CONVERGENCE OF POLYNOMIAL ERGODIC AVERAGES OF SEVERAL VARIABLES FOR SOME COMMUTING TRANSFORMATIONS

MICHAEL C. R. JOHNSON

Department of Mathematics
Northwestern University
Evanston, IL 60201

Abstract. Let $(X, \mathcal{B}, \mu)$ be a probability space and let $T_1, \ldots, T_l$ be l commuting invertible measure preserving transformations of X. We show that if the group of transformations generated by $T_1, \ldots, T_l$ is totally ergodic, then the averages $\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\prod_{i=1}^r T_1^{p_{i1}(u)}\ldots T_l^{p_{il}(u)}f_i$ converge in $L^2(\mu)$ for all polynomials $p_{ij}\colon \mathbb{Z}^d \to \mathbb{Z}$, all $f \in L^\infty(\mu)$ and all Følner sequences $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$.

1. INTRODUCTION

In 1996, Bergelson and Leibman proved the following generalization of Furstenberg's Multiple Recurrence Theorem [Fu1], corresponding to the multidimensional polynomial version of Szemerédi's theorem.

Theorem 1.1. [BL] *Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T_1, \ldots, T_l$ be commuting invertible measure preserving transformations of X, let $p_{ij}\colon \mathbb{Z} \to \mathbb{Z}$ be polynomials satisfying $p_{ij}(0) = 0$ for all $1 \le i \le r, 1 \le j \le l$, and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \mu\Big(\bigcap_{i=1}^{r} T_1^{-p_{i1}(n)}\ldots T_l^{-p_{il}(n)}A\Big) > 0.$$

Furstenberg's theorem corresponds to the case that $p_{ij}(n) = n$ for $i = j$, $p_{ij}(n) = 0$ for $i \neq j$ and each $T_i = T_1^i$. In this linear case, Host and Kra [HK1] showed that the lim inf is in fact a limit. Host and Kra [HK2] and Leibman [Le2] proved convergence in the polynomial case assuming all $T_i = T_1$. It is natural to ask whether the general commuting averages for polynomials in Theorem 1.1 converge.

Definition 1.2. We say $T_1, \ldots, T_l$ is a **totally ergodic** group of transformations of X if each $T_1^{c_1}T_2^{c_2}\ldots T_l^{c_l}$ is ergodic for any choice of $(c_1, \ldots, c_l) \neq (0, \ldots, 0)$.

We show that given a totally ergodic group of transformations, we obtain convergence in $L^2(\mu)$. We prove a statement replacing indicator functions with arbitrary functions in $L^\infty(\mu)$.

Theorem 1.3. *Let* $(X, \mathcal{B}, \mu)$ *be a probability space, let* $T_1, \dots, T_l$ *be a totally ergodic group of commuting invertible measure preserving transformations of* X*, and let* $p_{ij}\colon \mathbb{Z}^d \to \mathbb{Z}$ *for* $1 \le i \le r, 1 \le j \le l$ *be polynomials. For any* $f_1, \dots, f_r \in L^\infty(\mu)$ *and any Følner sequence* $\{\Phi_N\}_{N=1}^\infty$ *in* $\mathbb{Z}^d$*, the averages*

$$(1) \quad \frac{1}{|\Phi_N|} \sum_{u \in \Phi_N} \prod_{i=1}^{r} f_i(T_1^{p_{i1}(u)} \dots T_l^{p_{il}(u)} x)$$

converge in $L^2(\mu)$ *as* $N \to \infty$.

Without the assumption of total ergodicity, convergence for the above averages (1) remains open and is only known in the linear case. Frantzikinakis and Kra [FrK] showed that given $p_{ij}(n) = n$ for $i = j$ and $p_{ij}(n) = 0$ for $i \neq j$, if we assume that T_i is ergodic for each $i \in \{1, \dots, l\}$ and $T_i T_j^{-1}$ is ergodic for all $i \neq j$, we obtain convergence in $L^2(\mu)$. Tao [Ta] recently proved convergence in $L^2(\mu)$ for the general linear case without the ergodicity assumptions needed in [FrK].

In previous results, convergence was often shown by proving that the averages in (1) do not change by replacing each function with its conditional expectation on a certain characteristic factor, namely an inverse limit of nilsystems. This characteristic factor, is then shown to have algebraic structures for which convergence is known. We define these terms precisely in the section below. To prove our theorem, we combine this technique with PET-induction as introduced by Bergelson [Be].

2. Preliminaries

For simplicity, we assume all functions are real valued. All theorems and definitions hold for complex valued functions with obvious minor modifications. Throughout, we use the notation $Tf = f(T)$.

2.1. **Nilsystems.**

Definition 2.1. Let G be a k-step nilpotent Lie group, let Γ be a discrete cocompact subgroup of G, and let $X = G/\Gamma$. For each $g \in G$, let $T_g : G/\Gamma \to G/\Gamma$ be defined by $T_g(x\Gamma) = gx\Gamma$, and let μ be Haar measure, the unique invariant measure under left translations by elements in G. We call $(X, \mathcal{B}, \mu, (T_g, g \in G))$ a **nilsystem.**

Definition 2.2. A sequence of finite subsets $\{\Phi_N\}_{N=1}^{\infty}$ of a group G is a **Følner sequence** if for all $g \in G$,

$$\lim_{n\to\infty} \frac{|g\Phi_n \Delta \Phi_n|}{|\Phi_n|} = 0,$$

where Δ is the symmetric difference operation.

Ergodic averages in nilsystems have been well studied. We make use of the following theorem of Leibman:

Theorem 2.3. [Le1] *Let* $(X, \mathcal{B}, \mu, (T_g, g \in G))$ *be a nilsystem with* $X = G/\Gamma$, $g_1, \dots, g_l \in G$, *and* $p_1, \dots, p_l \colon \mathbb{Z}^d \to \mathbb{Z}$ *be polynomials. Then for any* $f \in C(X)$ *and any Følner sequence* $\{\Phi_N\}_{N=1}^{\infty}$ *in* $\mathbb{Z}^d$, *the averages*

$$\frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} T_{g_1}^{p_1(u)} \dots T_{g_l}^{p_l(u)} f$$

converge pointwise as $N \to \infty$.

Corollary 2.4. *Let* $(X, \mathcal{B}, \mu, (T_g, g \in G))$ *be a nilsystem with* $X = G/\Gamma$, $g_1, \dots, g_l \in G$, *and* $p_{ij} \colon \mathbb{Z}^d \to \mathbb{Z}$ *for* $1 \le i \le r, 1 \le j \le l$ *be polynomials. Then for any* $f_1, \dots, f_r \in L^\infty(\mu)$ *and any Følner sequence* $\{\Phi_N\}_{N=1}^{\infty}$ *in* $\mathbb{Z}^d$, *the averages*

$$\frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \prod_{i=1}^{r} T_{g_1}^{p_{i1}(u)} \dots T_{g_l}^{p_{il}(u)} f_i$$

converge in $L^2(\mu)$ *as* $N \to \infty$.

Proof. By applying Theorem 2.3 to X^{r+l},

$$\hat{T}_1 = \dots = \hat{T}_{r+l} = T_{g_1} \times \dots \times T_{g_l} \times Id_X \times \dots \times Id_X,$$

and $f = f_1 \otimes \dots \otimes f_r \otimes 1 \otimes \dots \otimes 1$, we obtain the pointwise averages for $f_1, \dots, f_r \in C(X)$. Using the density of $C(X)$ in $L^\infty(\mu)$, $L^2(\mu)$ convergence follows for arbitrary $f_1, \dots, f_r \in L^\infty(\mu)$. $\square$

2.2. **The Host-Kra seminorms $|\!|\!|\cdot|\!|\!|_k$.** We briefly review the construction of the Host-Kra seminorms on $L^\infty(\mu)$ from [HK1]. As our setting deals with multiple commuting transformations, we must specify which transformation is used. In this section, T is an ergodic measure preserving transformation of $(X, \mathcal{B}, \mu)$.

For each $k \ge 0$ we define a probability measure $\mu_T^{[k]}$ on $X^{[k]} = X^{2^k}$, invariant under $T^{[k]} = T \times \cdots \times T$ (2^k times).

Set $\mu_T^{[0]} = \mu$. For $k \geq 0$, let $\mathcal{I}_T^{[k]}$ be the σ-algebra of $T^{[k]}$-invariant subsets of $X^{[k]}$. Then define $\mu_T^{[k+1]} = \mu_T^{[k]} \times_{\mathcal{I}_T^{[k]}} \mu_T^{[k]}$ to be the relatively independent square of $\mu_T^{[k]}$ over $\mathcal{I}_T^{[k]}$. This means for $F, G \in L^\infty(\mu^{[k]})$

$$\int_{X^{[k+1]}} F(\mathbf{x}')G(\mathbf{x}'')d\mu_T^{[k+1]}(\mathbf{x}', \mathbf{x}'') = \int_{X^{[k]}} \mathbb{E}(F|\mathcal{I}_T^{[k]})\mathbb{E}(G|\mathcal{I}_T^{[k]})d\mu_T^{[k]}.$$

Using these measures, define

$$|||f|||_{k,T}^{2^k} = \int_{X^{[k]}} \prod_{j=0}^{2^k-1} f(x_j) d\mu_T^{[k]}(\mathbf{x})$$

for a bounded function $f \in L^\infty(\mu)$ and $k \geq 1$. It is shown in [HK1] that for every $k \geq 1$ and every ergodic T, $|||\cdot|||_{k,T}$ is a seminorm on $L^\infty(\mu)$. Also, for $f \in L^\infty(\mu)$, we have $|||f|||_{1,T} = |\int f d\mu|$ and for every $k \geq 1$, $|||f|||_{k,T} \leq |||f|||_{k+1,T} \leq \| f \|_{L^\infty(\mu)}$.

2.3. **The Host-Kra factors $Z_k(X)$.** We now define an increasing sequence of factors $\{Z_k(X,T) : k \geq 0\}$ as constructed in [HK1]. Let $\mathcal{Z}_k(X,T)$ be the T-invariant sub-σ-algebra characterized by the following property: for every $f \in L^\infty(\mu)$, $\mathbb{E}(f|\mathcal{Z}_k(X,T)) = 0$ if and only if $|||f|||_{k+1,T} = 0$. We define $Z_k(X,T)$ to be the factor of X associated to the sub-σ-algebra $\mathcal{Z}_k$. Thus $Z_0(X,T)$ is the trivial factor and $Z_1(X,T)$ is the Kronecker factor. *A priori*, these constructions depend on the transformation T.

Indeed, the following observation of Frantzikinakis and Kra shows that given basic assumptions, none of the previous constructions depend on the transformation T.

Proposition 2.5. [FrK] *Assume that T and S are ergodic commuting invertible measure preserving transformations of a space $(X, \mathcal{B}, \mu)$. Then for all $k \geq 1$ and all $f \in L^\infty(\mu)$, $|||f|||_{k,T} = |||f|||_{k,S}$ and $Z_k(X,T) = Z_k(X,S)$.*

Thus we discard T from our notation.

Definition 2.6. We call a probability space $(X, \mathcal{B}, \mu)$ with l invertible commuting measure preserving transformations $T_1, \ldots, T_l$, an **(invertible commuting measure preserving) system**. If the group of transformations is also totally ergodic, then we call it a **totally ergodic system**. We denote it as $(X, \mathcal{B}, \mu, (T_1, \ldots, T_l))$. A system $(X, \mathcal{B}, \mu, (T_1, \ldots, T_l))$ is an **inverse limit** of systems $(X, \mathcal{B}_i, \mu_i, (T_1, \ldots, T_l))$ if each $\mathcal{B}_i \subset \mathcal{B}_{i+1}$ and $\mathcal{B} = \bigvee_{i=1}^\infty \mathcal{B}_i$ up to sets of measure zero.

The main result of the Host-Kra theory is that each of the factors (Z_k, T_i) is isomorphic to an inverse limit of k-step nilsystems. However, such isomorphism *a priori* depends on the transformation T_i. In [FrK], they deal specifically with this technicality. We say that a system $(X, \mathcal{B}, \mu, (T_1, \ldots, T_l))$ has **order k** if $X = Z_k(X)$. (Note that by Proposition 2.5, $Z_k(X, T_i)$, does not depend on i).

Theorem 2.7. [FrK] *Any system $(X, \mathcal{B}, \mu, (T_1, \ldots, T_l))$ of order k is an inverse limit of a sequence of systems $(X, \mathcal{B}_i, \mu_i, (T_1, \ldots, T_l))$, each arising from k-step nilsystems, where $X = G_i/\Gamma_i$ and each transformation $T_1, \ldots, T_l$ is a left translation of G_i/Γ_i by an element in G_i.*

By combining Theorem 2.7 and Corollary 2.4, Theorem 1.3 is proved in the case that $X = Z_k(X)$ for some k.

2.4. Characteristic factors.

Definition 2.8. We say a sub-σ-algebra $\mathcal{X} \subseteq \mathcal{B}$ is a **characteristic factor for $L^2(\mu)$-convergence** of the averages

$$(1)\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\prod_{i=1}^{r} T_1^{p_{i1}(u)}\ldots T_l^{p_{il}(u)} f_i$$

if $\mathcal{X}$ is T_j invariant for all $1 \leq j \leq l$ and the averages in (1) converge to 0 in $L^2(\mu)$ for any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$ whenever $\mathbb{E}(f_i|\mathcal{X}) = 0$ for some $1 \leq i \leq r$.

Using the multilinearity of our averages in (1), it only remains to show that for some $k \in \mathbb{N}$, $Z_k(X)$ is a characteristic factor.

We say the set of polynomials $\{p_{ij} \colon \mathbb{Z}^d \to \mathbb{Z}$ for $1 \leq i \leq r, 1 \leq j \leq l\}$ is an **ED-set** if at least one polynomial p_{ij} is non-zero, and each $p_{i_0 j_0}$ and $p_{i_0 j_0} - p_{i_1 j_1}$ have positive degree or are identically zero for all i_0, j_0, i_1, j_1. We note that Theorem 1.3 is trivially true if all the polynomials are identically zero. So, by replacing each f_i with $T_1^{c_1} \ldots T_l^{c_l} f_i$ for some $c_1, \ldots, c_l \in \mathbb{Z}$, we may assume that our set of polynomials is an ED-set. Thus the main theorem is a consequence of the following:

Proposition 2.9. *Let $(X, \mathcal{B}, \mu, T_1, \ldots, T_l)$ be a totally ergodic system and $\{p_{ij} \colon \mathbb{Z}^d \to \mathbb{Z}$ for $1 \leq i \leq r, 1 \leq j \leq l\}$ be an ED-set of polynomials. Then there exists $k \in \mathbb{N}$ such that for any $f_1, \ldots, f_r \in L^\infty(\mu)$ with $|||f_m|||_k = 0$ for some $1 \leq m \leq r$, we have*

$$\limsup_{N\to\infty}\left\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}(\prod_{i=1}^{r} T_1^{p_{i1}(u)} T_2^{p_{i2}(u)} \ldots T_l^{p_{il}} f_i)\right\|_{L^2(\mu)} = 0$$

for any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$.

We note that the above integer k is only dependent on the set of polynomials $\{p_{ij}\}$ and not on the system $(X, \mathcal{B}, \mu, (T_1, \ldots, T_l))$ or the dimension d. By relabeling our polynomials and functions, we need only prove Proposition 2.9 in the case that $\|\!|f_1|\!\|_k = 0$ for some $k \in \mathbb{N}$.

3. Linear case

To prove proposition 2.9, we use PET-induction as introduced by Bergelson in [Be]. In this section we prove the base case of the induction.

Proposition 3.1. *Let $(X, \mathcal{B}, \mu, (T_1, \ldots, T_l))$ be a totally ergodic system and $\{p_{ij} \colon \mathbb{Z}^d \to \mathbb{Z}$ for $1 \leq i \leq r, 1 \leq j \leq l\}$ be an ED-set of linear functions. Then there exists a constant $C > 0$ dependent only on the set of polynomials, such that*

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} (\prod_{i=1}^{r} T_1^{p_{i1}(u)} T_2^{p_{i2}(u)} \ldots T_l^{p_{il}(u)} f_i) \Big\|_{L^2(\mu)} \leq C \min_{1\leq i\leq r} \|\!| f_i \|\!|_{r+1}$$

for any $f_1, \ldots, f_r \in L^\infty(\mu)$ with $\|f_i\|_{L^\infty(\mu)} \leq 1$ and any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$.

As a corollary, we get that $Z_r(X)$ is characteristic for the averages in (1). We use the following version of the van der Corput lemma in the inductive process to reduce each average to a previous step.

Lemma 3.2. [BMZ] *Let $\{g_u\}_{u\in G}$ be a bounded family of elements of a Hilbert space $\mathcal{H}$ indexed by elements of a finitely generated abelian group G and let $\{\Phi_N\}_{N=1}^\infty$ be a Følner sequence in G.*

(1) *For any finite set $F \subseteq G$,*

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} g_u \Big\|^2 \leq \limsup_{N\to\infty} \frac{1}{|F|^2} \sum_{v,w\in F} \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \langle g_{u+v}, g_{u+w} \rangle.$$

(2) *There exists a Følner sequence $\{\Theta_M\}_{M=1}^\infty$ in G^3 such that*

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} g_u \Big\|^2 \leq \limsup_{M\to\infty} \frac{1}{\Theta_M} \sum_{(u,v,w)\in\Theta_M} \langle g_{u+v}, g_{u+w} \rangle.$$

Leibman proved the following lemma in his proof of convergence for a single transformation [Le2]. We likewise use his lemma to prove the linear case for multiple commuting transformations.

Lemma 3.3. [Le2]

(1) *Let $p_i\colon \mathbb{Z}^d \to \mathbb{Z}$ be nonconstant linear functions for each $i = 1, \dots, l$. There exists a constant C, such that for any $f \in L^\infty(\mu)$ and any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$,*

$$\lim_{N\to\infty} \left\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} T_1^{p_1(u)} \dots T_l^{p_l(u)} f \right\|_{L^2(\mu)} \le C |\!|\!| f |\!|\!|_2.$$

(2) *Let $p_i\colon \mathbb{Z}^d \to \mathbb{Z}$ be nonconstant linear functions for each $i = 1, \dots, l$. There exists a constant C, such that for any $f \in L^\infty(\mu)$ and any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$,*

$$\lim_{N\to\infty} \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} |\!|\!| f \cdot T_1^{p_1(u)} \dots T_l^{p_l(u)} f |\!|\!|_k^{2^k} \le C |\!|\!| f |\!|\!|_{k+1}^{2^{k+1}}.$$

We note here that part (1) of Lemma 3.3 is similar to Lemma 7 in [Le2] but with multiple commuting transformations. The only step needed to alter his proof is to show our average also convergences to the conditional expection onto the appropriate sub-σ-algebra of f. But this follows from classical results on convergence for amenable group actions.

Proof of Proposition 3.1. To simplify notation, we write $T_1^{p_{i1}(u)} \dots T_l^{p_{il}(u)}$ as $S^{p_i(u)}$. Since each p_{ij} is a linear polynomial, we have $S^{p_i(u)} S^{p_j(u)} = S^{p_i + p_j(u)}$

We proceed by induction on r. For $r = 1$, we are done by Lemma 3.3. Assume the proposition holds for $r-1$ functions. Let $f_1, \dots, f_r$ be essentially bounded functions on X with $\|f_i\|_{L^\infty(\mu)} \le 1$ for all $1 \le i \le r$, and let $\{\Phi_N\}_{N=1}^\infty$ be a Følner sequence in $\mathbb{Z}^d$. By applying Lemma 3.2 to $g_u = S^{p_i(u)} f_1 \dots S^{p_r(u)} f_r$, for any finite $F \subseteq \mathbb{Z}^d$, we get

$$\begin{aligned}
&\limsup_{N\to\infty} \left\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \prod_{i=1}^r S^{p_i(u)} f_i \right\|_{L^2(\mu)}^2 \\
&\le \limsup_{N\to\infty} \frac{1}{|F|^2} \sum_{v,w\in F} \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \int_X \prod_{i=1}^r S^{p_i(u+v)} f_i \cdot \prod_{i=1}^r S^{p_i(u+w)} f_i d\mu \\
&= \limsup_{N\to\infty} \frac{1}{|F|^2} \sum_{v,w\in F} \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \int_X \prod_{i=1}^{r-1} S^{p_i(u)} S^{-p_r(u)} (S^{p_i(v)} f_i \\
&\qquad \cdot S^{p_i(w)} f_i) \cdot (S^{p_r(v)} f_r \cdot S^{p_r(w)} f_r) d\mu \\
&\le \frac{1}{|F|^2} \sum_{v,w\in F} \limsup_{N\to\infty} \left\| \frac{1}{\Phi_N} \sum_{u\in\Phi_N} \prod_{i=1}^{r-1} S^{p_i - p_r(u)} (S^{p_i(v)} f_i \cdot S^{p_i(w)} f_i) \right\|_{L^2(\mu)}.
\end{aligned}$$

By the induction process, there exists a constant C, independent of $f_1, \dots, f_r$ and $\{\Phi_N\}_{N=1}^{\infty}$, such that

$$\limsup_{N\to\infty} \Big\| \frac{1}{\Phi_N} \sum_{u\in\Phi_N} \prod_{i=1}^{r-1} S_i^{p_i(u)} S_r^{-p_r(u)} (S_i^{p_i(v)} f_i \cdot T_i^{p_i(w)} f_i) \Big\|_{L^2(\mu)} \\ \leq C \vert\!\vert\!\vert (S_i^{p_i(v)} f_i \cdot S_i^{p_i(w)} f_i) \vert\!\vert\!\vert_r$$

for all $v, w \in \mathbb{Z}^d$ and $i \in \{1, \dots, r\}$. Thus for any finite set $F \subset \mathbb{Z}^d$ and $i \in \{1, \dots, r\}$,

$$\begin{aligned} \limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \prod_{i=1}^{r} S_i^{p_i(u)} f_i \Big\|_{L^2(\mu)} & \\ \leq \Big(\frac{C}{|F|^2} \sum_{v,w\in F} \vert\!\vert\!\vert (S_i^{p_i(v)} f_i \cdot S_i^{p_i(w)} f_i) \vert\!\vert\!\vert_r \Big)^{1/2} & \\ \leq C^{1/2} \Big(\frac{1}{|F|^2} \sum_{v,w\in F} \vert\!\vert\!\vert (f_i \cdot S_i^{p_i(w-v)} f_i) \vert\!\vert\!\vert_r^{2^r} \Big)^{(1/2)^{r+1}} & . \end{aligned}$$

Let $\{\Psi_N\}_{N=1}^{\infty}$ be any Følner sequence in $\mathbb{Z}^d$. Thus $\{\Psi_N \times \Psi_N\}_{N=1}^{\infty}$ is a Følner sequence in $\mathbb{Z}^{2d}$. By Lemma 3.3 we have for each $i \in \{1, \dots, r\}$

$$\limsup_{M\to\infty} \frac{1}{|\Psi_M|^2} \sum_{v,w\in\Psi_M} \vert\!\vert\!\vert f_i \cdot S_i^{p_i(w-v)} f_i \vert\!\vert\!\vert_r^{2^r} \leq c \vert\!\vert\!\vert f_i \vert\!\vert\!\vert_{r+1}^{2^{r+1}}$$

with c independent of f_i. By replacing F with Ψ_N for each $N \in \mathbb{N}$, we get

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \prod_{i=1}^{r} S_i^{p_i(u)} f_i \Big\|_{L^2(\mu)} \leq C^{1/2} c^{(1/2)^{r+1}} \min_{i\leq r} \vert\!\vert\!\vert f_i \vert\!\vert\!\vert_{r+1}.$$

□

4. PET-Induction

We now deal with the inductive step. A set of polynomials $P = \{p_{ij} : 1 \leq i \leq r, 1 \leq j \leq l\}$ where each $p_{ij} \colon \mathbb{Z}^d \to \mathbb{Z}$ is called a **(integer) polynomial family**. We define the **degree** of a family P,

$$\deg(P) = \max_{i\leq r, j\leq l} \deg(p_{ij}).$$

We define the **column degree** of a polynomial family P to be the vector $C(P) = (c_1, \dots, c_l)$ where each

$$c_j = \max_{1\leq i\leq r} \deg(p_{ij}).$$

We say that two polynomials p, q are equivalent if $\deg(p) = \deg(q)$ and $\deg(p-q) < \deg(p)$. Thus any collection of polynomials can be partitioned into equivalence classes. We define the degree of an equivalence class of polynomials to be equal to the degree of any of its representatives.

We define the **weight** of a column j, to be the vector $w_j(P) = (w_{1j}, \dots, w_{\deg(P)j})$ with each w_{ij} equal to the number of equivalence classes in P of degree i in column j. Given two vectors $\mathbf{v} = (v_1, \dots, v_z)$, $\mathbf{v}' = (v'_1 \dots, v'_{z'})$, we say $\mathbf{v} < \mathbf{v}'$ if $z < z'$ or if $z = z'$ and there exists n_0 such that $v_{n_0} < v'_{n_0}$ and for each $n > n_0$, $v_n = v'_n$. Thus the set of weights and the set of column degrees become well ordered sets. A polynomial family $P = \{p_{ij}\}$ is said to be **standard** if it is an ED-set and $\deg(p_{1j}) = \deg(P)$ for some $1 \le j \le l$. We now state Proposition 2.9 in the case that P is standard.

Proposition 4.1. *Let $(X, \mathcal{B}, \mu, (T_1, \dots, T_l))$ be a totally ergodic system and $P = \{p_{ij} \colon 1 \le i \le r, 1 \le j \le l\}$ be a standard polynomial family. Then there exists $k \in \mathbb{N}$ such that for any $f_1, \dots, f_r \in L^\infty(\mu)$ with $|||f_1|||_k = 0$, we have*

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} (\prod_{i=1}^{r} T_1^{p_{i1}(u)} T_2^{p_{i2}(u)} \dots T_l^{p_{il}(u)} f_i) \Big\|_{L^2(\mu)} = 0$$

for any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$.

To prove Proposition 4.1, we construct a sequence of polynomial families that controls the above averages, where after finitely many steps, we obtain a family of smaller column degree. To accomplish this, we first fix a column j^* that contains a polynomial of maximal degree. Using PET-induction on the weight of this column, we show that indeed the column degree is eventually reduced. This process is similar to the PET-induction process used in [Le2] for a single transformation. We then proceed with a second induction on the column degree to construct a linear polynomial family that controls the above averages.

Example. We now illustrate this process for the following simple example. We start with the standard polynomial family

$$P_1 = \begin{pmatrix} n^2 & n^2 \\ n & 2n \end{pmatrix}$$

which has degree $\deg(P_1) = 2$, column degree $C(P_1) = (2,2)$ and column weights $w_1(P_1) = (1,1)$, and $w_2(P_1) = (1,1)$.

Next, we have the polynomial family

$$P_2 = \begin{pmatrix} n^2 - n - 2 & n^2 - 2n - 4 \\ n^2 + 3n + 2 & n^2 + 2n \end{pmatrix},$$

which has degree $\deg(P_2) = 2$, column degree $C(P_2) = (2, 2)$ and column weights $w_1(P_2) = (0, 1)$, and $w_2(P_2) = (0, 1)$.

Finally, the polynomial family

$$P_3 = \begin{pmatrix} -2n - 4 & -2n - 5 \\ -4n - 4 & -4n - 4 \\ 2n + 4 & 2n + 3 \end{pmatrix}$$

has degree $\deg(P_3) = 1$, column degree $C(P_3) = (1, 1)$ and column weights $w_1(P_3) = (3)$, and $w_2(P_3) = (3)$.

4.1. **Inductive Polynomial Families.** We begin by defining that a certain property holds for **almost all** $v \in \mathbb{Z}^d$ if the set of elements for which the property does not hold is contained in a set of zero density with respect to any Følner sequence in $\mathbb{Z}^d$. To show a property holds for almost all $v \in \mathbb{Z}^d$, we use the fact that a set of zeros of a nontrivial polynomial has zero density with respect to any Følner sequence in $\mathbb{Z}^d$.

Given any standard polynomial family P with $\deg(P) \geq 2$, for each $(v, w) \in \mathbb{Z}^{2d}$ we construct a new family $P_{v,w}$, as follows. Set j^* be the minimal j such that

$$\deg(p_{ij}) = \deg(P)$$

for some i. Define $\alpha_1 = j^*$, $\alpha_{j^*} = 1$, and $\alpha_j = j$ otherwise.

We now choose i_0 to be the row that we later subtract away to create the new polynomial family. Our goal is to choose i_0 appropriately such that $C(P)$ will be less for some future polynomial family.

If $\deg(p_{1j^*}) < \deg(P)$, choose i_0 to be the minimal i such that $\deg(p_{ij^*}) = \deg(P)$. When $\deg(p_{1j^*}) = \deg(P)$, the choice of i_0 is complicated by the fact that our column j^* may have many polynomials which are identically zero. We use the following notation to track such zeros. Let

$I_0 = \{i \in \{1, \ldots, r\} : p_{ij} = 0 \text{ for all } j = 1, \ldots, l\}$,
$I_1 = \{i \in \{1, \ldots, r\} : \deg(p_{ij}) \leq 1 \text{ for all } j = 1 \ldots, l\} \setminus I_0$, and
$I_2 = \{1, \ldots, r\} \setminus (I_0 \cup I_1)$.

Since P is an ED-set, I_1 is composed of nonzero rows containing only linear polynomials (or zero), while I_2 is composed of the rows that contain a polynomial of degree greater than 2. Define $H_0(P) = I_1 \cup I_2$ and inductively define

$$H_j(P) = \{i \in \{1, \ldots, r\} \colon p_{i\alpha_j} = 0\} \cap H_{j-1}(P)$$

for $1 \leq j \leq l$ (we omit the polynomial family P when there is no confusion which family we are dealing with). Thus, H_j records which non-identically zero rows have zeros in columns $\alpha_1, \ldots, \alpha_j$. Pick j_0 to be the smallest $j \geq 1$ such that $H_j = \emptyset$. In the case that column j^* has no zero entries, we see that $j_0 = 1$.

Then pick the smallest $i_0 \in H_{j_0-1}$ such that $p_{i_0\alpha_{j_0}}$ has minimal degree among all $p_{i\alpha_{j_0}}$ with $i \in H_{j_0-1}$. If all such $p_{i\alpha_{j_0}}$ have the same degree, pick the smallest i_0 so that $p_{i_0\alpha_{j_0}}$ is also not equivalent to $p_{1\alpha_{j_0}}$; if all such $p_{i\alpha_{j_0}}$ are equivalent, pick the smallest i_0 such that for some j, $\deg(p_{i_0j}) = \deg(P)$ and p_{i_0j} is not equivalent to p_{1j}. If all such p_{ij} are equivalent to p_{1j}, then choose the smallest i_0 such that there exists j such that $\deg(p_{1j}) = \deg(P)$, but $\deg(p_{i_0j}) < \deg(P)$. In the case that such i_0 does not exist, choose $i_0 = \min H_{j_0}$.

For all $(v, w) \in \mathbb{Z}^{2d}$, set $v^{ij} = w$ if $\deg(p_{ij}) = 1$, otherwise set $v^{ij} = v$. Therefore, given $(v, w) \in \mathbb{Z}^{2d}$, we define the new polynomial family

$$\begin{aligned}\bar{P}_{v,w} = \{p_{ij}(u + v^{ij}), p_{ij}(u + w) \colon i \in I_2, j = 1 \ldots, l\} \\ \bigcup \{p_{ij}(u + w) \colon i \in I_1, j = 1 \ldots, l\}\end{aligned}$$

where each of the polynomials $p_{ij}(u + v^{ij}), p_{ij}(u + w)$ are considered as polynomials in u.

We order the family

$$\bar{P}_{v,w} = \{q_{v,w,h,j} : 1 \leq h \leq s, 1 \leq j \leq l\}$$

in the following manner. We label each row

$$p_{i1}(u + v^{ij}), \ldots, p_{il}(u + v^{ij})$$

and

$$p_{i1}(u + w), \ldots, p_{il}(u + w)$$

as

$$q_{v,w,h,1}(u), \ldots, q_{v,w,h,l}(u)$$

for some unique $1 \leq h \leq s$ where $p_{1j}(u+v^{1j}) = q_{v,w,1,j}$ and $p_{i_0j}(u+w) = q_{v,w,s,j}(u)$.

Since for each vector (v, w) in $\mathbb{Z}^{2d}$, $p_{ij}(u+v), p_{ij}(u+w)$, and $p_{ij}(u)$ are all equivalent, the column degrees and the weight of column j for $(\bar{P}_{v,w})$, $(P_{v,w})$ are identical for all $1 \leq j \leq l$ and $(v, w) \in \mathbb{Z}^{2d}$. By construction, the first row of $\bar{P}_{v,w}$ also contains a polynomial of maximal degree and it is easy to check that $\bar{P}_{v,w}$ is an ED-set for each (v, w) outside a set of zeros of finitely many polynomials. Hence, $\bar{P}_{v,w}$ is a standard polynomial family for almost all $(v, w) \in \mathbb{Z}^{2d}$.

Next, for each $(v, w) \in \mathbb{Z}^{2d}$ we define the new family

$$P_{v,w} = \{q_{v,w,h,j} - q_{v,w,s,j} \colon 1 \leq h \leq s-1, 1 \leq j \leq l\}.$$

Example. In our previous example, we see that $P_2 = (P_1)_{v,w}$ with $(v, w) = (0, 2)$ and $P_3 = (P_2)_{v,w}$ with $(v, w) = (1, 0)$.

We would like to show that the new family $P_{v,w}$ is standard for almost all choices of (v, w). In fact, $P_{v,w}$ is an ED-set whenever $\bar{P}_{v,w}$ is. We now show that the first row in $P_{v,w}$ contains a polynomial of maximal degree. When $p_{i_0\alpha_1}$ is not equivalent to $p_{1\alpha_1}$, then $\deg(q_{v,w,1,\alpha_1} - q_{v,w,s,\alpha_1}) = \deg(p_{1\alpha_1}) = \deg(P_{v,w})$. However, when $p_{i_0\alpha_1}$ is equivalent to $p_{1\alpha_1}$, $\deg(q_{v,w,h,\alpha_1} - q_{v,w,s_a,\alpha_1}) = \deg(p_{1\alpha_1}) - 1$ for all $h \leq s-1$. However by the choice of i_0 either some other column j has $\deg(p_{1j}) = \deg(p_{1\alpha_1})$ or all columns have maximum degree $\deg(p_{1\alpha_1}) - 1$. In all cases, the new family $P_{v,w}$ is standard for almost all $(v, w) \in \mathbb{Z}^{2d}$. Also, since each new polynomial family is constructed only by subtracting polynomials within each column, the maximum degree in each column cannot increase. Therefore $C(P_{v,w}) \leq C(P)$. Thus, we have proved the following lemma.

Lemma 4.2. *Given any standard polynomial family P with* $\deg(P) \geq 2$*, for almost all $(v, w) \in \mathbb{Z}^{2d}$, $P_{v,w}$ is standard and has $C(P_{v,w}) \leq C(P)$.*

In the case where column j^* has any zero entries, we cannot expect that $w_{j^*}(P_{v,w}) < w_{j^*}(P)$ as in [Le2]. In the case that $P_{v,w}$ is standard and has $\deg(P_{v,w}) = 1$, then our process terminates. However, for each $(v_1, w_1) \in \mathbb{Z}^{2d}$ where the new polynomial family P_{v_1,w_1} is standard and has $\deg(P_{v_1,w_1}) \geq 2$, we may continue the inductive process to construct for each $(v_2, w_2) \in \mathbb{Z}^{2d}$ a new family

$$(P_{v_1,w_1})_{v_2,w_2}$$

which we notate as $P_{\mathbf{v}_2,\mathbf{w}_2}$, where $\mathbf{v}_2 = (v_1, v_2)$ and $\mathbf{w}_2 = (w_1, w_2)$. In the case that P_{v_1,w_1} is not standard, this leads our inductive process into a dead-end. However, Lemma 4.2 shows that this case is indeed exceptional, as for almost all choices of $(v_1, w_1) \in \mathbb{Z}^{2d}$, P_{v_1,w_1} is standard.

Similarly, when the family $P_{\mathbf{v}_t,\mathbf{w}_t}$ is standard with $\deg(P_{\mathbf{v}_t,\mathbf{w}_t}) \geq 2$, then for each $(v_{t+1}, w_{t+1}) \in \mathbb{Z}^{2d}$, we construct the new family $P_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}}$ where $\mathbf{v}_{t+1} = (v_1, \ldots, v_{t+1})$ and $\mathbf{w}_{t+1} = (w_1, \ldots, w_{t+1})$. We see from repeated application of Lemma 4.2 that at each step t in the process, for almost all choices of $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}$ is standard.

4.2. **Equivalence of Polynomial Families.** The set of equivalence classes, the degrees of each entry, and even the number of polynomials in $P_{\mathbf{v}_t,\mathbf{w}_t}$ depend on the choice of vector $(\mathbf{v}_t, \mathbf{w}_t)$. We will show that for almost all choices of $(\mathbf{v}_t, \mathbf{w}_t)$, the associated polynomial families share the above characteristics. For a standard polynomial family P and $t \in \mathbb{N}$, we say two polynomial families, $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t} = \{\bar{p}^t_{ij} \colon 1 \le i \le \bar{r}, 1 \le j \le l\}$ and $P_{\hat{\mathbf{v}}_t,\hat{\mathbf{w}}_t} = \{\hat{p}^t_{ij} \colon 1 \le i \le \hat{r}, 1 \le j \le l\}$ are **equivalent** if $\bar{r} = \hat{r}$, and for each i, j, there exists a polynomial $p^*_{ij}(u, \mathbf{v}_t, \mathbf{w}_t) \colon \mathbb{Z}^{(2t+1)d} \to \mathbb{Z}$, such that $\bar{p}^t_{ij}(u) = p^*_{ij}(u, \bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t)$ and $\hat{p}^t_{ij}(u) = p^*_{ij}(u, \hat{\mathbf{v}}_t, \hat{\mathbf{w}}_t)$.

Lemma 4.3. *Let P be a standard family with $\deg(P) \ge 2$. For each step t in the inductive process, there exists $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2td}$ such that $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$ is standard, and for almost all choices of $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}$ is standard and equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$.*

Proof. We proceed by induction on t. The case for $t = 0$ is trivial.

Suppose for some $t \in \mathbb{N}$, there exists $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2td}$ such that $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t} = \{p^*_{ij}(u, \bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t)\}$ is standard with $\deg(P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}) \ge 2$, and for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t} = \{p^*_{ij}(u, \mathbf{v}_t, \mathbf{w}_t)\}$ is standard and is equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$.

Given any polynomial $p(u, \mathbf{v}_t, \mathbf{w}_t)$, $\deg(p(u, \mathbf{v}_t, \mathbf{w}_t))$ is constant for almost all $(\mathbf{v}_t, \mathbf{w}_t)$ (as viewed as polynomials of u). We know that the union of finitely many sets of zero density with respect to any Følner sequence also has zero density. Hence for almost all $(\mathbf{v}_t, \mathbf{w}_t)$, both $\deg(p_{ij}(u, \mathbf{v}_t, \mathbf{w}_t))$ and $\deg(p_{i_1j_1}(u, \mathbf{v}_t, \mathbf{w}_t) - p_{i_2j_2}(u, \mathbf{v}_t, \mathbf{w}_t))$ are constant. Since $p^t_{ij}(u, \mathbf{v}_t, \mathbf{w}_t)$ have the same degree as $p^t_{ij}(u, \hat{\mathbf{v}}_t, \hat{\mathbf{w}}_t)$ and $p^t_{i_1j_1}(u, \mathbf{v}_t, \mathbf{w}_t)$ is equivalent to $p^t_{i_2j_2}(u, \mathbf{v}_t, \mathbf{w}_t)$ if and only if $p^t_{i_1j_1}(u, \hat{\mathbf{v}}_t, \hat{\mathbf{w}}_t)$ is equivalent to $p^t_{i_2j_2}(u, \hat{\mathbf{v}}_t, \hat{\mathbf{w}}_t)$ for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, our choice of i_0 is the same for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$.

Choose $(\hat{\mathbf{v}}_t, \hat{\mathbf{w}}_t)$ to be any such vector satisfying the above with $P_{\hat{\mathbf{v}}_{t+1},\hat{\mathbf{w}}_{t+1}}$ standard for some $(\hat{v}_{t+1}, \hat{w}_{t+1}) \in \mathbb{Z}^{2d}$.

Since our choice of i_0 is the same for almost all polynomial families, we get that $P_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}}$ is a standard family equivalent to $P_{\hat{\mathbf{v}}_{t+1},\hat{\mathbf{w}}_{t+1}}$ for almost all $(\mathbf{v}_{t+1}, \mathbf{w}_{t+1}) \in \mathbb{Z}^{2d(t+1)}$. ☐

4.3. **Reduction of Column Degree.**

Proposition 4.4. *Given any standard polynomial family P with $\deg(P) \ge 2$, there exists a step $t \in \mathbb{N}$ and $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2td}$ such that $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$ is a standard polynomial family of linear terms, and for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}$ is a standard linear family and equivalent to $P^t_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$.*

By applying Proposition 3.1, we get the following corollary.

Corollary 4.5. *Given any totally ergodic system, Z_{r_t} is characteristic for the averages of all such $P_{\mathbf{v}_t,\mathbf{w}_t}$, where r_t corresponds to number of rows in $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$.*

Proof. We proceed by induction on the column degrees of our families. Given any standard polynomial family P with $\deg(P) \geq 2$, it suffices to show that there exists a step t and $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2td}$ with $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$ standard, such that for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}$ is standard and equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$, and $C(P_{\mathbf{v}_t,\mathbf{w}_t}) < C(P)$. To show after finitely many steps the column degree indeed reduces, we employ a second induction on the weight of column j^*, the minimal j such that

$$\deg(p_{ij}) = \deg(P)$$

for some i.

When $\deg(p_{1j^*}) < \deg(P)$, our choice of i_0 is such that for almost all $(v_1, w_1) \in \mathbb{Z}^{2d}$, P_{v_1,w_1} is standard and the polynomial in the first row of column j^* is of maximal degree. When $\deg(p_{1j^*}) = \deg(P)$, for each $(v_1, w_1) \in \mathbb{Z}^{2d}$, the polynomial in the first row of column j^* is of maximal degree unless $p_{i_0 j^*}$ is equivalent to p_{1j^*}. In this case $C(P_{v_1,w_1}) < C(P)$ and we are done. As such, we may assume that $\deg(p_{1j^*}) = \deg(P)$.

Case 1: Suppose $r = 1$. Clearly $w_j(P_{\mathbf{v}_1,\mathbf{w}_1}) < w_j(P)$ for all $(\mathbf{v}_1, \mathbf{w}_1) \in \mathbb{Z}^{2d}$ and $j = 1, \ldots, l$ as $\deg(p_{1j}(u + v_1^{1j}) - p_{1j}(u + w_1)) < \deg(p_{1j})$ for all $p_{1j} \neq 0$.

Case 2: Suppose $r \geq 2$ and the set $H_1(P) = \emptyset$. This means that all polynomials in column j^* are nonzero. Then for all $(v_1, w_1) \in \mathbb{Z}^{2d}$, the equivalence classes and their degrees in column j^* of $P_{\mathbf{v}_1,\mathbf{w}_1}$ remain the same as in $\bar{P}_{\mathbf{v}_1,\mathbf{w}_1}$ except the classes in column j^* of $\bar{P}_{\mathbf{v}_1,\mathbf{w}_1}$ with identical degree as q_{v_1,w_1,s,j^*}. Each equivalence class not equivalent to q_{v_1,w_1,s,j^*} becomes a new equivalence class of the same degree, while the equivalence class equivalent to q_{v_1,w_1,s,j^*} splits into new classes of lower degree. Thus $w_{j^*}(P_{\mathbf{v}_1,\mathbf{w}_1}) < w_{j^*}(P)$ for all $(\mathbf{v}_1, \mathbf{w}_1) \in \mathbb{Z}^{2d}$.

Case 3: Suppose $r \geq 2$ and $H_1(P) \neq \emptyset$. Here, at least one polynomial in column j^* is identically zero. We show that after finitely many steps, the weight of column j^* is reduced. To do this we show that there exists a step t and $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2td}$ with $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$ standard, such that for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}$ is standard and equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$, $H_1(P_{\mathbf{v}_t,\mathbf{w}_t}) = \emptyset$, and $w_{\alpha_1}(P_{\mathbf{v}_t,\mathbf{w}_t}) \leq w_{\alpha_1}(P)$.

Since $H_1(P) \neq \emptyset$, we have $H_{j_0} = \emptyset$ but $H_{j_0 - 1} \neq \emptyset$.

For each polynomial family $P_{\mathbf{v}_t,\mathbf{w}_t}$ and integer $x = 1, \ldots, l-1$, we define the sub-polynomial family

$$P^x_{\mathbf{v}_t,\mathbf{w}_t} = \{p_{ij} \colon i \in H_x(P_{\mathbf{v}_t,\mathbf{w}_t}), x < \alpha_j \leq l\}.$$

We note that the entries in the first column in $P^{j_0-1}_{\mathbf{v}_t,\mathbf{w}_t}$ are the entries of column α_{j_0} of $P_{\mathbf{v}_t,\mathbf{w}_t}$ from rows whose polynomials are all identically zero in columns $\alpha_1, \ldots, \alpha_{j_0-1}$.

Claim: For every polynomial family P with $\deg(P) \geq 2$ and $j_0 \neq 1$, there exists a step t such that for almost all choices of $(\mathbf{v}_t, \mathbf{w}_t)$, $P^{j_0-1}_{\mathbf{v}_t,\mathbf{w}_t}$ is a linear family.

To prove the claim, we induct on j_0. Suppose $j_0 = l$. Thus, P^{j_0-1} has only one column. Then, by our choice of i_0, for all $(v_1, w_1) \in \mathbb{Z}^{2d}$, we have $w_1(P^1_{v_1,w_1}) < w_1(P^1)$. This is because the equivalence classes and their degrees of $P^1_{\mathbf{v}_1,\mathbf{w}_1}$ remain the same as in $\bar{P}^1_{\mathbf{v}_1,\mathbf{w}_1}$ except the classes with identical degree as q_{v_1,w_1,s,α_l}. Each equivalence class not equivalent to q_{v_1,w_1,s,α_l} become a new equivalence class of the same degree, while the equivalence class equivalent to q_{v_1,w_1,s,α_l} splits into new classes of lower degree.

Since row i_0 has zeros in each column $\alpha_1, \ldots, \alpha_{j_0-1}$, for each $(v_1, w_1) \in \mathbb{Z}^{2d}$, the number and degrees of each equivalence class in columns $\alpha_1, \ldots, \alpha_{j_0-1}$ is the same in P_{v_1,w_1} as in P. Thus for almost all $(v_1, w_1) \in \mathbb{Z}^{2d}$, j^* is still the minimal j such that $\deg(p_{ij}) = \deg(P)$ for some i, and $w_{\alpha_j}(P) = w_{\alpha_j}(P_{v_1,w_1})$, for all $j = \alpha_1, \ldots, \alpha_{j_0-1}$. Thus, we may continue our inductive process until we have $P^{j_0-1}_{\mathbf{v}_t,\mathbf{w}_t}$ is linear.

Suppose our claim holds for all polynomial families with $\bar{j} < j_0 \leq l$. To complete the induction, we will show the claim holds for all polynomial families with $j_0 = \bar{j}$. To do this, we employ an induction on the weight of column j_0. As in the base case, since $H_{j_0} = \emptyset$ but $H_{j_0-1} \neq \emptyset$, we have $w_1(P^{j_0}_{v_1,w_1}) < w_1(P^{j_0})$. When $H_{j_0}(P_{v_1,w_1}) = \emptyset$, similarly we can continue to reduce the weight of column 1 in $P^{j_0}_{v_1,w_1}$.

However, if some new zero appears in column j_0, and $H_{j_0}(P_{v_1,w_1}) \neq \emptyset$, we must first reduce the remaining polynomials in all rows $i \in H_{j_0}(P_{v_1,w_1})$ to zero. Thus, we consider the subfamily $P^{j_0}_{v_1,w_1}$. By the induction hypothesis, there exists some step t such that $P_{\mathbf{v}_t,\mathbf{w}_t}$ has all linear terms in subfamily $P^{j_0}_{\mathbf{v}_t,\mathbf{w}_t}$. Hence for each $(v_{t+1}, w_{t+1}) \in \mathbb{Z}$, $H_{j_0}(P_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}}) = \emptyset$. Thus as previously, we then may reduce the weight of column α_{j_0} further.

Hence after finitely many steps t', we will get that every entry in column 1 of $P^{j_0-1}_{\mathbf{v}_{t'},\mathbf{w}_{t'}}$ is identically zero. Thus $H_{j_0} = H_{j_0+1} \neq \emptyset$. Thus by the induction hypothesis, we may find a step t such that $P^{j_0}_{\mathbf{v}_t,\mathbf{w}_t}$ is linear. Since, every entry in column 1 of $P^{j_0-1}_{\mathbf{v}_{t'},\mathbf{w}_{t'}}$ is identically zero, $P^{j_0-1}_{\mathbf{v}_t,\mathbf{w}_t}$ is also linear, thus proving the claim.

Thus, given polynomial family P, we may find a step t such that $P^{j_0-1}_{\mathbf{v}_t,\mathbf{w}_t}$ is linear. Thus, $P^{j_0-1}_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}}$ has no nonzero polynomials. Thus, in each row of $P_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}}$ with zeros in columns, $\alpha_1,\dots,\alpha_{j_0-1}$, it also has zeros in columns $\alpha_{j_0}\dots,\alpha_l$. Thus, $H_{j_0-1}(P_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}})=\emptyset$.

We now may repeat the above process until $H_1(P_{\mathbf{v}_t,\mathbf{w}_t})=\emptyset$ for some t. Thus by the previous cases, we have $w_{j^*}(P_{\mathbf{v}_{t+1},\mathbf{w}_{t+1}})<w_{j^*}(P)$.

Thus, by repeating either step 1,2, or 3, we may show that there is a step t with $C(P_{\mathbf{v}_t,\mathbf{w}_t})<C(P)$. By Lemma 4.3 this step t is the same for almost all choices of $(\mathbf{v}_t,\mathbf{w}_t)\in\mathbb{Z}^{2td}$. At this point column j^* may no longer contain a polynomial of maximal degree. We, then now choose a new j^*, which again by Lemma 4.3 will be the same for almost all choices of $(\mathbf{v}_t,\mathbf{w}_t)\in\mathbb{Z}^{2td}$. We then may repeat this process to further reduce the column degree until we reach the desired step where almost all such families are linear families.

□

4.4. **Reduction to Linear Case.**

Proof of Proposition 4.1. We now reduce to the linear case. For polynomial families of degree 1, the result is given by Proposition 3.1. Let $P=\{p_{ij}\colon 1\le i\le r, 1\le j\le l\}$ be a standard polynomial family of degree at least 2 with column degree $C(P)$. Now let $f_1,\dots,f_r\in L^\infty(\mu)$ and let $\{\Phi_N\}_{N=1}^\infty$ be a Følner sequence in $\mathbb{Z}^d$. Without loss of generality we may assume that $\|f_i\|_{L^\infty(\mu)}\le 1$ for all $1\le i\le r$. Thus by Lemma 3.2 and Hölder's inequality we have for any finite set $F\subset\mathbb{Z}^d$,

$$\begin{aligned}
(2)\quad & \limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}(\prod_{i=1}^{r}T_1^{p_{i1}(u)}\dots T_l^{p_{il}(u)}f_i)\Big\|_{L^2(\mu)}\\
&\le \limsup_{N\to\infty}\frac{1}{|F|^2}\sum_{v_1,w_1\in F}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X\prod_{i=1}^{r}T_1^{p_{i1}(u+v_1)}\dots\\
&\qquad T_l^{p_{il}(u+v_1)}f_i\cdot\prod_{i=1}^{r}T_1^{p_{i1}(u+w_1)}\dots T_l^{p_{il}(u+w_1)}f_i\,d\mu\\
&\le \limsup_{N\to\infty}\frac{1}{|F|^2}\sum_{v_1,w_1\in F}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X\prod_{h=1}^{s}T_1^{q_{v_1,w_1,h,1}(u)}\\
&\qquad\dots T_l^{q_{v_1,w_1,h,l}(u)}b_{v_1,w_1,h}\,d\mu\\
&\le\frac{1}{|F|^2}\sum_{v_1,w_1\in F}\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\prod_{h=1}^{s-1}T_1^{(q_{v_1,w_1,h,1}-q_{v_1,w_1,s,1})(u)}\\
&\qquad\dots T_l^{(q_{v_1,w_1,h,l}-q_{v_1,w_1,s,l})(u)}b_{v_1,w_1,h}\Big\|_{L^2(\mu)}
\end{aligned}$$

for $(v_1, w_1) \in \mathbb{Z}^{2d}$, where the $b_{v_1,w_1,h}$ represent any of the following bounded functions:

- $T_1^{p_{i1}(v_1 - v_1^{i1})} \dots T_l^{p_{il}(v_1 - v_1^{il})} f_i$ for $i \in I_2$
- $f_i \cdot T_1^{p_{i1}(v_1) - p_{i1}(w_1)} \dots T_l^{p_{il}(v_1) - p_{il}(w_1)} f_i$ for $i \in I_1$

Since P is a standard polynomial family of degree at least 2, $1 \in I_2$ and we have that $b_{v_1,w_1,1} = T_1^{a_1} \dots T_l^{a_l} f_1$ for some $a_1, \dots, a_l \in \mathbb{Z}$. Thus for all k and all $(v_1, w_1) \in \mathbb{Z}^{2d}$,

$$|\!|\!| b_{v_1,w_1,1} |\!|\!|_k = |\!|\!| f_1 |\!|\!|_k.$$

By Lemma 4.3, at each step t there exists $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2d}$ such that for almost all choices of $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}$ is standard and equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}$.

Let G_t be the good set of $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$ satisfying the above, and B_t be the complement. For each $(\mathbf{v}_t, \mathbf{w}_t) \in G_t$, when $P_{\mathbf{v}_t,\mathbf{w}_t} = \{p_{ij}^t : 1 \le i \le r_t, 1 \le j \le l\}$ has degree at least 2, we repeat the argument in (2) to get that for any finite $F \subset \mathbb{Z}^d$,

$$\begin{aligned}
&\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \big(\prod_{i=1}^{r} T_1^{p_{i1}(u)} \dots T_l^{p_{il}(u)} f_i\big) \Big\|_{L^2(\mu)} \\
&\le \frac{1}{|F|^{2t}} \sum_{(\mathbf{v}_t,\mathbf{w}_t)\in F^{2t}\cap G_t} \limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \big(\prod_{i=1}^{r_t} T_1^{p_{i1}^t(u)} \dots T_l^{p_{il}^t(u)} f_i^t\big) \Big\|_{L^2(\mu)} \\
&\qquad + \frac{1}{|F|^{2t}} \sum_{(\mathbf{v}_t,\mathbf{w}_t)\in F^{2t}\cap B_t} 1
\end{aligned}$$

where $f_1^t = T_1^{a_1} \dots T_l^{a_l} f_1$ for some $a_1, \dots, a_l \in \mathbb{Z}$ and all others are functions bounded by 1. This follows because in each step t in the induction process,

$$\limsup_{N\to\infty} \Big\| \frac{1}{|\Phi_N|} \sum_{u\in\Phi_N} \big(\prod_{i=1}^{r_t} T_1^{p_{i1}^t(u)} \dots T_l^{p_{il}^t(u)} f_i^a\big) \Big\|_{L^2(\mu)}$$

is bounded by 1.

By Proposition 4.4, there exists a step $t \in \mathbb{N}$ and $(\bar{\mathbf{v}}_t, \bar{\mathbf{w}}_t) \in \mathbb{Z}^{2td}$ such that for almost all $(\mathbf{v}_t, \mathbf{w}_t) \in \mathbb{Z}^{2td}$, $P_{\mathbf{v}_t,\mathbf{w}_t}^t$ is a standard polynomial family of linear terms and equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}^t$.

By Corollary 4.5, when $P_{\mathbf{v}_t,\mathbf{w}_t}^t$ is equivalent to $P_{\bar{\mathbf{v}}_t,\bar{\mathbf{w}}_t}^t$, Z_{r_t} is characteristic for the averages using $P_{\mathbf{v}_t,\mathbf{w}_t}^t$. In other words, if we assume $|\!|\!| f |\!|\!|_{r_t+1} = 0$

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}(\prod_{i=1}^{r_t}T_1^{p^t_{i1}(u)}\dots T_l^{p^t_{il}(u)}f^t_i)\Big\|_{L^2(\mu)}=0.$$

As Z_{r_t} is characteristic for the averages of $P^t_{\mathbf{v}_t,\mathbf{w}_t}$ for all $(\mathbf{v}_t,\mathbf{w}_t)\in G_t$ and the exceptional set B_t has zero density with respect to any Følner sequence in $\mathbb{Z}^{2td}$, we get that

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}(\prod_{i=1}^{r}T_1^{p_{i1}(u)}\dots T_l^{p_{il}(u)}f_i)\Big\|_{L^2(\mu)}$$
$$\le\inf_F\Big(\frac{1}{|F|^{2t}}\sum_{(\mathbf{v_t},\mathbf{w_t})\in F^{2t}\cap G_t}\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\prod_{i=1}^{r_t}T_1^{p^t_{i1}(u)}\dots T_l^{p^t_{il}(u)}f^t_i\Big\|_{L^2(\mu)}$$
$$+\frac{1}{|F|^{2t}}\sum_{(\mathbf{v_t},\mathbf{w_t})\in F^{2t}\cap B_t}1\Big)=0.$$

□

4.5. **Reduction to the standard case.**

Proof of Proposition 2.9. We now reduce the general case to one involving standard systems. Let $P=\{p_{ij}\colon 1\le i\le r, 1\le j\le l\}$ be a (nonstandard) ED-set of polynomials of degree less than b, let $f_1,\dots,f_r\in L^\infty(\mu)$, and let $\{\Phi_N\}_{N=1}^\infty$ be a Følner sequence in $\mathbb{Z}^d$. By Lemma 3.2, there exists a Følner sequence $\{\Theta_N\}_{N=1}^\infty$ in $\mathbb{Z}^{3d}$ such that

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}(\prod_{i=1}^{r}T_1^{p_{i1}(u)}\dots T_l^{p_{il}(u)}f_i)\Big\|_{L^2(\mu)}$$
$$\le\limsup_{M\to\infty}\frac{1}{\Theta_M}\sum_{(u,v,w)\in\Theta_M}\int_X\prod_{i=1}^{r}T_1^{p_{i1}(u+v)+q(u)}$$
$$\dots T_l^{p_{il}(u+v)}f_i\prod_{i=1}^{r}T_1^{p_{i1}(u+w)+q(u)}\dots T_l^{p_{il}(u+w)}f_i d\mu$$
$$\le\limsup_{M\to\infty}\Big\|\frac{1}{\Theta_M}\sum_{(u,v,w)\in\Theta_M}\prod_{i=1}^{r}T_1^{p_{i1}(u+v^{i1})+q(u)}\dots T_l^{p_{il}(u+v^{il})}(T_1^{p_{i1}(v-v^{i1})}$$
$$\dots T_l^{p_{il}(v-v^{il})}f_i)\prod_{i=1}^{r}T_1^{p_{i1}(u+w)+q(u)}\dots T_l^{p_{il}(u+w)}f_i\Big\|_{L^2(\mu)}$$

where $q : \mathbb{Z}^{3d} \to \mathbb{Z}$ is any polynomial of degree b. We note the set $\{p_{i1}(u+v^{i1})+q(u), p_{i1}(u+w)+q(u), p_{ij}(u+v^{ij}), p_{ij}(u+w)\colon 1 \le i \le r, 2 \le j \le l\}$ of polynomials $\mathbb{Z}^{3d} \to \mathbb{Z}$ is a standard family of degree b, thus there exists $k \in \mathbb{N}$ (that depends only on the original polynomial family P) such that

$$\limsup_{M\to\infty} \Big\| \frac{1}{\Theta_M} \sum_{(u,v,w)\in\Theta_M} \prod_{i=1}^{r} T_1^{p_{i1}(u+v^{i1})+q(u)} \dots T_l^{p_{il}(u+v^{il})} (T_1^{p_{i1}(v-v^{i1})} \dots T_l^{p_{il}(v-v^{il})} f_i) \prod_{i=1}^{r} T_1^{p_{i1}(u+w)+q(u)} \dots T_l^{p_{il}(u+w)} f_i \Big\|_{L^2(\mu)} = 0.$$

□

REFERENCES

[Be] Bergelson, V. Weakly mixing PET. *Erg. Th. and Dyn. Sys.* **7** (1987), 337-349.

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

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

[CL] Conze, J.P., Lesigne, E. Théoré mes ergodiques pour des mesures diagonales. *Bull. Soc. Math France.* **112** (1984), 143-175.

[FrK] Frantzikinakis, N., Kra, B. Convergence of multiple ergodic averages for some commuting transformations. *Erg. Th. and Dyn. Sys.,* **25** (2005) 799-809.

[Fu1] Furstenberg, H. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* **31** (1977), 204-256.

[Fu2] Furstenberg, H. Recurrence in ergodic theory and combinatorial number theory. *Princeton University Press, Princeton, NJ.*, (1981).

[HK1] Host, B., Kra, B. Nonconventional ergodic averages and nilmanifolds. *Annals of Math,* **161** (2005), 397-488.

[HK2] Host, B., Kra, B. Convergence of polynomial ergodic averages. *Isr. J. Math,* **149** (2005) 1-19.

[Le1] Leibman, A. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translation on a nilmanifold. *Erg. Th. and Dyn. Sys.* **25** (2005), no. 1, 215-225.

[Le2] Leibman, A. Convergence of multiple ergodic averages along polynomials of several variables, *Isr. J. Math,* **146** (2005), 303-315.

[Ta] Tao, T. Norm convergence of multiple ergodic averages for commuting transformaions, *Erg. Th. and Dyn. Sys.* **10** (2008), 657-688.