}$.

3.21. *A Y-unitary operator T* on a Y-Hilbert space $\mathcal{H}$ is an invertible measure preserving transformation $T: Y \longrightarrow Y$ and an invertible linear transformation $T: \mathcal{H} \longrightarrow \mathcal{H}$ satisfying $\langle Tf, Tg\rangle = T\langle f, g\rangle$ and $T(hf) = ThTf$ for any $f, g \in \mathcal{H}$ and $h \in L^\infty(Y)$. Note that such T is also a unitary operator with respect to the Hilbert structure of $\mathcal{H}$. *A Y-unitary action* U of a group G on a Y-Hilbert space $\mathcal{H}$ is a homomorphism U of G into the group of Y-unitary operators on $\mathcal{H}$.

If V, W are finite dimensional vector spaces, $\varphi: V \longrightarrow W$ is a polynomial and U is a Y-unitary action of W on $\mathcal{H}$, we call the mapping $U(\varphi)$ a *polynomial Y-unitary action* of V on $\mathcal{H}$.

3.22. Given $\varepsilon > 0$, we say that a set $Q \subseteq \mathcal{H}$ is an *ε-net for $f \in \mathcal{H}$ over Y* if $\int \inf_{g\in Q} \|f - g\|_y^2 \, d\nu < \varepsilon^2$. We say that a set $Q \subseteq \mathcal{H}$ is *an ε-net for $C \subseteq \mathcal{H}$ over Y* if Q is an ε-net over Y for all $f \in C$. We say that a set $C \subseteq \mathcal{H}$ is *Y-precompact* if for any $\varepsilon > 0$ there exists a finite ε-net for C over Y.

3.23. We will need the following fact:

Lemma. *Let $\{g_1, \ldots, g_k\}$ be an ε-net for f over Y. Then for any $g \in \mathcal{H}$,*

$$\big\|\langle f, g\rangle\big\|_{L^1(Y)} \le \sum_{i=1}^{k} \big\|\langle g_k, g\rangle\big\|_{L^1(Y)} + \varepsilon\|g\|_Y.$$

Proof. For $y \in Y$, let $i(y)$ be the first i, $1 \le i \le k$, such that $\|f - g_{i(y)}\|_y$ is minimized. Then

$$\int \big|\langle f, g\rangle_y\big| d\nu - \sum_{i=1}^{k} \int \big|\langle g_k, g\rangle_y\big| d\nu \le \int \big|\langle f, g\rangle_y\big| d\nu - \int \big|\langle g_{i(y)}, g\rangle_y\big| d\nu \le \int \big|\langle f - g_{i(y)}, g\rangle_y\big| d\nu$$
$$\le \int \|f - g_{i(y)}\|_y \cdot \|g\|_y d\nu \le \Big(\int \|f - g_{i(y)}\|_y^2 d\nu\Big)^{1/2} \Big(\int \|g\|_y^2 d\nu\Big)^{1/2} \le \varepsilon \Big(\int \|g\|_y^2 d\nu\Big)^{1/2}.$$

■

3.24. Lemma. *Let C_1, C_2 be two sets of Y-unitary operators on a Y-Hilbert space $\mathcal{H}$ with $TS = ST$ for all $T \in C_1$, $S \in C_2$, and let $f \in \mathcal{H}$ be such that the sets $C_1 f$, $C_2 f$ are Y-precompact. Then $C_1 C_2 f$ is Y-precompact.*

3.25. Let U be a mapping of a group G into the group of Y-unitary operators on a Y-Hilbert space $\mathcal{H}$. We say that U is *compact on $f \in \mathcal{H}$ relative to Y* or that f is *compact relative to Y (with respect to U)* if the set $U(G)f$ is precompact relative to Y. We say that U is *compact on $\mathcal{H}$ relative to Y* if U is compact relative to Y on all $f \in \mathcal{H}$.

3.26. Corollary of Lemma 3.24. *Let φ_1, φ_2 be mappings of a group V into an abelian group G and let U be a Y-unitary action of G on a Y-Hilbert space $\mathcal{H}$. If both $U(\varphi_1)$ and $U(\varphi_2)$ are compact on $\mathcal{H}$ relative to Y then $U(\varphi_1 + \varphi_2)$ is compact on $\mathcal{H}$ relative to Y.*

3.27. We say that a Y-unitary action of a group G on a Y-Hilbert space $\mathcal{H}$ is *weakly mixing relative to* Y if the induced unitary action of G on the Hilbert space $\mathcal{H} \otimes_Y \overline{\mathcal{H}}$ is ergodic.

Theorem. *Let $\mathcal{K}$ be the space of U-invariant vectors in $\mathcal{H} \otimes_Y \overline{\mathcal{H}}$ and let $f \in \mathcal{H}$. Then $f \otimes \overline{f} \perp \mathcal{K}$ iff $f \otimes \overline{g} \perp \mathcal{K}$ for all $g \in \mathcal{H}$ iff $\underset{u \in G}{\text{D-lim}} \langle U(u)f, f \rangle = 0$ in $L^1(Y)$ iff $\underset{u \in G}{\text{D-lim}} \langle U(u)f, g \rangle = 0$ for all $g \in \mathcal{H}$. In particular, U is weakly mixing relative to Y if and only if $\underset{u \in G}{\text{D-lim}} \langle U(u)f, g \rangle = 0$ in $L^1(Y)$ for all $f, g \in \mathcal{H}$.*

3.28. Let U be a Y-unitary action of an abelian group G on a Y-Hilbert space $\mathcal{H}$. For $f \in \mathcal{H}$ we say that U *is weakly mixing on f relative to Y* if $\underset{u \in G}{\text{D-lim}} \langle U(u)f, g \rangle = 0$ in $L^1(Y)$ for all $g \in \mathcal{H}$, and define

$$\mathcal{H}_Y^{\text{c}}(U) = \{ f \in \mathcal{H} : \ U \text{ is compact on } f \text{ relative to } Y \},$$
$$\mathcal{H}_Y^{\text{wm}}(U) = \{ f \in \mathcal{H} : \ U \text{ is weakly mixing on } f \text{ relative to } Y \}.$$

Theorem. $\mathcal{H} = \mathcal{H}_Y^{\text{c}}(U) \oplus \mathcal{H}_Y^{\text{wm}}(U)$.

3.29. We say that a polynomial Y-unitary action of a vector space V on a Y-Hilbert space $\mathcal{H}$ is *weakly mixing relative to* Y if the induced polynomial unitary action of V on the Hilbert space $\mathcal{H} \otimes_Y \overline{\mathcal{H}}$ is ergodic.

Theorem. *Let $\mathcal{K}$ be the space of $U(\varphi)$-invariant vectors in $\mathcal{H} \otimes_Y \overline{\mathcal{H}}$ and let $f \in \mathcal{H}$. Then $f \otimes \overline{f} \perp \mathcal{K}$ iff $f \otimes \overline{g} \perp \mathcal{K}$ for all $g \in \mathcal{H}$ iff $\underset{u \in V}{\text{D-lim}} \langle U(\varphi(u))f, f \rangle = 0$ iff $\underset{u \in V}{\text{D-lim}} \langle U(\varphi(u))f, g \rangle = 0$ for all $g \in \mathcal{H}$. In particular, $U(\varphi)$ is weakly mixing on $\mathcal{H}$ relative to Y if and only if $\underset{u \in V}{\text{D-lim}} \langle U(\varphi(u))f, g \rangle = 0$ in $L^1(Y)$ for all $f, g \in \mathcal{H}$.*

3.30. Lemma. *Let U be a Y-unitary action of W on a Y-Hilbert space $\mathcal{H}$ and let φ_1, φ_2 be polynomials $V \longrightarrow W$. If $U(\varphi_1)$ is compact on $\mathcal{H}$ relative to Y and $U(\varphi_2)$ is weakly mixing on $\mathcal{H}$ relative to Y, then $U(\varphi_1 + \varphi_2)$ is weakly mixing on $\mathcal{H}$ relative to Y.*

3.31. We say that a set Q in a Y-Hilbert space $\mathcal{H}$ is *Y-bounded* if there exists $b \in \mathbb{R}$ such that for every $f \in Q$, $\|f\|_y < b$ for almost all $y \in Y$. The following is the "Y-analogue" of Lemma 2.10.

Lemma. *Let G be an infinite abelian group and let $\{f_u\}_{u \in G}$ be a Y-bounded set in a Y-Hilbert space $\mathcal{H}$ indexed by the elements of G. Assume that there exists an infinite subgroup G^* of G such that $\underset{v^* \in G^*}{\text{D-lim}} \ \underset{u \in G}{\text{D-limsup}} \big\| \langle f_{u+v^*}, f_u \rangle \big\|_{L^1(Y)} = 0$. Then $\underset{u \in G}{\text{D-lim}} \langle f_u, g \rangle = 0$ in $L^1(Y)$.*

3.32. Let $U(\varphi)$ be a polynomial Y-unitary action of V on a Y-Hilbert space $\mathcal{H}$. For $f \in \mathcal{H}$ we say that $U(\varphi)$ is *weakly mixing on f relative to Y* if $\underset{u \in V}{\text{D-lim}} \langle U(\varphi(u))f, g \rangle = 0$ for all $g \in \mathcal{H}$, and define

$$\mathcal{H}_Y^{\text{c}}(U(\varphi)) = \{ f \in \mathcal{H} : \ U(\varphi) \text{ is compact on } f \text{ relative to } Y \},$$
$$\mathcal{H}_Y^{\text{wm}}(U(\varphi)) = \{ f \in \mathcal{H} : \ U(\varphi) \text{ is weakly mixing on } f \text{ relative to } Y \}.$$

Theorem. *Let U be a Y-unitary action of W on a Y-Hilbert space $\mathcal{H}$, let $\varphi: W \longrightarrow V$ be a polynomial with zero constant term and let $W' = \operatorname{Span}(\operatorname{Ran}(\varphi))$. Then*

(i) $\mathcal{H}_Y^{\mathrm{c}}(U(\varphi)) = \mathcal{H}_Y^{\mathrm{c}}(U_{|W'})$;

(ii) $\mathcal{H}_Y^{\mathrm{wm}}(U(\varphi)) = \mathcal{H}_Y^{\mathrm{wm}}(U_{|W'})$;

(iii) $\mathcal{H} = \mathcal{H}_Y^{\mathrm{wm}}(U(\varphi)) \oplus \mathcal{H}_Y^{\mathrm{c}}(U(\varphi))$.

The proof is completely analogous to the proof of Theorem 3.17.

4. Measure preserving actions and proof of main theorem

4.1. A mapping $\pi: X \longrightarrow Y$ of a probability measure space $(X, \mathcal{B}, \mu)$ into a probability measure space $(Y, \mathcal{D}, \nu)$ is called a *factor map* if $\pi^{-1}(\mathcal{D}) \subseteq \mathcal{B}$ and for any $D \in \mathcal{D}$, $\mu(\pi^{-1}(D)) = \nu(D)$. If this is the case, X is called an *extension of* Y and Y is called a *factor of* X. In a slight abuse of terminology we will often refer to the factor map itself as an "extension". A factor map π induces an isometry homomorphism $\pi^*: L^2(Y) \longrightarrow L^2(X)$ by $\pi^* f(x) = f(\pi(x))$; we will identify $L^2(Y)$ with $\pi^*(L^2(Y)) \subseteq L^2(X)$.

If $\mathcal{B}'$ is a sub-σ-algebra of $\mathcal{B}$, the identity mapping $X \longrightarrow X$ defines an extension $(X, \mathcal{B}, \mu) \longrightarrow (X, \mathcal{B}', \mu)$; we will call $\mathbf{X}' = (X, \mathcal{B}', \mu)$ an *inner factor* of $(X, \mathcal{B}, \mu)$.

If X is an extension of X' and X' is an extension of Y, then X is an extension of Y with X' being a *subextension* of X.

4.2. Let $(X, \mathcal{B}, \mu) \longrightarrow (Y, \mathcal{D}, \nu)$ be an extension. For $f \in L^2(X)$ let $E_Y(f)$ denote the *conditional expectation of f given Y*, namely the projection of f onto $L^2(Y)$. The homomorphism $E_Y: L^2(X) \longrightarrow L^2(Y)$ maps $L^\infty(X)$ onto $L^\infty(Y)$ and is extendible by continuity to an epimorphism $L^1(X) \longrightarrow L^1(Y)$. Putting $\langle f, g\rangle = E_Y(f\bar{g}) \in L^1(Y)$ for $f, g \in L^2(X)$ turns $L^2(X)$ into a Y-Hilbert space (see 3.19 above).

4.3. *A transformation T of an extension* $\pi: X \longrightarrow Y$ is a pair (T_X, T_Y), where T_X and T_Y are measure preserving transformations of X and Y respectively, satisfying $\pi \circ T_X = T_Y \circ \pi$. Depending on the context, we may use the symbol T to denote either T_X or T_Y.

If T is a measure preserving transformation of a measure space $\mathbf{X} = (X, \mathcal{B}, \mu)$ and an inner factor $\mathbf{Y} = (X, \mathcal{B}', \mu)$ of $\mathbf{X}$ is such that $L^2(\mathbf{Y})$ is T-invariant, then T induces a transformation of the extension $\mathbf{X} \longrightarrow \mathbf{Y}$.

4.4. Let T be an invertible transformation of an extension $X \longrightarrow Y$. T induces an automorphism of $L^1(X)$, $(Tf)(x) = f(Tx)$, which we will identify by the same symbol T. T preserves multiplication, $T(hf) = T(h)T(f)$ for $f \in L^1(X)$, $h \in L^\infty(X)$. $T_{|L^2(X)}$ is a unitary operator on the Hilbert space $L^2(X)$; this operator will also be denoted by T. $L^2(Y)$ is invariant under T, therefore $L^2(Y)^\perp$ is also invariant, so that $E_Y(Tf) = T(E_Y f)$ for all $f \in L^2(X)$. This implies $E_Y(Th) = T(E_Y h)$ for all $h \in L^1(X)$; in particular, $E_Y(Tf\overline{Tg}) = TE_Y(f\bar{g})$ for $f, g \in L^2(X)$. It follows that, in terms of 3.21, T is a Y-unitary operator on the Y-Hilbert space $L^2(X)$.

4.5. *A measure preserving action* U of a group G on an extension $X \longrightarrow Y$ is a homomorphism of G into the group of invertible transformations of $X \longrightarrow Y$. Such an action defines a Y-unitary action of G on the Y-Hilbert space $L^2(X)$, which we will identify by the same symbol U.

Let U be a unitary action of a group G on an extension $X \longrightarrow Y$. We say that U is *compact on X relative to Y* if the corresponding action of G on $L^2(X)$ is compact relative to Y. Note that U is always compact on $L^2(Y)$ relative to Y; we say that U is *weakly mixing on X relative to Y* if U is weakly mixing relative to Y on the orthocomplement $L^2(Y)^\perp$ of $L^2(Y)$ in $L^2(X)$.

If U is a unitary action of a vector space W on an extension $X \longrightarrow Y$ and $\varphi\colon V \longrightarrow W$ is a polynomial, we call $U(\varphi)$ a *polynomial action of V on $X \longrightarrow Y$*. We say that $U(\varphi)$ is *compact on X relative to Y* if $U(\varphi)$ is compact relative to Y on $L^2(X)$, and that $U(\varphi)$ is *weakly mixing relative to Y* if $U(\varphi)$ is weakly mixing relative to Y on $L^2(Y)^\perp$ (see 3.27).

4.6. Let $U(\varphi)$ be a polynomial action on an extension $(X,\mathcal{B},\mu) \longrightarrow (Y,\mathcal{D},\nu)$; denote $L^2(X)$ by $\mathcal{H}$. One can check that for any $f,g \in \mathcal{H}_Y^{\mathrm{c}}(U(\varphi))$, one has $\max\{f,g\} \in \mathcal{H}_Y^{\mathrm{c}}(U(\varphi))$. It follows that $\mathcal{H}_Y^{\mathrm{c}}(U(\varphi)) = L^2(\mathbf{X}')$ for some inner factor $\mathbf{X}' = (X,\mathcal{B}',\mu)$ of $(X,\mathcal{B},\mu)$. Since $L^2(Y) \subseteq \mathcal{H}_Y^{\mathrm{c}}(U(\varphi))$, $\mathbf{X}'$ is an extension of Y. Since $\mathcal{H}_Y^{\mathrm{c}}(U(\varphi))$ is U-invariant, the action U and the polynomial action $U(\varphi)$ are defined on $\mathbf{X}'$. (See [F2], page 137 for more details.)

4.7. We fix now a vector space W over a countable field F. Let A be a finite subset of W and let $n \in \mathbb{N}$; define Λ, Ω and $*$ as in 3.8. Let U be an action of W on an extension $X \longrightarrow Y$. We say that $X \longrightarrow Y$ is *primitive* (with respect to U, n and A) if there exists a space $L \in \Lambda$ such that, for any $L' \in \Lambda$, $U_{|L'}$ is compact on X relative to Y if $L' \subseteq L$ and is weakly mixing on X relative to Y otherwise.

If $X \longrightarrow Y$ is a primitive extension then, by Theorem 3.32, for any $\varphi \in \Omega$, $U(\varphi)$ is either compact relative to Y or weakly mixing relative to Y: if $\mathrm{Ran}(\varphi - \varphi(0)) \subseteq L$, $U(\varphi)$ is compact; if $\mathrm{Ran}(\varphi - \varphi(0)) \not\subseteq L$, $U(\varphi)$ is weakly mixing.

4.8. Theorem. *Any nontrivial extension has a nontrivial primitive subextension.*

Proof. Let $X \longrightarrow Y$ be a nontrivial extension; put $\mathcal{H} = L^2(X)$. Utilizing 3.8(i), choose a maximal element L of Λ (possibly, $L = \{0\}$) with the property that $\mathcal{H}_Y^{\mathrm{c}}\big(U_{|L}\big) \neq L^2(Y)$. Denote $\mathcal{H}_Y^{\mathrm{c}}\big(U_{|L}\big)$ by $\mathcal{L}$. By 4.6, $\mathcal{L} = L^2(X')$ for some subextension $X' \longrightarrow Y$ of $X \longrightarrow Y$. Since $\mathcal{L}$ is U-invariant, the action of U on X' is defined. For any $L' \subseteq L$, $U_{|L'}$ is compact on X' relative to $\mathbf{Y}$. If $L' \in \Lambda$ and $L' \not\subseteq L$, then there cannot be a vector in $\mathcal{L} \setminus L^2(Y)$ which is compact for $U_{|L'}$ relative to Y, since otherwise we could extend L to $L + L'$. Therefore, by Theorem 3.32, $U_{|L'}$ is weakly mixing on X' relative to Y in this case. Hence, $X' \longrightarrow Y$ is a primitive extension. ■

4.9. Now let $\mathcal{P}$ be a finite family of polynomials from a finite dimensional space V into W, let A be the set of all coefficients of the elements of $\mathcal{P}$ and let $n = \max\big\{\deg\varphi : \varphi \in P\big\}$. Let U be an action of W on an extension $X \longrightarrow Y$, and assume that $X \longrightarrow Y$ is primitive with respect to U, n and A. Let $\mathcal{P} = \bigcup_{i=1}^{l} \mathcal{P}_i$ be the partition of $\mathcal{P}$ into the classes of elements whose differences are compact relative to Y. Thus, for any i and $\varphi_1, \varphi_2 \in \mathcal{P}_i$, $U(\varphi_1 - \varphi_2)$ is compact on X relative to Y (this is equivalent to $\mathrm{Ran}\big(\varphi_1 - \varphi_2 - \varphi_1(0) + \varphi_2(0)\big) \subseteq L$) and for any $i_1 \neq i_2$, $\varphi_1 \in \mathcal{P}_{i_1}$ and $\varphi_2 \in \mathcal{P}_{i_2}$, $U(\varphi_1 - \varphi_2)$ is weakly mixing on X relative to

Y (this is equivalent to $\mathrm{Ran}\big(\varphi_1-\varphi_2-\varphi_1(0)+\varphi_2(0)\big)\not\subseteq L$).

Theorem. *For any set of functions $\{f_\varphi\}_{\varphi\in\mathcal{P}}\subset L^\infty(X)$,*

$$\operatorname*{D-lim}_{u\in V}\Big(E_Y\big(\prod_{\varphi\in\mathcal{P}}U(\varphi(u))f_\varphi\big)-\prod_{i=1}^{l}E_Y\big(\prod_{\varphi\in\mathcal{P}_i}U(\varphi(u))f_\varphi\big)\Big)=0\quad \mathit{in}\ L^1(Y).$$

4.10. We start with the following lemma.

Lemma. *Let $(X,\mathcal{B},\mu)\longrightarrow(Y,\mathcal{D},\nu)$ be an extension, let $C_1,\ldots,C_k\subset L^\infty(X)$ be Y-precompact sets of uniformly bounded functions on X and let $T_1,\ldots,T_k$ be mappings of an abelian group V into the group of transformations of $\mathbf{X}\longrightarrow\mathbf{Y}$ satisfying*

$$\operatorname*{D-lim}_{u\in V}\Big(E_Y\big(\prod_{j=1}^{k}T_j(u)f_j\big)-\prod_{j=1}^{k}E_Y\big(T_j(u)f_j\big)\Big)=0\quad \mathit{in}\ L^1(Y) \tag{4.1}$$

for all $f_j\in L^\infty(X)$, $j=1,\ldots,k$. Then for any mappings $\alpha_j\colon V\longrightarrow C_j$, $j=1,\ldots,k$, one has

$$\operatorname*{D-lim}_{u\in V}\Big(E_Y\big(\prod_{j=1}^{k}T_j(u)\alpha_j(u)\big)-\prod_{j=1}^{k}E_Y\big(T_j(u)\alpha_j(u)\big)\Big)=0\quad \mathit{in}\ L^1(Y). \tag{4.2}$$

Proof. Because of the multilinearity of (4.2) we may replace each C_j by $\big\{f-E_Y(f):f\in C_j\big\}$, $j=1,\ldots,k$, and assume that $E_Y(\alpha_j(u))=0$ for all $u\in V$ and $j=1,\ldots,k$. Indeed, for each $f\in L^\infty(X)$ let $f'=f-E_Y(f)$. Then, writing for convenience $\alpha'_j(u)$ for $(\alpha_j(u))'$, we have

$$\begin{aligned}
&E_Y\Big(\prod_{j=1}^{k}T_j(u)\alpha_j(u)\Big)-\prod_{j=1}^{k}E_Y\big(T_j(u)\alpha_j(u)\big)\\
&=E_Y\Big(\prod_{j=1}^{k}T_j(u)\big(\alpha'_j(u)+E_Y(\alpha_j(u))\big)\Big)-\prod_{j=1}^{k}T_j(u)E_Y(\alpha_j(u))\\
&=\sum_{\substack{I\subseteq\{1,\ldots,k\}\\ I\neq\emptyset}}E_Y\Big(\prod_{j\in I}T_j(u)\alpha'_j(u)\cdot\prod_{j\notin I}T_j(u)E_Y(\alpha_j(u))\Big)\\
&=\sum_{\substack{I\subseteq\{1,\ldots,k\}\\ I\neq\emptyset}}\prod_{j\notin I}T_j(u)E_Y(\alpha_j(u))\cdot E_Y\Big(\prod_{j\in I}T_j(u)\alpha'_j(u)\Big),
\end{aligned}$$

and we are done if we show that $\operatorname*{D-lim}_{u\in V}E_Y\big(\prod_{j\in I}T_j(u)\alpha'_j(u)\big)=0$ for any $I\subseteq\{1,\ldots,k\}$, $I\neq\emptyset$.

After a renumeraion of indices we may assume that $I=\{1,\ldots,l\}$. Without loss of generality $\sup|f|\leq 1$ for all $f\in C_1\cup\ldots\cup C_l$. Fix an $\varepsilon>0$ and choose an ε-net $\{f_{1,1},\ldots,f_{1,m_1}\}$ for C_1 over Y, an $\frac{\varepsilon}{m_1}$-net $\{f_{2,1},\ldots,f_{2,m_2}\}$ for C_2 over Y, ..., and an

$\frac{\varepsilon}{m_1\ldots m_l}$-net $\{f_{l,1},\ldots,f_{l,m_l}\}$ for C_l over Y. Then for any $u\in V$, $\{T_j(u)f_{j,1},\ldots,T_j(u)f_{j,m_j}\}$ is an $\frac{\varepsilon}{m_1\ldots m_j}$-net for $T_j(u)C_j$ over Y, and by Lemma 3.23 we have

$$
\begin{aligned}
&\Bigl\|E_Y\bigl(\prod_{j=1}^{l}T_j(u)\alpha_j(u)\bigr)\Bigr\|_{L^1(Y)}\\
&\le\sum_{i_1=1}^{m_1}\Bigl\|E_Y\bigl(T_1(u)f_{1,i_1}\cdot\prod_{j=2}^{l}T_j(u)\alpha_j(u)\bigr)\Bigr\|_{L^1(Y)}+\varepsilon\Bigl\|\prod_{j=2}^{l}T_j(u)\alpha_j(u)\Bigr\|_Y\\
&<\sum_{i_1=1}^{m_1}\Bigl\|E_Y\bigl(T_1(u)f_{1,i_1}\cdot\prod_{j=2}^{l}T_j(u)\alpha_j(u)\bigr)\Bigr\|_{L^1(Y)}+\varepsilon\\
&\le\sum_{i_1=1}^{m_1}\sum_{i_2=1}^{m_2}\Bigl\|E_Y\bigl(T_1(u)f_{1,i_1}\cdot T_2(u)f_{2,i_2}\cdot\prod_{j=3}^{l}T_j(u)\alpha_j(u)\bigr)\Bigr\|_{L^1(Y)}+m_1\tfrac{\varepsilon}{m_1}+\varepsilon\\
&\le\ldots\le\sum_{i_1=1}^{m_1}\cdots\sum_{i_l=1}^{m_l}\Bigl\|E_Y\bigl(T_1(u)f_{1,i_1}\cdot\ldots\cdot T_l(u)f_{l,i_l}\bigr)\Bigr\|_{L^1(Y)}+l\varepsilon.
\end{aligned}
$$

Since ε is arbitrary and, by the assumption of the lemma,

$$
\begin{aligned}
&\operatorname*{D-lim}_{u\in V}E_Y\Bigl(\prod_{j=1}^{l}T_j(u)f_{j,i_j}\Bigr)=\operatorname*{D-lim}_{u\in V}E_Y\Bigl(\prod_{j=1}^{l}T_j(u)f_{j,i_j}\cdot\prod_{j=l+1}^{k}T_j(u)1\Bigr)\\
&=\operatorname*{D-lim}_{u\in V}\prod_{j=1}^{l}E_Y\bigl(T_j(u)f_{j,i_j}\bigr)\cdot\prod_{j=l+1}^{k}E_Y\bigl(T_j(u)1\bigr)=\operatorname*{D-lim}_{u\in V}\prod_{j=1}^{l}T_j(u)E_Y\bigl(f_{j,i_j}\bigr)=0
\end{aligned}
$$

for all $i_1\in\{1,\ldots,m_1\}$, …, $i_l\in\{1,\ldots,m_l\}$, we are done. ■

4.11. We will also use so-called *PET-induction*. Let us say that polynomials $\varphi_1,\varphi_2\colon V\longrightarrow W$ are *equivalent* if $\operatorname{Deg}\varphi_1=\operatorname{Deg}\varphi_2$ and $\operatorname{Deg}(\varphi_1-\varphi_2)<\operatorname{Deg}\varphi_1$. Given a finite set $\mathcal{P}$ of polynomials $V\longrightarrow W$ for $m\in\mathbb{Z}_+$ we define

$$\omega_m(\mathcal{P})=\ \#\text{ of equivalence classes of polynomials of degree } m \text{ in } \mathcal{P}$$

and set $\omega(\mathcal{P})=\bigl(\omega_0(\mathcal{P}),\omega_1(\mathcal{P}),\ldots\bigr)\in\mathcal{K}$, where $\mathcal{K}$ is the set of sequences of the form $(k_0,k_1,\ldots,k_r,0,0,\ldots)$ with $r\in\mathbb{N}$, $k_0,\ldots,k_r\in\mathbb{Z}_+$. $\mathcal{K}$ is well ordered by writing $(k_0,k_1,k_2,\ldots)<(l_0,l_1,l_2,\ldots)$ if there exists $s\in\mathbb{Z}_+$ such that $k_m=l_m$ for all $m>s$ and $k_s<l_s$. We say that $\mathcal{P}'$ *precedes* $\mathcal{P}$ if $\omega(\mathcal{P}')<\omega(\mathcal{P})$. *PET-induction* is induction on $\omega(\mathcal{P})$: if $S(\mathcal{P})$ is a statement to be proved, we check $S(\{0\})$ and prove $S(\mathcal{P})$ under the assumption that $S(\mathcal{P}')$ is true for all $\mathcal{P}'$ preceding $\mathcal{P}$.

4.12. Lemma. *Let $\mathcal{P}$ be a finite set of polynomials $V\longrightarrow W$.*
(i) *Let $\varphi\in\mathcal{P}$ and $v\in V$; define $\varphi'(u)=\varphi(u+v)$ and $\mathcal{P}'=\mathcal{P}\cup\{\varphi'\}$. Then $\omega(\mathcal{P}')=\omega(\mathcal{P})$.*
(ii) *Assume that $\mathcal{P}$ does not contain constant polynomials; let ψ be a polynomial of minimal degree in $\mathcal{P}$ and let $\mathcal{P}''=\bigl\{\varphi-\psi:\varphi\in\mathcal{P}\bigr\}$. Then $\omega(\mathcal{P}'')<\omega(\mathcal{P})$.*

Proof. (i) Clearly, φ' is equivalent to φ, so $\mathcal{P}\cup\{\varphi'\}$ has the same equivalence classes as $\mathcal{P}$.

(ii) If $\varphi_1,\varphi_2\in\mathcal{P}$ are equivalent to each other but are not equivalent to ψ, then $\varphi_1-\psi$ and $\varphi_2-\psi$ are equivalent and $\mathrm{Deg}(\varphi_1-\psi)=\mathrm{Deg}\,\varphi_1$. If φ is equivalent to ψ, then $\mathrm{Deg}(\varphi-\psi)<\mathrm{Deg}\,\varphi$. Hence, subtracting ψ transforms equivalence classes in $\mathcal{P}$ into equivalence classes of the same degree; the only exception is the class of ψ, which falls into classes of smaller degree. ■

4.13. Proof of Theorem 4.9. For each $i\in\{1,\dots,l\}$ fix $\varphi_i\in\mathcal{P}_i$. Since $\prod_{\varphi\in\mathcal{P}_i}U(\varphi(u))f_\varphi$ $=U(\varphi_i)\bigl(\prod_{\varphi\in\mathcal{P}_i}U\bigl((\varphi-\varphi_i)(u)\bigr)f_\varphi\bigr)$ and the sets $\bigl\{\prod_{\varphi\in\mathcal{P}_i}U\bigl((\varphi-\varphi_i)(u)\bigr)f_i : u\in V\bigr\}$ are Y-precompact, an application of Lemma 4.10 reduces the problem to the case $|\mathcal{P}_1|=\ldots=|\mathcal{P}_l|=1$. We therefore have to prove that

$$\operatorname*{D-lim}_{u\in V}\Bigl(E_Y\bigl(\prod_{i=1}^{l}U(\varphi_i(u))f_i\bigr)-\prod_{i=1}^{l}E_Y\bigl(U(\varphi_i(u))f_i\bigr)\Bigr)=0\quad\text{in } L^1(Y)\tag{4.3}$$

for any $f_1,\dots,f_l\in L^\infty(X)$. After replacing f_i by $U(\varphi_i(0))f_i$, we may assume that $\varphi_i(0)=0$, $i=1,\dots,l$. Because of the multilinearity of (4.3), we may also replace f_i by $f_i-E_Y(f_i)$, $i=1,\dots,l$, and prove

$$\operatorname*{D-lim}_{u\in V}E_Y\bigl(\prod_{i=1}^{l}U(\varphi_i(u))f_i\bigr)=0\quad\text{in } L^1(Y)$$

under the assumption that $E_Y(f_i)=0$, $i=1,\dots,l$.

We will use PET-induction on $\mathcal{P}=\{\varphi_1,\dots,\varphi_l\}$. The statement of the theorem is trivial for $\mathcal{P}=\{0\}$, which gives the base of induction. Let φ_1 be of minimal degree in $\mathcal{P}$. If $\varphi_1\neq 0$, then for any $u\in V$

$$\begin{aligned}\Bigl\|E_Y\bigl(\prod_{i=1}^{l}U(\varphi_i(u))f_i\bigr)\Bigr\|_{L^1(Y)}&=\Bigl\|U(\varphi_1(u))\Bigl(E_Y\bigl(\prod_{i=1}^{l}U(\varphi_i(u)-\varphi_1(u))f_i\bigr)\Bigr)\Bigr\|_{L^1(Y)}\\&=\Bigl\|E_Y\bigl(\prod_{i=1}^{l}U\bigl((\varphi_i-\varphi_1)(u)\bigr)f_i\bigr)\Bigr\|_{L^1(Y)},\end{aligned}$$

so we may replace $\mathcal{P}$ by the family $\{0,\varphi_2-\varphi_1,\dots,\varphi_l-\varphi_1\}$, which precedes $\mathcal{P}$ by Lemma 4.12(ii). We may therefore assume that $\varphi_1=0$. Then $U(\varphi_i)=U(\varphi_i-\varphi_1)$ are weakly mixing for all $i=2,\dots,l$, and we have to prove that

$$\operatorname*{D-lim}_{u\in V}E_Y\bigl(f_1\cdot\prod_{i=2}^{l}U(\varphi_i(u))f_i\bigr)=0\quad\text{in } L^1(Y).\tag{4.4}$$

For $i=2,\dots,l$, let λ_i be the linear part of φ_i. If $U(\varphi_i-\lambda_i)$ is compact relative to Y, we may write

$$U(\varphi_i(u))f_i=U(\lambda_i(u))\bigl(U((\varphi_i-\lambda_i)(u))f_i\bigr),$$

apply Lemma 4.10 once again and replace φ_i by λ_i. So, we may assume that for any $i=2,\ldots,l$, either φ_i is linear, or $U(\varphi_i-\lambda_i)$ is weakly mixing relative to Y. In the latter case, $\mathrm{Ran}(\varphi_i-\lambda_i)\not\subseteq L$, which implies by 3.8(iv) that $\mathrm{Span}\big(\mathrm{Ran}(D_{v^*}\varphi_i)\big)\not\subseteq L$, hence $D_{v^*}\varphi_i$ is weakly mixing relative to Y for almost all $v\in V$.

Let us consider two cases.

Case 1: Not all φ_i, $i=2,\ldots,l$, are linear.
For $v\in V$, define polynomials $\varphi_{i,v}$, $i=2,\ldots,l$, $v\in V$, by $\varphi_{i,v}(u)=\varphi_i(u+v^*)$. By Lemma 3.31, (4.4) would follow from

$$\operatorname*{D-lim}_{v^*\in V^*}\operatorname*{D-limsup}_{u\in V}\Big\|E_Y\big(\prod_{i=2}^{l}U(\varphi_i(u))f_i\cdot\prod_{i=2}^{l}U(\varphi_{i,v}(u))\bar{f}_i\big)\Big\|_{L^1(Y)}=0, \tag{4.5}$$

which we will now establish.

Fix $v\in V$ and define $\mathcal{P}_v=\{\varphi_i,\ \varphi_{i,v}\,:\,i=2,\ldots,l\}$. Let us investigate the weakly mixing/compact properties of differences of the elements of $\mathcal{P}_v$. If φ_i is linear, then $U(\varphi_{i,v}-\varphi_i)=U(\varphi_i(v^*))$ is constant. If φ_i is non-linear, then $U(\varphi_{i,v}-\varphi_i)=U(D_{v^*}\varphi_i)U(\varphi_i(v^*))$ is weakly mixing relative to Y for almost all $v\in V$. For $i\neq j$, $U(\varphi_i-\varphi_j)$ is weakly mixing relative to Y and so, $U(\varphi_{i,v}-\varphi_{j,v})$ is weakly mixing relative to Y for all $v\in V$. If $i\neq j$ and φ_i is linear, then $U(\varphi_{i,v}-\varphi_j)=U(\varphi_i-\varphi_j)U(\varphi_i(v^*))$ is weakly mixing relative to Y for all $v\in V$. Finally, let φ_i be non-linear and assume that there exists $w\in V$ such that $U(\varphi_{i,w}-\varphi_j)$ is compact relative to Y. Since for $u,v\in V$

$$(\varphi_{i,v}-\varphi_{i,w})(u)=\varphi_i(u+v^*)-\varphi_i(u+w^*)=D_{v^*-w^*}\varphi_i(u+w^*)+\varphi_i(v^*-w^*),$$

and since $U(D_{v^*-w^*}\varphi_i)$ is weakly mixing relative to Y for almost all $v\in V$, $U(\varphi_{i,v}-\varphi_{i,w})$ is weakly mixing relative to Y for almost all $v\in V$. By Lemma 3.30, $U(\varphi_{i,v}-\varphi_j)=U\big((\varphi_{i,v}-\varphi_{i,w})+(\varphi_{i,w}-\varphi_j)\big)$, is weakly mixing relative to Y for almost all $v\in V$. Hence, for almost all $v\in V$ the only case when $U(\varphi-\psi)$ with $\varphi,\psi\in\mathcal{P}_v$ is compact relative to Y is when $\varphi=\varphi_i$ is linear and $\psi=\varphi_{i,v}$.

Assume that $\varphi_2,\ldots,\varphi_k$ are linear and $\varphi_{k+1},\ldots,\varphi_l$ are non-linear; since we have at least one non-linear polynomial, $k<l$ and φ_l is non-linear. Since $0\neq\mathcal{P}_v$, it follows from Lemma 4.12(i) that $\mathcal{P}_v$ precedes $\mathcal{P}$. So, by the induction hypothesis, we have

$$\begin{aligned}\operatorname*{D-lim}_{u\in V}\Big(&E_Y\big(\prod_{i=2}^{l}U(\varphi_i(u))f_i\cdot\prod_{i=2}^{l}U(\varphi_{i,v}(u))\bar{f}_i\big)\\ &-\prod_{i=2}^{k}E_Y\big(U(\varphi_i(u))f_i\cdot U(\varphi_{i,v}(u))\bar{f}_i\big)\cdot\prod_{i=k+1}^{l}E_Y\big(U(\varphi_i(u))f_i\big)\cdot\prod_{i=k+1}^{l}E_Y\big(U(\varphi_{i,v}(u))\bar{f}_i\big)\Big)=0\\ &\text{in } L^1(Y)\end{aligned}$$

for almost all $v\in V$. Since $E_Y(f_{k+1})=\ldots=E_Y(f_l)=0$,

$$\begin{aligned}&\prod_{i=2}^{k}E_Y\big(U(\varphi_i(u))f_i\cdot U(\varphi_{i,v}(u))\bar{f}_i\big)\cdot\prod_{i=k+1}^{l}E_Y\big(U(\varphi_i(u))f_i\big)\cdot\prod_{i=k+1}^{l}E_Y\big(U(\varphi_{i,v}(u))\bar{f}_i\big)\\ &\quad=E_Y\big(U(\varphi_i(u))f_i\cdot U(\varphi_{i,v}(u))\bar{f}_i\big)\cdot\prod_{i=k+1}^{l}U(\varphi_i(u))E_Y(f_i)\cdot\prod_{i=k+1}^{l}U(\varphi_{i,v}(u))E_Y(\bar{f}_i)=0.\end{aligned}$$

Hence, $\operatorname*{D-lim}_{u\in V} E_Y\big(\prod_{i=2}^l U(\varphi_i(u))f_i \cdot \prod_{i=2}^l U(\varphi_{i,v}(u))\bar{f}_i\big) = 0$ in $L^1(Y)$ for almost all $v \in V$, so that (4.5) holds.

Case 2: $\varphi_2, \dots, \varphi_l$ are all linear.
Then for any $u, v \in V$

$$E_Y\big(\prod_{i=2}^l U(\varphi_i(u))f_i \cdot \prod_{i=2}^l U(\varphi_i(u+v))\bar{f}_i\big) = E_Y\big(\prod_{i=2}^l U(\varphi_i(u))(f_i \cdot U(\varphi_i(v))\bar{f}_i)\big),$$

and by induction on l we get for any $v \in V$

$$\operatorname*{D-lim}_{u\in V}\Big(E_Y\big(\prod_{i=2}^l U(\varphi_i(u))f_i \cdot \prod_{i=2}^l U(\varphi_i(u+v))\bar{f}_i\big) - \prod_{i=2}^l E_Y\big(f_i \cdot U(\varphi_i(v))\bar{f}_i\big)\Big) = 0 \quad \text{in } L^1(Y).$$

Since φ_i, $i = 2, \dots, l$, are Y-weakly mixing, $\operatorname*{D-lim}_{v\in V}\big\|E_Y\big(f_i \cdot U(\varphi_i(v))\bar{f}_i\big)\big\|_{L^1(Y)} = 0$, which implies

$$\operatorname*{D-lim}_{v\in V}\operatorname*{D-limsup}_{u\in V}\Big\|E_Y\big(\prod_{i=2}^l U(\varphi_i(u))f_i \cdot \prod_{i=2}^l U(\varphi_i(u+v))\bar{f}_i\big)\Big\|_{L^1(Y)} = 0.$$

By Lemma 3.31, (4.4) follows. ■

4.14. We now pass to the main result of the paper:

Theorem. *Let V, W be finite dimensional vector spaces over F, let U be a measure preserving action of W on a probability measure space $(X, \mathcal{B}, \mu)$ and let $\mathcal{P}$ be a finite family of polynomials $V \longrightarrow W$ with zero constant term. Then for any $B \in \mathcal{B}$ with $\mu(B) > 0$ there exists $c > 0$ such that the set $\big\{u \in V : \mu\big(\bigcap_{\varphi\in\mathcal{P}} U(\varphi(u))B\big) > c\big\}$ is syndetic in V.*

4.15. The remainder of this section constitutes a proof of Theorem 4.14. The strategy of the proof is as follows. First, we define a property of measure preserving systems that states, more or less, that a form of Theorem 4.14 holds in the system for any family of polynomials of at most some fixed formal degree. This property obviously holds for the trivial system, and we shall show, after disposing of a few preparatory lemmas, that the property passes to primitive extensions. (This is the main part of the proof.) Moreover, as one may also easily show that the property holds in systems generated by factors possessing the property, this is sufficient to prove P-VSz in general, as one may pass from the trivial system to an arbitrary system via a transfinite chain of primitive and/or limiting extensions.

4.16. We fix a vector space W over F, $n \in \mathbb{N}$ and a finite set $A \subset W$.

Let U be a measure preserving action of W on a probability measure space $\mathbf{Y} = (Y, \mathcal{D}, \nu)$. We will say that $\mathbf{Y}$ has *the VSZ property (with respect to U, n and A)* if for any finite dimensional vector space V over F, any finite family $\mathcal{P}$ of polynomials $V \longrightarrow W$ with coefficients in $\mathrm{Span}(A)$, of formal degree $\leq n$ and with zero constant term, and any $B \in \mathcal{B}$

with $\mu(B) > 0$, there exists $c > 0$ such that the set $\{u \in V : \mu(\bigcap_{\varphi\in\mathcal{P}} U(\varphi(u))B > c\}$ is syndetic.

We must show that any measure space has the VSZ property with respect to any U, n and A.

4.17. Given a probability measure space $\mathbf{X} = (X, \mathcal{B}, \mu)$, the set of inner factors of $\mathbf{X}$ is partially ordered by the rule $(X, \mathcal{B}_1, \mu) \le (X, \mathcal{B}_2, \mu)$ if $\mathcal{B}_1 \subseteq \mathcal{B}_2$.

Theorem. *Given a unitary action U of W on $\mathbf{X} = (X, \mathcal{B}, \mu)$, the set of U-invariant inner factors of $\mathbf{X}$ which have the VSZ property (with respect to U, n and A) has a maximal element.*

Proof. We mimic the proof of Proposition 3.3 in [FK1]. In light of Zorn's lemma, it suffices to prove that if $\{\mathcal{B}_\beta\}$ is a linearly ordered family of U-invariant sub-σ-algebras of $\mathcal{B}$ such that $(X, \mathcal{B}_\beta, \mu)$ has the VSZ property for all β, then $(X, \bigcup \mathcal{B}_\beta, \mu)$ has the VSZ property.

Let $B \in \bigcup \mathcal{B}_\beta$, $\mu(B) > 0$, and let a finite set $\mathcal{P}$ of polynomials $V \longrightarrow W$ (with coefficients in $\mathrm{Span}(A)$, of formal degree $\le n$ and with zero constant term) be given. Find $B' \in \mathcal{B}_\beta$ for some β with $\mu(B' \triangle B) < \frac{\mu(B)}{4|\mathcal{P}|}$. Let $\mu = \int \mu_x \, d\mu$ be the decomposition of μ with respect to the factor $(X, \mathcal{B}_\beta, \mu)$. Define $D = \{x \in X : \mu_x(B) \ge 1 - \frac{1}{2|\mathcal{P}|}\}$; then $D \in \mathcal{B}_\beta$ and $\mu(D) > 0$. (Since $\mu_x(B') = 1$ for $x \in B'$, one would have

$$\mu(B' \setminus B) \ge \frac{1}{2|\mathcal{P}|}\mu(B') > \frac{1}{2|\mathcal{P}|} \cdot \frac{\mu(B)}{2} = \frac{\mu(B)}{4|\mathcal{P}|}$$

otherwise.)

Since $(X, \mathcal{B}_\beta, \mu)$ has the VSZ property, there exists $c > 0$ such that the set $S = \{u \in V : \mu(\bigcap_{\varphi\in\mathcal{P}} U(-\varphi(u))D) > c\}$ is syndetic. Then for any $u \in S$ and any $x \in \bigcap_{\varphi\in\mathcal{P}} U(-\varphi(u))D$ we have

$$\mu_x\Big(\bigcap_{\varphi\in\mathcal{P}} U(\varphi(u))B\Big) \ge 1 - \sum_{\varphi\in\mathcal{P}} \big(1 - \mu_x(U(\varphi(u))B)\big) = 1 - \sum_{\varphi\in\mathcal{P}} \big(1 - \mu_{U(\varphi(u))x}(B)\big)$$
$$\ge 1 - \frac{|\mathcal{P}|}{2|\mathcal{P}|} = \frac{1}{2}.$$

So, $\mu(\bigcap_{\varphi\in\mathcal{P}} U(\varphi(u))B) > c/2$ for all $u \in S$. ■

4.18. It follows that in order to prove Theorem 4.14 it suffices to establish the following fact: if $\mathbf{Y}$ is a proper U-invariant inner factor of $\mathbf{X}$ which has the VSZ property, then there exists a U-invariant inner factor $\mathbf{X}'$ of $\mathbf{X}$ satisfying $\mathbf{X}' > \mathbf{Y}$ and having the VSZ property. Indeed, if this were the case, then any maximal element of the family of U-invariant factors of $\mathbf{X}$ having the VSZ property would of necessity coincide with $\mathbf{X}$. Therefore, in light of Theorem 4.8, it is enough to prove the following proposition:

4.19. Proposition. *Let an extension $(X, \mathcal{B}, \mu) \longrightarrow (Y, \mathcal{D}, \nu)$ be primitive with respect to U, n and A. If $(Y, \mathcal{D}, \nu)$ has the VSZ property with respect to U, n and A, then so does $(X, \mathcal{B}, \mu)$.*

4.20. We will need the following combinatorial fact.

Proposition. *Let $\varphi_1,\dots,\varphi_k$ be polynomials $V\longrightarrow W$ of formal degree $\le n$, with coefficients in $A\subseteq W$ and with zero constant term. There exist $N\in\mathbb{N}$ and a finite family $\mathcal{Q}$ of polynomials $V^N\longrightarrow W$ of formal degree $\le n$, with coefficients in* $\mathrm{Span}(A)$ *and with zero constant term such that*

(i) *for any $\Phi\in\mathcal{Q}$,* $\mathrm{Ran}(\Phi)\subseteq\mathrm{Span}\big(\bigcup_{i=1}^k\mathrm{Ran}(\varphi_i)\big)$*;*

(ii) *for any partition $\bigcup_{s=1}^r\mathcal{Q}_s=\mathcal{Q}$ there exist $s\in\{1,\dots,r\}$, $\Phi\in\mathcal{Q}_s$ and a nonempty set $\alpha\subseteq\{1,\dots,N\}$ such that $\Phi(v_1,\dots,v_N)+\varphi_i\big(\sum_{t\in\alpha}v_t\big)\subseteq\mathcal{Q}_s$ for all $i\in\{1,\dots,k\}$.*

Proof. For a polynomial $\varphi\colon V\longrightarrow W$ with $\varphi(0)=0$, let us define $\varphi^{(1)}=\varphi$ and, for $m\ge 2$, $\varphi^{(m)}\colon V^m\longrightarrow W$ by

$$\varphi^{(m)}(v_1,\dots,v_m)=\varphi(v_1+\ldots+v_m)-\sum_{l=1}^{m-1}\sum_{1\le t_1<\ldots<t_l\le m}\varphi^{(l)}(v_{t_1},\dots,v_{t_l}).$$

Then for any $m\in\mathbb{N}$, $\mathrm{Span}\big(\mathrm{Ran}(\varphi^{(m)})\big)\subseteq\mathrm{Span}\big(\mathrm{Ran}(\varphi)\big)$ and

$$\varphi(v_1+\ldots+v_m)=\sum_{l=1}^{m}\sum_{1\le t_1<\ldots<t_l\le m}\varphi^{(l)}(v_{t_1},\dots,v_{t_l}).$$

As one easily checks, $\varphi^{(m)}(v_1,\dots,v_m)=D_{v_1}\ldots D_{v_{m-1}}\varphi(v_m)$ and so, $\varphi^{(m)}=0$ for $m>\mathrm{Deg}\,\varphi$. Therefore, if $\mathrm{Deg}\,\varphi\le n$, we may also write

$$\varphi(v_1+\ldots+v_m)=\sum_{l=1}^{n}\sum_{1\le t_1<\ldots<t_l\le m}\varphi^{(l)}(v_{t_1},\dots,v_{t_l}).$$

Fix N; we will say momentarily how large it must be. For $i\in\{1,\dots,k\}$ and $\mathbf{t}=(t_1,\dots,t_n)\in\{1,\dots,N\}^n$, if $t_1<t_2<\ldots<t_m=t_{m+1}=\ldots=t_n$ for some $m\le n$ we define $\pi(\{(i,\mathbf{t})\})=\varphi_i^{(m)}(v_{t_1},\dots,v_{t_m})$, and put $\pi(\{(i,\mathbf{t})\})=0$ otherwise. π maps the family of singleton subsets of $\{1,\dots,k\}\times\{1,\dots,N\}^n$ into the set of polynomials $V^N\longrightarrow W$ (in variables $v_1,\dots,v_N$). π may be uniquely extended to a mapping from the power set $\mathcal{F}$ of $\{1,\dots,k\}\times\{1,\dots,N\}^n$ into the set of polynomials $V^N\longrightarrow W$ in such a way that $\pi(\gamma_1\cup\gamma_2)=\pi(\gamma_1)+\pi(\gamma_2)$ for $\gamma_1\cap\gamma_2=\emptyset$. By construction, for any $l\in\{1,\dots,k\}$ and nonempty $\alpha\subseteq\{1,\dots,N\}$ we have

$$\pi\big(\{i\}\times\alpha^n\big)=\sum_{t_1,\dots,t_n\in\alpha}\pi\big(\{(i,(t_1,\dots,t_n))\}\big)=\sum_{l=1}^{n}\sum_{\substack{t_1,\dots,t_l\in\alpha\\ t_1<\ldots<t_l}}\varphi_i^{(l)}(v_{t_1},\dots,v_{t_l})=\varphi_i\big(\sum_{t\in\alpha}v_t\big).$$

Put $\mathcal{Q}=\mathrm{Ran}(\pi)$. Suppose now $\mathcal{Q}=\bigcup_{s=1}^r\mathcal{Q}_s$. Put $Q_s=\pi^{-1}(\mathcal{Q}_s)$. Then $\bigcup_{s=1}^rQ_s=\mathcal{F}$. By the polynomial Hales-Jewett theorem (see [BL2]), if N is large enough (depending on k, n and r), then there exist $s\in\{1,\dots,r\}$, $\gamma\in Q_s$ and a nonempty $\alpha\subseteq\{1,\dots,N\}$ such that $\gamma\cap\big(\{1,\dots,k\}\times\alpha^n\big)=\emptyset$ and $\gamma\cup\big(\{i\}\times\alpha^n\big)\in Q_s$, $i=1,\dots,k$. Putting $\Phi=\pi(\gamma)$ we are done. ∎

4.21. In our proof of Proposition 4.19 we may and will assume without loss of generality that X is *a regular space* (see [F2], page 103). One then has *a decomposition of measures*: for almost every $y \in Y$ a measure μ_y on X is defined (μ_y "is concentrated" on the fiber $\pi^{-1}(y)$) so that $\mu = \int \mu_y \, d\nu$. For $f \in L^1(X)$ and almost all $y \in Y$ we then have $E_Y(f)(y) = \int f \, d\mu_y$.

4.22. To start the proof of Proposition 4.19 we need one more ingredient. It can be proven (cf. [Z2]) that a Y-unitary action of a group G on a Y-Hilbert space $\mathcal{H}$ is compact relative to Y if and only if $\mathcal{H}$ is decomposable into (a possibly infinite) sum of G-invariant Y-subspaces which have finite rank as L^∞-modules. Instead, we will follow the softer approach suggested in [FK1], which is based on the notion of an almost periodic function.

4.23. Let U be an action of a countable abelian group G on an extension $\pi \colon (X, \mathcal{B}, \mu) \longrightarrow (Y, \mathcal{D}, \nu)$; though it is not necessary, we will assume that X is a regular space. We say that a function $f \in L^2(X)$ *is almost periodic relative to* Y (with respect to U) if for any $\varepsilon > 0$ there exists a finite set $R \in L^2(X)$, called a *fiberwise a.e. ε-net* for $U(G)f$, such that for any $u \in G$ and almost every $y \in Y$ there exists $g \in R$ with $\left\| U(u)f - g \right\|_y < \varepsilon$.

Proposition. *If U is compact on X relative to Y then $L^2(X)$ has a dense set of functions almost periodic relative to Y with respect to U.*

Proof. For $H \in L^\infty(X) \otimes L^\infty(X)$ and $g \in L^2(X)$ define $H * g(x) = \int H(x, x') g(x') \, d\mu_{\pi(x)}(x')$. Then $g \mapsto H * g$ is a bounded operator on $L^2(X, \mu)$ and a compact operator on $L^2(X, \mu_y)$ and for almost all $y \in Y$.

Let $\mathcal{H} = L^2(X)$, let $\mathcal{K}$ be the subspace of $\mathcal{H} \otimes_Y \mathcal{H}$ consisting of the U-invariant functions and let $\mathcal{K}^\infty \subseteq \mathcal{K}$ be the subspace of bounded functions from $\mathcal{K}$. By using the identity $\langle f, H * g \rangle = \langle f \otimes g, H \rangle$, one sees that if $\langle f, H * g \rangle = 0$ for all $H \in \mathcal{K}^\infty$ and all $g \in \mathcal{H}$ then one has $\langle f \otimes \bar{f}, H \rangle = 0$ for all $H \in \mathcal{K}^\infty$. Since $\mathcal{K}^\infty$ is dense in $\mathcal{K}$, this implies by Theorem 3.27 that U is weakly mixing on f relative to Y and so, $f = 0$. It follows that the set $\left\{ H * g : H \in \mathcal{K}^\infty, \ g \in \mathcal{H} \right\}$ is dense in $\mathcal{H}$, and we need only show that any $H * g$ may be approximated by an almost periodic function.

Let now $G_1 \subseteq G_2 \subseteq \ldots$ be an exhaustive sequence of finite subsets of G. Fix $H \in \mathcal{K}^\infty$ and $g \in \mathcal{H}$, and put $f = H * g$. For any $v \in U$ one has $U(v)f = (U(v)H) * (U(v)g) = H * (U(v)g)$, and so the set $\left\{ U(v)f : v \in G \right\}$ is precompact in the $\| \cdot \|_y$-norm for a.e. $y \in Y$. For $\delta > 0$ and $y \in Y$ let $M = M(y, \delta)$ be the minimal positive integer such that the set $\left\{ U(u)f : u \in G_M \right\}$ is δ-dense in $\left\{ U(v)f : v \in G \right\}$ in the $\| \cdot \|_y$-norm. Next, for $k \in \mathbb{N}$ let M_k be large enough that $\nu\left(\{ y : M(y, \frac{1}{k}) > M_k \} \right) < \frac{1}{2^k}$. For $K \in \mathbb{N}$ let $E_K = Y \setminus \left(\bigcup_{k=K+1}^\infty \{ y : M\left(y, \frac{1}{k}\right) > M_k \} \right)$. Note that $\nu(E_K) \geq 1 - 2^{-K}$. Finally, let $\widetilde{E}_K = \bigcup_{v \in G} U(v)^{-1} E_K$ and let $f_K = f \cdot 1_{\widetilde{E}_K}$. Note that for K large, f_K approximates f closely. We claim that $h = f_K$ is AP.

To see this, let $\varepsilon > 0$ and let k be large enough that $k > K$ and $\frac{1}{k} < \varepsilon$. For $u \in G_{M_k}$ let $h_u(x)$ be equal to $f\big(U(u)x\big)$ if $\pi(x) \in E_K$, zero if $\pi(x) \in Y \setminus \widetilde{E}_K$, and for $y \in \widetilde{E}_K \setminus E_K$ choose (measurably) $v_y \in G$ with $U(v_y)y \in E_K$ and set $h_u(x)$ equal to $f\big(U(v_{\pi(x)})U(u)x\big)$ if $\pi(x) \in \widetilde{E}_K \setminus E_K$. We claim that $\{0\} \cup \left\{ h_u : u \in G_{M_k} \right\}$ is a fiberwise a.e. ε-net for h.

Indeed, let $v \in G$ and $y \in Y$. First, let $y \in E_K$. Then on the fiber over y, $h = f$, $h_u = U(u)f$ for $u \in G_{M_k}$, and the set $\{h_u : u \in G_{M_k}\}$ is ε-dense in $\{U(v) : v \in G\}$ with respect to the $\|\cdot\|_y$-norm. Thus, for some $u \in G_{M_k}$ we have $\|U(v)h - h_u\|_y < \varepsilon$. The complement of $\widetilde{E}_K$ is an invariant set, therefore if $y \in Y \setminus \widetilde{E}_K$ then $U(u)h$ is zero on the fiber over y. Finally if $y \in \widetilde{E}_K \setminus E_K$ then since $U(v_y)y \in E_K$, for some $u \in G_{M_k}$ we have $\|U(v)U(v_y)^{-1}f - U(u)f\|_{U(v_y)y} < \varepsilon$. But this implies that $\|U(v)h - h_u\|_y < \varepsilon$. ■

4.24. Proposition. *If U is compact on X relative to Y, then for any $f \in \mathcal{H}$ and any $\delta > 0$ there exists $D \subseteq Y$ with $\nu(D) > 1 - \delta$ such that $1_D \cdot f$ is almost periodic relative to Y with respect to U.*

Proof. Let $f \in \mathcal{H}$ and let $\delta > 0$ be given. Choose a sequence $(\varepsilon_n)_{n=1}^{\infty}$ of positive numbers such that $\sum_{n=1}^{\infty} \sqrt{\varepsilon_n} < \delta$. For every $n = 1, 2, \ldots$ choose an almost periodic f_n with $\|f_n - f\| < \varepsilon_n$ and let R_n be a fiberwise a.e. ε_n-net for $U(G)f_n$. Let $C_n = \{y : \|f_n - f\|_y \geq \sqrt{\varepsilon_n}\}$. Then $\nu(C_n) < \sqrt{\varepsilon_n}$, so that $D = Y \setminus \bigcup_{n=1}^{\infty} C_n$ satisfies $\nu(D) > 1 - \delta$.

Now given $\varepsilon > 0$, pick n with $2\varepsilon_n < \varepsilon$. Let $u \in G$ and $y \in Y$. If $U(u)y \in D$ then $\|U(u)f - U(u)f_n\|_y = \|f - f_n\|_{U(u)y} < \varepsilon_n$, which implies that there exists $g \in R_n$ such that $\|U(u)f - g\|_y < \varepsilon$. If on the other hand $U(u)y \notin D$, then $\|U(u)f - 0\|_y = \|f\|_{U(u)y} = 0$. Hence $R_n \cup \{0\}$ is a fiberwise a.e. ε-net for $U(G)f$. ■

4.25. Proof of Proposition 4.19. Let $L \in \Lambda$ be the space ensuring the primitivity of the extension $(X, \mathcal{B}, \mu) \longrightarrow (Y, \mathcal{D}, \nu)$, as defined in 4.7. Let V be a finite dimensional vector space over F, let $\mathcal{P}$ be a finite family $\mathcal{P}$ of polynomials $V \longrightarrow W$ of formal degree $\leq n$, with coefficients in $\mathrm{Span}(A)$ and with zero constant term. Let $\mathcal{G}$ be the group of polynomials $V \longrightarrow W$ of formal degree $\leq n$ and with zero constant term, let $\mathcal{G}_0 = \{\varphi \in \mathcal{G} : \mathrm{Ran}(\varphi) \subseteq L\}$, and let $\mathcal{W}$ be a set of coset representatives for $\mathcal{G}/\mathcal{G}_0$. Choose $\varphi_1, \ldots, \varphi_k \in \mathcal{W}$ and $\psi_1 = 0, \psi_2, \ldots, \psi_l \in \mathcal{G}_0$ such that $\mathcal{P} \subseteq \{\varphi_i + \psi_j : 1 \leq i \leq k,\ 1 \leq j \leq l\}$. Note that $U(\psi_1), \ldots, U(\psi_l)$ are compact on X relative to Y, while $U(\varphi_i - \varphi_j)$, $i \neq j$, are weakly mixing on X relative to Y.

Let $B \in \mathcal{B}$, $\mu(B) > 0$. By Proposition 4.24 we may assume that $f = 1_B$ is almost periodic relative to Y with respect to $U|_L$. Pick $b > 0$ and $D \in \mathcal{D}$ with $\nu(D) > 0$ such that $\mu_y(B) > b$ for all $y \in D$. Put $\varepsilon = \frac{b^2}{8l^2}$. Let $\{g_1, \ldots, g_r\}$ be a fiberwise a.e. ε-net for $U(L)f$.

We now apply Proposition 4.20 to the family $\{(\varphi_i, \psi_j) : 1 \leq i \leq k,\ 1 \leq j \leq l\}$ of polynomials $V \longrightarrow W \times W$ and find $N \in \mathbb{N}$ and a finite family $\mathcal{Q}$ of polynomials $(\Phi, \Psi)\colon V^N \longrightarrow W \times W$ of formal degree $\leq n$, with coefficients in $\mathrm{Span}(A) \times \mathrm{Span}(A)$ and with zero constant term such that for any $(\Phi, \Psi) \in \mathcal{Q}$, $\mathrm{Ran}(\Psi) \subseteq \mathrm{Span}\big(\bigcup_{j=1}^{k} \mathrm{Ran}(\psi_j)\big) \subseteq L$ and for any partition $\bigcup_{s=1}^{r} \mathcal{Q}_s = \mathcal{Q}$ there are $s \in \{1, \ldots, r\}$, $(\Phi, \Psi) \in \mathcal{Q}$ and a nonempty $\alpha \subseteq \{1, \ldots, N\}$ such that

$$\Big(\Phi(v_1, \ldots, v_N) + \varphi_i\big(\sum_{t \in \alpha} v_t\big), \Psi(v_1, \ldots, v_N) + \psi_j\big(\sum_{t \in \alpha} v_t\big)\Big) \in \mathcal{Q}_s \quad \text{for all } i \in \{1, \ldots, k\},\ j \in \{1, \ldots, l\}.$$

Since $(Y, \mathcal{D}, \nu)$ has the VSZ property, there exists $d > 0$ such that the set

$$S = \Big\{\mathbf{v} \in V^N : \nu\Big(\bigcap_{(\Phi,\Psi)\in\mathcal{Q}} U\big(\Phi(\mathbf{v}) + \Psi(\mathbf{v})\big)D\Big) > d\Big\}$$

is syndetic. We put $c = \frac{b^k d}{2^{k+N+1} r|\mathcal{Q}|}$.

Let $T \subseteq V$ be an arbitrary thick set. For $u \in T$ put

$$C_u = \Big\{y \in Y : \Big|\mu_y\Big(\bigcap_{i=1}^k \bigcap_{j=1}^l U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) - \prod_{i=1}^k \mu_y\Big(\bigcap_{j=1}^l U\big(\varphi_i(u) + \psi_j(u)\big)B\Big)\Big| > \frac{b^k}{2^{k-1}}\Big\}$$

$$= \Big\{y \in Y : \Big|\int \prod_{i=1}^k \prod_{j=1}^l U\big(-\varphi_i(u) - \psi_j(u)\big)f\,d\mu_y - \prod_{i=1}^k \int \prod_{j=1}^l U(-\varphi_i(u) - \psi_j(u))f\,d\mu_y\Big| > \frac{b^k}{2^{k-1}}\Big\}.$$

By Theorem 4.9 the set $C = \big\{u \in V : \nu(C_u) < \frac{d}{2^{N+1}|\mathcal{Q}|}\big\}$ has Banach density 1. By Lemma 2.7(vi), $T' = T \cap C$ is thick. Utilizing Lemma 2.8, fix $\mathbf{v} = (v_1, \ldots, v_N) \in S$ such that $\mathrm{FS}\big(\{v_1, \ldots, v_N\}\big) \subset T'$. Let

$$D_1 = \Big(\bigcap_{(\Phi,\Psi)\in\mathcal{Q}} U\big(\Phi(\mathbf{v}) + \Psi(\mathbf{v})\big)D\Big) \setminus \Big(\bigcup_{(\Phi,\Psi)\in\mathcal{Q}} \bigcup_{u\in\mathrm{FS}(\{v_1,\ldots,v_N\})} U\big(\Phi(\mathbf{v}) + \Psi(\mathbf{v})\big)C_u\Big).$$

Then $\nu(D_1) > \frac{d}{2}$.

For $y \in D$ and $s \in \{1, \ldots, r\}$ let

$$\mathcal{Q}_{y,s} = \Big\{(\Phi, \Psi) \in \mathcal{Q} : \big\|U(-\Psi(\mathbf{v}))f - g_s\big\|_{U(-\Phi(\mathbf{v}))y} < \varepsilon\Big\};$$

then $\bigcup_{s=1}^r \mathcal{Q}_{y,s} = \mathcal{Q}$ for a.e. $y \in Y$. By the choice of $\mathcal{Q}$, for a.e. $y \in Y$ there exist $s \in \{1, \ldots, r\}$, $(\Phi, \Psi) \in \mathcal{Q}$ and a nonempty $\alpha \in \{1, \ldots, N\}$ such that for $u = \sum_{t\in\alpha} v_t \in \mathrm{FS}(\{v_1, \ldots, v_N\})$ one has $\big(\Phi(\mathbf{v}) + \varphi_i(u), \Psi(\mathbf{v}) + \psi_j(u)\big) \in \mathcal{Q}_{y,s}$, that is,

$$\big\|U\big(-\Psi(\mathbf{v}) - \psi_j(u)\big)f - g_s\big\|_{U(-\Phi(v)-\varphi_i(u))y} < \varepsilon \tag{4.6}$$

for all $i = 1, \ldots, k$, $j = 1, \ldots, l$. Since there are at most $2^N r|\mathcal{Q}|$ choices for α, s and (Φ, Ψ), we may find a set $D_2 \subseteq D_1$ with $\nu(D_2) > \frac{d}{2^{N+1} r|\mathcal{Q}|}$ such that (4.6) holds for some α, s and (Φ, Ψ) not depending on $y \in D_2$. We fix these α, s and (Φ, Ψ).

For $y \in D_2$, let $\tilde{y} = U\big(-\Phi(\mathbf{v}) - \Psi(\mathbf{v})\big)y$. By (4.6),

$$\big\|U\big(-\varphi_i(u) - \psi_j(u)\big)f - U\big(-\psi_j(u) + \Psi(\mathbf{v})\big)g_s\big\|_{\tilde{y}} < \varepsilon$$

for all $i = 1, \ldots, k$, $j = 1, \ldots, l$. Since $\psi_1 = 0$, this implies

$$\left\|U\big(-\varphi_i(u) - \psi_j(u)\big)f - U(-\varphi_i(u))f\right\|_{\tilde{y}} < 2\varepsilon,$$

which is equivalent to

$$\mu_{\tilde{y}}\Big(U\big(\varphi_i(u) + \psi_j(u)\big)B \triangle U(\varphi_i(u))B\Big) < \sqrt{2\varepsilon}.$$

It follows that for all $i = 1, \ldots, k$,

$$\mu_{\tilde{y}}\Big(\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) > \mu_{\tilde{y}}(B) - l\sqrt{2\varepsilon}.$$

By construction of $\tilde{y}$ and D_1, $\tilde{y} \in D$. Therefore $\mu_{\tilde{y}}(B) > b$ and so,

$$\mu_{\tilde{y}}\Big(\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) > b - l\sqrt{2\varepsilon} = \frac{b}{2}$$

for all $i = 1, \ldots, k$. Hence,

$$\prod_{i=1}^{k} \mu_{\tilde{y}}\Big(\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) > \frac{b^k}{2^k}.$$

On the other hand, $\tilde{y} \notin C_u$, hence

$$\left|\mu_{\tilde{y}}\Big(\bigcap_{i=1}^{k}\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) - \prod_{i=1}^{k} \mu_{\tilde{y}}\Big(\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big)\right| < \frac{b^k}{2^{k-1}}.$$

We therefore have

$$\mu_{\tilde{y}}\Big(\bigcap_{i=1}^{k}\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) > \frac{b^k}{2^{k-1}} - \frac{b^k}{2^k} = \frac{b^k}{2^k}.$$

Finally, as this holds for all $\tilde{y} \in U\big(-\Phi(\mathbf{v}) - \Psi(\mathbf{v})\big)D_2$ and $\nu(D_2) \geq \frac{d}{2^{N+1}r|\mathcal{Q}|}$, we have

$$\mu\Big(\bigcap_{i=1}^{k}\bigcap_{j=1}^{l} U\big(\varphi_i(u) + \psi_j(u)\big)B\Big) > \frac{b^k d}{2^{k+N+1}r|\mathcal{Q}|} = c.$$

Since $\mathcal{P} \subseteq \{\varphi_i + \psi_j : 1 \leq i \leq k,\ 1 \leq j \leq l\}$, this implies $\bigcap_{\varphi\in\mathcal{P}} U\big(\varphi(u)\big)B\big) > c$. Since u was chosen from an arbitrary thick set T, we have shown that the set $\big\{u \in V : \mu\big(\bigcap_{\varphi\in\mathcal{P}} U\big(\varphi(u)(u)\big)B\big) > c\big\}$ is syndetic. ■

5. Applications of main theorem

5.1. In this chapter we derive some combinatorial corollaries from our main theorem. We start by formulating a version of Furstenberg's correspondence principle, which serves as a bridge between multiple recurrence and density combinatorics:

Theorem. (Cf. [F2], page 152, and [B2], Theorem 4.17.) *Given an abelian group G and $E \subseteq G$, there exist a probability space $(X, \mathcal{B}, \mu)$, a set $A \subseteq \mathcal{B}$ with $\mu(A) = d^*(E)$ and a measure preserving action U of G on X such that for any $k \in \mathbb{N}$ and any $v_1, \ldots, v_k \in G$ one has $d^*\big(E \cap (E + v_1) \cap \ldots \cap (E + v_k)\big) \geq \mu\big(A \cap U(v_1)A \cap \ldots \cap U(v_k)A\big)$.*

5.2. The following combinatorial statement now follows immediately from Theorem 4.14 and the correspondence principle:

Theorem. *Let F be a countable field, let V and W be finite dimensional vector spaces over F and let $\mathcal{P}$ be a finite family of polynomials $V \longrightarrow W$ with zero constant term. For any $E \subseteq W$ with $d^*(E) > 0$ there exist $u \in V$, $u \neq 0$, and $w \in E$ such that $w + \varphi(u) \in E$ for all $\varphi \in \mathcal{P}$. Moreover, the set of u which satisfy $w + \varphi(u) \in E$ for some $w \in E$ and all $\varphi \in \mathcal{P}$ is syndetic in V.*

5.3. As a matter of fact, one may derive Theorem 4.14 from Theorem 5.2 by arguing as follows. Let $B \in \mathcal{B}$ with $\mu(B) > 0$. Then for almost all $x \in B$ the set $E_x = \big\{w \in W : U(w)x \in B\big\}$ satisfies $d^*(E_x) > 0$. By Theorem 5.2, there exist $w_x \in E_x$ and nonzero $u_x \in V$ such that $w_x + \varphi(u_x) \in E_x$ for all $\varphi \in \mathcal{P}$, so that $U(w_x)x \in B$ and $U(\varphi(u_x))\big(U(w_x)x\big) \in B$ for all $\varphi \in \mathcal{P}$. Choose $u \in V$ and $w \in W$ such that the set $B = \big\{x \in X : u_x = u \text{ and } w_x = w\big\}$ has positive measure. Then $U(w)B \subseteq B$ and $U(\varphi(u))\big(U(w)B\big) \subseteq B$ for all $\varphi \in \mathcal{P}$.

5.4. The following is a "finitary" version of Theorem 5.2.

Corollary. *Let $\mathcal{P}$ be a finite family of polynomials $V \longrightarrow W$ with zero constant term and let $\{\Phi_n\}$ be a Følner sequence in W. For any $\alpha > 0$ there exists $N \in \mathbb{N}$ such that whenever $n \geq N$ and $E \subseteq \Phi_n$ with $|E| \geq \alpha|\Phi_n|$, there exist $u \in V$, $u \neq 0$, and $w \in E$ such that $w + \varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

Proof. Assume, without loss of generality, that $0 \in \mathcal{P}$ and that $|\mathcal{P}| \geq 3$. Supposing the conclusion fails, we choose $E_n \subseteq \Phi_n$, $|E_n| > \alpha|\Phi_n|$, $n \in \mathbb{N}$, not containing any $\{w + \varphi(u) : \varphi \in \mathcal{P}\}$, $u \neq 0$. Having chosen $w_i \in W$, $i = 1, \ldots, n-1$, let

$$B_n = \Big\{u \in V : \text{there exist } \varphi_1, \varphi_2 \in \mathcal{P} \text{ with } \varphi_1(u) - \varphi_2(u) \in \Big(\bigcup_{i=1}^{n-1}(w_i + E_i) - \bigcup_{i=1}^{n-1}(w_i + E_i)\Big) \cup (E_n - E_n)\Big\}.$$

Note that B_n is finite. Choose w_n outside the (finite) set

$$\bigcup_{i=1}^{n-1}(w_i + E_i) - \big\{\varphi(u) : u \in B_n,\ \varphi \in \mathcal{P}\big\} + \big\{\varphi(u) : u \in B_n,\ \varphi \in \mathcal{P}\big\} - E_n.$$

Again $E = \bigcup_{i=1}^{\infty}(w_n + E_n)$ satisfies $d^*(E) \geq \alpha$, so E contains a configuration $\{w + \varphi(u) : \varphi \in \mathcal{P}\}$, $u \neq 0$. Let $\varphi_1, \varphi_2, \varphi_3$ be distinct members of $\mathcal{P}$ and suppose there are $w \in W$, $0 \neq$

$u \in V$ and $l \le m \le n$ with $w+\varphi_1(u) \in w_l+E_l$, $w+\varphi_2(u) \in w_m+E_m$, $w+\varphi_3(u) \in w_n+E_n$. One checks that necessarily $u \in B_n$, hence it cannot be the case that $m < n$ or $l < m = n$, as in either of these cases we would have $w \in \bigcup_{i=1}^{n-1}(w_i + E_i) - \{\varphi(u) : u \in B_n,\ \varphi \in \mathcal{P}\}$ and $w_n \in w + \{\varphi(u) : u \in B_n,\ \varphi \in \mathcal{P}\} - E_n$, a contradiction. Hence $l = m = n$ and, as the $\varphi_1, \varphi_2, \varphi_3$ were arbitrary members of $\mathcal{P}$, in fact $\{w + \varphi(u) : \varphi \in \mathcal{P}\} \subseteq w_n + E_n$, a contradiction. ■

5.5. We will now extend our multiple recurrence theorems from vector spaces to modules over integral domains. Let K be a ring and let M be a module over K; we call a mapping $\varphi\colon K^d \longrightarrow M$ *a polynomial* if it has the form $\varphi(x_1, \ldots, x_d) = \sum_{i=1}^k x_1^{n_{i,1}} \ldots x_d^{n_{i,d}} a_i$, $a_i \in M$.

Theorem. *Let K be a countable integral domain, let M be a K-module, let U be a measure preserving action of M on a probability measure space $(X, \mathcal{B}, \mu)$ and let $\mathcal{P}$ be a finite family of polynomials $K \longrightarrow M$ with zero constant term. For any $B \in \mathcal{B}$ with $\mu(B) > 0$ there exists $u \in K$, $u \neq 0$, such that $\mu\big(\bigcap_{\varphi\in\mathcal{P}} U(\varphi(u))B\big) > 0$.*

5.6. The following is an equivalent combinatorial version of Theorem 5.5. (One can establish the equivalence with the help of the correspondence principle and by an argument similar to that used in 5.3.)

Theorem. *Let K be a countable integral domain, let M be a K-module and let $\mathcal{P}$ be a finite family of polynomials $K \longrightarrow M$ with zero constant term. For any $E \subseteq M$ with $d^*(E) > 0$ there exist $u \in K$, $u \neq 0$, and $w \in E$ such that $w + \varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

5.7. We will be proving the finitary version of Theorem 5.6:

Theorem. *Let K be a countable integral domain, let M be a K-module, let $\mathcal{P}$ be a finite family of polynomials $K \longrightarrow M$ with zero constant term and let $\{\Phi_n\}$ be a Følner sequence in M. For any $\alpha > 0$ there exists $N \in \mathbb{N}$ such that whenever $n \ge N$ and $E \subseteq \Phi_n$ with $|E| \ge \alpha|\Phi_n|$, there exist $u \in K$, $u \neq 0$, and $w \in E$ such that $w + \varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

Proof. Denote by F the field of quotients of K. Let $\mathcal{P} = \{\varphi_1, \ldots, \varphi_k\}$, where $\varphi_i(x) = \sum_{j=1}^{d_i} x^j a_{i,j}$, $a_{i,j} \in M$. We put $W = F^{d_1+\ldots+d_k}$ and $R = K^{d_1+\ldots+d_k} \subseteq W$. Let ζ be the homomorphism $R \longrightarrow M$ defined by $\zeta(v) = \sum_{i=1}^k \sum_{j=1}^{d_i} v_{i,j} a_{i,j}$ for $v = (v_{1,1}, \ldots, v_{k,d_k})$. We define polynomial mappings $\pi_1, \ldots, \pi_k\colon F \longrightarrow W$ by $\big(\pi_i(u)\big)_{l,j} = \begin{cases} u^j \text{ if } i = l \\ 0 \text{ otherwise.} \end{cases}$ Then $\pi_i(K) \subseteq R$ and $\zeta \circ \pi_i = \varphi_i$, $i = 1, \ldots, k$.

Let G be a countable abelian group and let $v_1, \ldots, v_r \in G$. For each $l = 1, \ldots, r$, if $\langle v_l\rangle \cap \langle v_1, \ldots, v_{l-1}\rangle \neq \{0\}$ let N_l be the minimal positive integer for which $N_l v_l \in \langle v_1, \ldots, v_{l-1}\rangle$, and let N_l be an arbitrary positive integer otherwise. We will call a set $P \subseteq G$ of the form $P = \{n_1 v_1 + \ldots + n_r v_r : n_l \in \{0, 1, \ldots, N_l - 1\},\ l = 1, \ldots, r\}$ an *admissible parallelepiped.* One can check that there always exists a Følner sequence consisting of admissible parallelepipeds, and that the following holds:

5.8. Lemma. *Given a Følner sequence $\{\Psi_n\}$, a admissible parallelepiped P in G and $\beta > 0$, for any n large enough there exist $w_1, \ldots, w_m \in \Psi_n$ such that $\bigcup_{j=1}^m (w_j + P) \subseteq \Psi_n$, $(w_i + P) \cap (w_j + P) = \emptyset$ for $i \neq j$ and $\big|\bigcup_{j=1}^m (w_j + P)\big| > (1 - \beta)|\Psi_n|$.*

Sketch of the proof. Let $H=\langle v_1,\ldots,v_r\rangle$ and let $Z\subset G$ be a set of representatives of G/H. Let $Y=\{z+k_1N_1v_1+\ldots+k_rN_rv_r : z\in Z,\ k_1,\ldots,k_r\in\mathbb{Z}\}$. Then $\bigcup_{w\in Y}(w+P)$ is a partition of G. Let $Y_n=\{w\in Y : w+P\subseteq\Psi_n\}$; if n is large enough then $\{w\in\Psi_n : w\pm P\subseteq\Psi_n\}\approx\Psi_n$ and so, $\bigcup_{w\in Y_n}(w+P)\approx\Psi_n$. ∎

Let $\{Q_n\}$ be a Følner sequence of admissible parallelepipeds in W. Using Corollary 5.4, find n such that whenever $C\subseteq Q_n$ with $|C|\geq\frac{\alpha}{2}|Q_n|$, there exist $u\in W$, $u\neq 0$, and $w\in C$ such that $w+\pi_1(u),\ldots,w+\pi_k(u)\in C$. Let $h\in K$ be such that $hw\in R$ for all $w\in Q_n$, and such that $hu\in K$ for all $u\in F$ having the property that $\pi_i(u)=Q_n-Q_n$ for all $i=1,\ldots,k$. Define $\eta\colon W\longrightarrow R$ by $\big(\eta(v)\big)_{i,j}=h^jv_{i,j}$, $i=1,\ldots,k$, $j=1,\ldots,d_i$, $v=(v_{1,1},\ldots,v_{k,d_k})$; then $\eta\big(\pi_i(v)\big)=\pi_i(hv)$, $v\in W$, $i=1,\ldots,k$.

Define $Q=\eta(Q_n)$; since η is injective, Q is a admissible parallelepiped in R. Assume that $D\subseteq Q$, $|D|\geq\frac{\alpha}{2}|Q|$. Then $C=\eta^{-1}(D)\subseteq Q_n$ and $|C|\geq\frac{\alpha}{2}|Q|$. Thus there exist $u\in F$, $u\neq 0$, and $w\in C$ such that $w+\pi_i(u)\in C$ for all $i=1,\ldots,k$. Since $w\in Q_n$, one has $hw\in R$. Since $\pi_i(u)\in Q_n-Q_n$, $i=1,\ldots,k$, one has $hu\in K$. Finally we have $\eta(w)\in D$ and $\eta\big(w+\pi_i(u)\big)=\eta(w)+\pi_i(hu)\in D$ for all $i=1,\ldots,k$.

Let $L=\operatorname{Ran}\zeta$; choose a Følner sequence $\{P_n\}$ of admissible parallelepipeds in L. Choose a Følner sequence $\{\Psi'_m\}$ in $\ker(\zeta)$ and a (not necessarily homomorphic) section $\tau\colon L\longrightarrow R$ of ζ.

5.9. Lemma. *There exists a sequence $\{m_n\}$ of positive integers such that $\Psi_n=\tau(\mathcal{P}_n)+\Psi'_{m_n}$, $n\in\mathbb{N}$, is a Følner sequence in R.*

Now let $\{\Phi_n\}$ be a Følner sequence in M. Define $\{\Psi_n\}$ as in Lemma 5.9 and let, in accordance with Lemma 5.8, i and $v_1,\ldots,v_m\in\Psi_i$ be such that $v_1+Q,\ldots,v_m+Q\subseteq\Psi_i$, the sets v_j+Q are pairwise disjoint and $\big|\bigcup_{j=1}^m(v_j+Q)\big|>\big(1-\frac{\alpha}{4}\big)|\Psi_i|$. Once again applying Lemma 5.8, find N such that for any $n\geq N$ there exist $w_1,\ldots,w_l\in\Phi_n$ for which $w_1+P_1,\ldots,w_l+P_l\subseteq\Phi_n$, the sets w_j+P_i are pairwise disjoint and $\big|\bigcup_{j=1}^l(w_j+P_i)\big|>\big(1-\frac{\alpha}{4}\big)|\Phi_n|$.

Now let $n\geq N$ and $E\subseteq\Phi_n$ with $|E|\geq\alpha|\Phi_n|$. Then for some $w=w_j$, $\big|E\cap(w+P_i)\big|>\frac{3\alpha}{4}|P_i|$. Putting $D=\zeta^{-1}\big((E-w'')\cap P_i\big)$ we have $|D|\geq\frac{3\alpha}{4}|\Psi_i|$, thus $\big|D\cap(v+Q)\big|\geq\frac{\alpha}{2}|Q|$ for some $v=v_j\in\Psi_i$, so $\big|(D-v)\cap Q\big|\geq\frac{\alpha}{2}|Q|$. We can therefore find $u\in K$, $u\neq 0$, and $v'\in(D-v)$ such that $v'+\pi_i(u)\in(D-v)$ for all $i=1,\ldots,k$. It follows that $v'+v\in D$ and $(v'+v)+\pi_i(u)\in D$ for all $i=1,\ldots,k$. Hence, $w+\zeta(v'+v)\in E$ and

$$w+\zeta\big(v'+v+\pi_i(u)\big)=w+\zeta(v'+v)+\zeta(\pi_i(u))=w+\zeta(v'+v)+\varphi_i(u)\in E,$$

$i=1,\ldots,k$. ∎

5.10. Corollary of the proof. *Let M be a module over a countable integral domain K. For any $k,d\in\mathbb{N}$ and $\alpha>0$ there exists a finite set $J\subset K$, $0\notin J$, such that, whenever $\varphi_1,\ldots,\varphi_k$ are polynomials $K\longrightarrow M$ of formal degree $\leq d$ having zero constant term, and $E\subseteq M$ with $d^*(E)\geq\alpha$, there exist $u\in J$ and $w\in E$ such that $w+\varphi_1(u),\ldots,w+\varphi_k(u)\in E$.*

Indeed, in the proof of Theorem 5.7 u is chosen from a finite set, which depends only on $d_1, \dots, d_k$ and α.

5.11. An equivalent "ergodic" statement is

Theorem. *Let M be a module over a countable integral domain K. For any $\alpha > 0$ and $k, d \in \mathbb{N}$, there exists a finite set $J \subset K$, $0 \notin J$, and $\delta > 0$ such that for any measure preserving action U of M on a probability measure space $(X, \mathcal{B}, \mu)$, polynomials $\varphi_1, \dots, \varphi_k\colon K \longrightarrow M$ of formal degree $\leq d$ and with zero constant term and $B \in \mathcal{B}$ with $\mu(B) \geq \alpha$ there exists $u \in J$ such that $\mu\big(\bigcap_{i=1}^k U(\varphi_i(u))B\big) > \delta$.*

Proof. One need only to modify slightly the argument given in 5.3. By ergodic decomposition, there exists a subset $B' \subseteq B$ such that $\mu(B') \geq \frac{\mu(B)}{2}$ having the property that $d^*(E_x) \geq \frac{\mu(B)}{2}$ for $x \in B'$. Choose J from Corollary 5.10 (with $\alpha = \frac{\mu(B)}{2}$). For all $x \in B'$ there exist $w_x \in E_x$ and $u_x \in J$ such that $w_x + \varphi_i(u_x) \in E_x$, $i = 1, 2, \dots, k$, hence for some fixed $u \in J$, $\mu\big(\{x \in B' : u_x = u\}\big) \geq \frac{\mu(B)}{2|J|}$. It follows that $\mu\big(\bigcap_{i=1}^k U(\varphi_i(u))B\big) \geq \frac{\mu(B)}{2|J|} = \delta$. ■

5.12. We may also restate Theorem 5.6 in a more geometric form (cf. Theorem PSZ from the introduction):

Theorem. *Let K be a countable integral domain, let M be a K-module and let φ be a polynomial $K^d \longrightarrow M$ with $\varphi(0) = 0$. For any finite set $Z \subset K^d$ and any $E \subseteq M$ with $d^*(E) > 0$ there exist $u \in K$, $u \neq 0$, and $w \in E$ such that $w + \varphi(uZ) \subset E$.*

Indeed, it suffices to apply Theorem 5.6 to the set $\{\varphi_z(u) = \varphi(uz),\ z \in Z\}$ of polynomials $K \longrightarrow M$.

5.13. In Theorem 5.6 one would like to know "how many" $u \in K$ satisfy the conclusion of the theorem. Our proof of Theorem 5.6 does not allow us to guarantee that such u form a syndetic set in K, but nevertheless we can show that, in a sense, this set is large:

Theorem. *Let K be a countable integral domain, let M be a K-module and let $\varphi_1, \dots, \varphi_k$ be polynomials $K \longrightarrow M$ with zero constant term. For any $E \subseteq M$ with $d^*(E) > 0$ there exists $r \in K$ such that the set*

$$S = \big\{u \in K :\ \textit{there exists } w \in E \textit{ such that } w + \varphi_i(u) \in E,\ i = 1, \dots, k\big\}$$

has positive upper density with respect to any Følner sequence in the ideal rK.

Proof. Given $a \in K$, consider the polynomials $\varphi_i'(x) = \varphi_i(ax)$, $i = 1, \dots, k$. By Corollary 5.10, there exists a finite set $J \subset K$, which does not depend on a, such that $\varphi(au) + w = \varphi_i'(u) + w \in E$ for some $u = u(a) \in J$ and $w \in E$. Let $r \in J$ be such that the set $P = \{a \in K : u(a) = r\}$ has positive upper density in K. Then $rP \subseteq S$ and has positive upper density in rK. ■

In particular, S is infinite. If K has the property that any of its principal ideals has finite index (which is the case, for example, when $K=\mathbb{Z}$ or $K=F[x]$, where F is a finite field), then S has positive upper density in K.

5.14. One more question arising in connection with Theorem 5.6 is whether one can get "nontrivial" configurations of the form $\{w, w+\varphi_1(u), \ldots, w+\varphi_k(u)\}$ in E. Indeed, the condition $u \neq 0$ does not guarantee yet that $\varphi_i(u) \neq 0$ in M. In some situations the existence of a nontrivial configuration is clear: if, say, M is a free module, the polynomials $\varphi_1, \ldots, \varphi_k$ have only finitely many roots, whereas the set S of those u which satisfy the conclusion of the theorem is infinite by Theorem 5.13.

5.15. We will now derive a couple of corollaries from Theorem 5.7. (The first one was proved in [BL1], the second one is new.)

Corollary. *Let $\mathcal{P}$ be a finite family of polynomials with integer coefficients and zero constant term. For any $\alpha>0$ there exists $C \in \mathbb{N}$ such that whenever $p \in \mathbb{N}$, $p \geq C$, and $E \subseteq \{1, \ldots, p\}$, $|E| \geq \alpha p$, there exist $u \in \{1, \ldots, p\}$ and $w \in E$ such that $w+\varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

5.16. Corollary. *Let K be a countable or finite integral domain and let $\mathcal{P}$ be a finite family of polynomials over the ring $K[x]$ with zero constant term. For any $\alpha>0$ there exists $d \in \mathbb{N}$ such that whenever $E \subseteq K$ with $d^*(E) \geq \alpha$ there exist $u \in K$, $u \neq 0$, $\deg u \leq d$, and $w \in E$ such that $w+\varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

5.17. We conclude the paper by formulating some new applications to finite fields.

Theorem. *Let L be a finite field, let $\mathcal{P}$ be a finite family of polynomials over L with zero constant term and let $\alpha>0$. There exists $m \in \mathbb{N}$ such that whenever F is a finite extension of L with $|F| \geq m$ and $E \subseteq F$ with $|E| \geq \alpha|F|$, there exist $u \in F$, $u \neq 0$, and $w \in E$ such that $w+\varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

Proof. Any finite extension of L has the form $F=L[x]/(v)$, where $v \in L[x]$ is an irreducible polynomial. We have $|F|=|L|^{\deg v}$; assume that $|F|>|L|^d$, where d is the constant from Corollary 5.16. Then $\deg v>d$. Let π be the factorization mapping $L[x] \longrightarrow F$. If $E \subseteq F$ with $|E| \geq \alpha|F|$, then $d^*(\pi^{-1}(E)) \geq \alpha$ and so, by Corollary 5.16, there exist $u' \in L[x]$, $u' \neq 0$ and $\deg u' \leq d$, and $w' \in \pi^{-1}(E)$ such that $w'+\varphi(u') \in \pi^{-1}(E)$ for all $\varphi \in \mathcal{P}$. Since $\deg u' \leq d<\deg v$, $u' \notin (v)$, so $\pi(u') \neq 0$. Hence, $u=\pi(u')$ and $w=\pi(w')$ satisfy the conclusion of the theorem. ■

5.18. Combining Theorem 5.17 and Corollary 5.15, we get:

Theorem. *Let $\mathcal{P}$ be a finite family of polynomials with integer coefficients and zero constant term and let $\alpha>0$. There exists $N \in \mathbb{N}$ such that whenever F is a field with $|F| \geq N$ and $E \subseteq F$ with $|E| \geq \alpha|F|$, there exist $u \in F$, $u \neq 0$, and $w \in E$ such that $w+\varphi(u) \in E$ for all $\varphi \in \mathcal{P}$.*

Proof. Let C be the constant from Corollary 5.15, let $\{p_1=2, p_2=3, \ldots, p_k\}$ be the set of all prime numbers $<C$, and for each p_i, $i=1, \ldots, k$, let m_i be the number whose

existence is guaranteed by Theorem 5.17 for $L = \mathbb{Z}_{p_i}$. Now let F be a finite field with $|F| \geq \max\{m_1, \ldots, m_k\}$ and let $E \subseteq F$ with $|E| \geq \alpha|F|$. If $p = \operatorname{char} F \geq C$, partition $F = \bigcup_{j=1}^{r} J_j$, where J_j are cosets of $\mathbb{Z}_p \in F$. Then for one of these cosets one has $|J_j \cap E| \geq \alpha|J_j|$, and the existence of the desired $u \in \mathbb{Z}_p$ and $w \in E$ follows from Corollary 5.15 in this case. If $\operatorname{char} F < C$, the existence of u and w follows from Theorem 5.17. ■

Bibliography

[B1] V. Bergelson, Weakly mixing PET, *Ergodic Theory and Dynam. Syst.* **7** (1987), 337–349.

[B2] V. Bergelson, Ergodic theory and Diophantine problems, *Topics in symbolic dynamics and applications,* London Math. Soc. Lecture Note Series **279** (2000), 167–205.

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

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

[BM] V. Bergelson and R. McCutcheon, Uniformity in polynomial Szemerédi theorem, *Ergodic Theory of $\mathbb{Z}^d$-actions,* London Math. Soc. Lecture Note Series **228** (1996), 273–296.

[BMZ] V. Bergelson, R. McCutcheon and Q. Zhang, A Roth theorem for amenable groups, *Amer. J. Math.* **119** (1997), 1173–1211.

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

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

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

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

[FK3] H. Furstenberg and Y. Katznelson, A density version of the Hales-Jewett theorem, *J. d'Analyse Math.* **57** (1991), 64–119.

[G] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11,** no. 3 (2001), 465–588.

[La] P. Larick, Ph.D. Thesis, The Ohio State University, 1998.

[Le] A. Leibman, Multiple recurrence theorem for measure preserving actions of a nilpotent group, *Geom. and Funct. Anal.* **8** (1998), 853–931.

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

[Z1] R. Zimmer, Extensions of ergodic group actions, *Illinois J. Math.* **20** (1976), 373-409.

[Z2] R. Zimmer, Ergodic actions with generalized discrete spectrum, *Illinois J. Math.* **20** (1976), 555-588.