MULTIPLE RECURRENCE AND CONVERGENCE RESULTS ASSOCIATED TO $\mathbb{F}_p^\omega$-ACTIONS

By

VITALY BERGELSON,[1] TERENCE TAO,[2] AND TAMAR ZIEGLER[3]

Abstract. Using an ergodic inverse theorem obtained in our previous paper, we obtain limit formulae for multiple ergodic averages associated with the action of $\mathbb{F}_p^\omega = \oplus \mathbb{F}_p$. From this we deduce multiple Khintchine-type recurrence results analogous to those for $\mathbb{Z}$-systems obtained by Bergelson, Host, and Kra, and also present some new counterexamples in this setting.

1 Introduction

The celebrated theorem of Szemerédi [29] stating that any set of positive density in $\mathbb{Z}$ contains arbitrary long progressions has a natural analogue for "large" sets in the group $\mathbb{F}_p^\omega = \oplus \mathbb{F}_p$, the direct sum of countably many copies of a finite field of prime order p. While the content of Szemerédi's theorem can be succinctly expressed by the maxim "large sets in $\mathbb{Z}$ are AP-rich" (where AP stands for Arithmetic Progression), the $\mathbb{F}_p^\omega$ analogue states that any "large" set in $\mathbb{F}_p^\omega$ is AS-rich, that is, it contains arbitrarily large Affine Subspaces. This analogy extends to the similarity between various proofs of these two theorems and is especially interesting when one studies the $\mathbb{F}_p^\omega$ analogues of various aspects of the ergodic approach to Szemerédi's theorem introduced by Furstenberg in [13].

Given an invertible probability measure preserving system $(X, \mathcal{X}, \mu, T)$, a set $A \in \mathcal{X}$ with $\mu(A) > 0$, and an integer $k \in \mathbb{N}$, let $\phi(n) = \mu(A \cap T^n A \cap \cdots \cap T^{kn} A)$. The sequence $\phi(n)$ can be viewed as a generalized positive definite sequence. The analysis of the properties and the asymptotic behavior of $\phi(n)$ leads to the proof and enhancements of Szemerédi's theorem. Not surprisingly, the study of the $\mathbb{F}_p^\omega$ analogue of $\phi(n)$ leads to a better understanding and enhancement of the $\mathbb{F}_p^\omega$

[1]The first author is supported by support NSF grant DMS-1162073.
[2]The second author is supported by NSF grant DMS-0649473 and by a Simons Investigator Award.
[3]The third author is supported by ISF grant 407/12.

JOURNAL D'ANALYSE MATHÉMATIQUE, Vol. 127 (2015)
DOI 10.1007/s11854-015-0033-1

analogue of Szemerédi's theorem. It also throws new light on the various related facts belonging to the realm of ergodic theory.

In this paper, we describe (in Theorem 1.6 below) the characteristic factor for certain multiple ergodic averages on measure-preserving systems in the case where the underlying group G is an infinite-dimensional vector space $\mathbb{F}_p^\omega$ over a finite field; this is the analogue of the well-known description in [23], [40] of characteristic factors for multiple ergodic averages of $\mathbb{Z}$-actions. Using this description, we obtain explicit formulae for the limit of such multiple ergodic averages. As an application of these formulae, we establish multiple recurrence theorems of Khintchine type in some cases, and exhibit counterexamples to such theorems in other cases.

The detailed statements of the main results of our paper are formulated at the end of the introduction.

1.1 Convergence of multiple ergodic averages and limit formulae.

Before we can properly state our main results, we need to set up a certain amount of notation regarding measure-preserving G-systems and their characteristic factors.

Let $G = (G, +)$ be a countable abelian group, and let $(X, \mathcal{X}, \mu)$ be a probability space, which we always assume to be *separable*[1] in the sense that the σ-algebra $\mathcal{X}$ is countably generated modulo μ-null sets; in most applications, one can reduce to this situation without difficulty, so this is not a serious restriction in practice. An **invertible measure-preserving transformation** on X is an invertible map $T : X \to X$ with T and T^{-1} both measurable such that $\mu(T_g(E)) = \mu(E)$ for all $E \in \mathcal{X}$. A **measure-preserving G-action** on X is a family $(T_g)_{g\in G}$ of invertible measure-preserving transformations $T_g : X \to X$, such that $T_g T_h = T_{g+h}$ and $T_0 = \mathrm{id}$ μ-almost everywhere for all $g, h \in G$. We refer to the quadruplet $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ as a **measure-preserving G-system**, or **G-system** for short. We abbreviate the (complex-valued) Lebesgue spaces $L^p(X, \mathcal{X}, \mu)$ for $1 \le p \le \infty$ as $L^p(\mathrm{X})$. We adopt the usual convention of identifying two functions in $L^p(\mathrm{X})$ if they agree μ-almost everywhere; in particular, this makes $L^2(\mathrm{X})$ a separable Hilbert space.

A **Følner sequence** in G is a sequence $(\Phi_n)_{n=1}^\infty$ of finite nonempty subsets of G such that

$$\lim_{n\to\infty} \frac{|(g + \Phi_n)\Delta\Phi_n|}{|\Phi_n|} = 0$$

[1]This assumption is not used explicitly in this paper but is used in the paper [6] on whose results we rely, in order to perform certain measurable selections (see Appendix C of that paper) as well as disintegrations of measures.

for all $g \in G$, where $|H|$ denotes the cardinality of a finite set H. Note that we do not require the Φ_n to be nested or to exhaust all of G. The class of Følner sequences is translation-invariant in the sense that if $(\Phi_n)_{n=1}^\infty$ is a Følner sequence, then $(g_n + \Phi_n)_{n=1}^\infty$ is also a Følner sequence for any $g_1, g_2, \ldots \in G$. It is a classical fact that every countable abelian group is amenable [34] and hence has at least one Følner sequence [12]; for instance, if $G = \mathbb{Z}$, one can take $\Phi_n := \{1, \ldots, n\}$.

The classical Mean Ergodic Theorem[2] asserts, among other things, that if $G = (G, +)$ is a countable abelian group with Følner sequence $(\Phi_n)_{n=1}^\infty$, $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ is a G-system and $f \in L^2(\mathrm{X})$, then the limit

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} T_g f \tag{1.1}$$

converges strongly in $L^2(\mathrm{X})$ norm, where we use the averaging notation $\mathbb{E}_{h\in H} := \frac{1}{|H|}\sum_{h\in H}$ for any nonempty finite set H and also write $T_g f$ for $f \circ T_g$. Since strong convergence in $L^2(\mathrm{X})$ implies weak convergence, we obtain as a corollary that the limit

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0 T_g f_1 \, d\mu \tag{1.2}$$

exists for all $f_0, f_1 \in L^2(\mathrm{X})$.

The Mean Ergodic Theorem not only gives existence of these limits, but provides a formula for the value of these limits. To describe this formula, we need some more notation. Define a **factor** $(Y, \mathcal{Y}, \nu, (S_g)_{g\in G}, \pi) = (\mathrm{Y}, \pi)$ of a G-system $(X, \mathcal{X}, \mu, (T_g)_{g\in G})$ to be another G-system $\mathrm{Y} = (Y, \mathcal{Y}, \nu, (S_g)_{g\in G})$, together with a measurable map $\pi : X \to Y$ which respects the measure in the sense that $\mu(\pi^{-1}(E)) = \nu(E)$ for all $E \in \mathcal{Y}$ (or equivalently, $\pi_* \mu = \nu$), and also respects the G-action in the sense that $S_g \circ \pi = \pi \circ T_g$ μ-a.e. for all $g \in G$. For instance, if $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{X}$ which is invariant with respect to the G-action $(T_g)_{g\in G}$, then $(X, \mathcal{B}, \mu \downharpoonright_{\mathcal{B}}, (T_g)_{g\in G}, \mathrm{id})$ is a factor of $(X, \mathcal{X}, \mu, (T_g)_{g\in G})$, where $\mu \downharpoonright_{\mathcal{B}}$ is the restriction of $\mathcal{X}$ to $\mathcal{B}$. By abuse of notation, we refer to an invariant sub-σ-algebra $\mathcal{B}$ as a factor of $(X, \mathcal{X}, \mu, (T_g)_{g\in G})$. We call two factors $(Y, \mathcal{Y}, \nu, (S_g)_{g\in G}, \pi)$, $(Y', \mathcal{Y}', \nu', (S'_g)_{g\in G}, \pi')$ **equivalent** if the sub-σ-algebras

$$\{\pi^{-1}(E) : E \in \mathcal{Y}\}, \ \{(\pi')^{-1}(E) : E \in \mathcal{Y}'\}$$

of $\mathcal{X}$ that they generate agree modulo null sets. It is clear that every factor is equivalent to a unique invariant (modulo null sets) sub-σ-algebra of $\mathcal{X}$, so one may think of factors as invariant sub-σ-algebras if it is convenient to do so.

[2]This theorem is usually stated in textbooks for $\mathbb{Z}$-systems, but the proof extends without difficulty to actions by other amenable groups. See, for example, [3, Theorem 6.4.15].

For a factor $(Y, \mathcal{Y}, \nu, (S_g)_{g\in G}, \pi) = (\mathrm{Y}, \pi)$, we define the **pullback map** $\pi^* : L^2(\mathrm{Y}) \to L^2(\mathrm{X})$ by $\pi^* f := f \circ \pi$. We define the **pushforward map** $\pi_* : L^2(\mathrm{X}) \to L^2(\mathrm{Y})$ to be the adjoint of this map. In the case when the factor arises from an invariant sub-σ-algebra $\mathcal{B}$ of $\mathcal{X}$, the pushforward $\pi_* f$ is the same as the conditional expectation $\mathbb{E}(f|\mathcal{B})$ of f with respect to $\mathcal{B}$.

Given a G-system $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$, we define the **invariant factor**

$$(\mathrm{Z}_0, \pi_0) = (\mathrm{Z}_0(\mathrm{X}), \pi_0) = (Z_0, \mathcal{Z}_0, \mu_0, (T_g)_{g\in G}, \pi_0)$$

of X to be (up to equivalence[3]) the factor associated to the invariant σ-algebra $\mathcal{X}^T := \{E \in \mathcal{X} : T_g E = E \text{ for all } g \in G\}$. This factor is a *characteristic factor* for the averages (1.1), (1.2), in the sense that the limit in (1.1) converges strongly in $L^2(\mathrm{X})$ to 0 whenever $(\pi_0)_* f$ vanishes, and similarly the limit in (1.2) converges to 0 when either $(\pi_0)_* f_0$ or $(\pi_0)_* f_1$ vanishes (see [16]). As a consequence, to compute the limits in (1.1), one may freely replace f by $(\pi_0)_* f$ (and descend from X to the factor Z_0), and similarly for (1.2). On the characteristic factor Z_0, the action of G is essentially trivial; and as a conclusion one obtains the well-known limit formulae

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} T_g f = (\pi_0)^* (\pi_0)_* f$$

and

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0 T_g f_1 \, d\mu = \int_{Z_0} ((\pi_0)_* f_0) ((\pi_0)_* f_1) \, d(\pi_0)_* \mu.$$

The situation is particularly simple when the G-system X is **ergodic**, which means that the invariant σ-algebra $\mathcal{X}^T$ consists only of sets of full measure or zero measure, or equivalently that the invariant factor Z_0 is a point. In this case, $(\pi_0)_* f = \int_X f d\mu$; and so

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} T_g f = \int_X f d\mu$$

and

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0 T_g f_1 \, d\mu = \left(\int_X f_0 d\mu\right) \left(\int_X f_1 d\mu\right).$$

[3] One can, of course, define (Z_0, π_0) canonically actually to *be* the factor $(X, \mathcal{X}^T, \mu \restriction_{\mathcal{X}^T}, (T_g)_{g\in G}, \mathrm{id})$, but it can be convenient to allow Z_0 only to be defined up to equivalence, in order to take advantage of other models of the invariant factor which may be more convenient to compute with.

This concludes our discussion of the classical ergodic averages. We now consider the more general **multiple ergodic averages**

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n}(T_{c_1 g} f_1)(T_{c_2 g} f_2)\dots(T_{c_k g} f_k) \tag{1.3}$$

and

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n}\int_X (T_{c_0 g} f_0)\dots(T_{c_k g} f_k)\,d\mu \tag{1.4}$$

associated to a G-system $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$, where $k \ge 1$ and $c_0,\dots,c_k$ are integers, and (to avoid absolute integrability issues) $f_0,\dots,f_k$ are now assumed to lie in $L^\infty(\mathrm{X})$ rather than $L^2(\mathrm{X})$. Note that in (1.4) we may collect terms if necessary and reduce to the case when the $c_0,\dots,c_k$ are distinct. Similarly, in (1.3) we may reduce to the case when the $c_1,\dots,c_k$ are distinct and nonzero (since zero coefficients can simply be factored out). The reader can keep the model case $c_i = i$ in mind for this discussion, though for technical reasons it is convenient to consider more general coefficients c_i as well.

The convergence and recurrence properties of these averages have been extensively studied in the literature, particularly in the model case $G = \mathbb{Z}$. For instance, the celebrated Furstenberg multiple recurrence theorem [13] asserts the lower bound

$$\liminf_{n\to\infty} \mathbb{E}_{g\in\Phi_n}\int_X f(T_g f)\cdots(T_{kg} f)\,d\mu > c > 0$$

in the $G = \mathbb{Z}$ case whenever $k \ge 1$ and $f \in L^\infty(\mathrm{X})$ is nonnegative and not identically 0, where c does not depend on the choice of $(\Phi_n)_{n=1}^\infty$. The same result holds for all countable abelian groups G [15]. On the other hand, the original proofs of the multiple recurrence theorem did not actually establish the existence of the limit in (1.3) or (1.4) for general $k > 1$. In the case of $\mathbb{Z}$-actions, this was first achieved for $k = 2$ in [13], for $k = 3$ in [37] (building upon a sequence of partial results in [7, 8, 9, 16]). The case $k = 4$ was established in [21] (see also [22]) and independently in [38]. The methods in [21], [38] were generalized to cover all $k \ge 1$ first in [23] and then in [40]. After the general convergence of (1.3), (1.4) for $G = \mathbb{Z}$ was established, a number of additional proofs of this result (as well as generalizations thereof) have appeared in the literature [30], [1], [33], [20], [36]. The argument in [36] is extremely general and extends to averages over arbitrary countable abelian groups G, with the shifts $g,\dots,(k-1)g$ replaced by polynomial functions of g (see also [42]).

Now we turn to the question of understanding the nature of the limit in (1.3) or (1.4) for higher values of k than $k = 1$. For simplicity, we focus on the case of ergodic G-systems $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$; the results discussed here can then be extended to the nonergodic case by ergodic decomposition (see, e.g., [35]).

The case $k = 2$ can be analysed by spectral theory. Define an **eigenfunction** of an ergodic G-system X to be a nonzero function $f \in L^2(\mathrm{X})$ such that for each $g \in G$, one has $T_g f = \lambda_g f$ for some complex number λ_g. Define the **Kronecker factor** $(\mathrm{Z}_1, \pi_1) = (\mathrm{Z}_1(\mathrm{X}), \pi_1) = (Z_1, \mathbb{Z}_1, \mu_1, (T_{1,g})_{g\in G}, \pi_1)$ of X to be the factor (up to equivalence) associated to the sub-σ-algebra of $\mathcal{X}$ generated by the eigenfunctions of X. The Kronecker factor is (up to equivalence) given by an abelian group rotation $\mathrm{Z}_1 = (U, \mathcal{U}, m_U, (S_g)_{g\in G})$, where $U = (U, +)$ is a compact abelian group with Borel σ-algebra $\mathcal{U}$ and Haar probability measure m_U, and each $S_g : U \to U$ is a group translation $S_g(x) := x + \alpha_g$, where $g \mapsto \alpha_g$ is a homomorphism from G to U (see [41] for a general form of this theorem). Furthermore, this factor is ergodic (which is equivalent to the image of the homomorphism $g \mapsto \alpha_g$ being dense in U). It is known (see, e.g., [6]) that the Kronecker factor Z_1 is characteristic for the $k = 2$ averages (1.3), (1.4), in the sense that the former average converges to 0 in $L^2(\mathrm{X})$ norm when at least one of $(\pi_1)_* f_1, (\pi_1)_* f_2$ vanishes, and the latter average converges to 0 when at least one of $(\pi_1)_* f_0, (\pi_1)_* f_1, (\pi_1)_* f_2$ vanishes. From this, one can effectively replace each function f_i by its pushforward $(\pi_1)_* f_i$ in the limits (1.3), (1.4) (replacing X with Z_1). These limits can then be evaluated by harmonic analysis on U, resulting in the limit formula

$$\text{(1.5)}\quad \lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0(T_g f_1)(T_{2g} f_2)\,d\mu = \int_U \int_U (\pi_1)_* f_0(h)(\pi_1)_* f_1(h+t)(\pi_1)_* f_2(h+2t)\,d\mu_U(h)\,d\mu_U(t)$$

for (1.4) (in the model case $c_i = i$), and hence (by duality, and existence of the limit) a similar formula for (1.3); similarly for other choices of coefficients c_i. We can rewrite this formula as

$$\text{(1.6)}\quad \lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0(T_g f_1)(T_{2g} f_2)\,d\mu = \int_{HP_{0,1,2}(U)} (\pi_1)_* f_0(h_0)(\pi_1)_* f_1(h_1)(\pi_1)_* f_2(h_2)\,dm_{HP_{0,1,2}(U)}(h_0, h_1, h_2),$$

where $HP_{0,1,2}(U) \subset U^3$ is the closed subgroup

$$HP_{0,1,2}(U) := \{(h, h+t, h+2t) : h, t \in U\}$$

of U^3 and $m_{HP_{0,1,2}(U)}$ is the Haar probability measure on $HP_{0,1,2}(U)$. (The reason for the notation $HP_{0,1,2}$ will be made clearer later.)

In the case of $\mathbb{Z}$-actions, the limit of (1.3), (1.4) for higher values of k is also understood; see [39], [40], [4]. For each value of k, a characteristic factor Z_k associated to the averages (1.3), (1.4) which (up to equivalence) is an inverse limit of nilsystems of step at most $k-1$ was constructed in [23] (see also [40]). By projecting onto each such nilsystem and using the equidistribution theory on such nilsystems (see [25], [39]), a limit formula generalizing (1.5), (1.6) (but for $\mathbb{Z}$-actions) was established; see [40]. A closely related analysis was also performed in [4], which (among other things) led to the following Khintchine-type recurrence result: if $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in\mathbb{Z}})$ is an ergodic $\mathbb{Z}$-system and $A \in \mathcal{X}$ has positive measure, then for every $\varepsilon > 0$ and all $k = 1, 2, 3$, the sets[4]

$$\{n \in \mathbb{Z} : \mu(A \cap T_{-n}A \cap \cdots \cap T_{-kn}A) \geq \mu(A)^{k+1} - \varepsilon\}$$

are syndetic. Surprisingly, this type of result fails for $k > 3$; see [4] for details.

The arguments in [4] also give a structural result for the correlation sequences[5]

$$I_{c_0,\dots,c_k;f_0,\dots,f_k}(g) := \int_X (T_{c_0g}f_0)(T_{c_1g}f_1)\dots(T_{c_kg}f_k)d\mu \tag{1.7}$$

for $k \geq 1$, $f_0, \dots, f_k \in L^\infty(\mathrm{X})$ and $g \in \mathbb{Z}$ and distinct integers $c_1, \dots, c_k$. To state these results, recall that a $(k-1)$**-step nilsequence** a uniform limit of sequences of the form $n \mapsto F(\theta^n\Gamma)$ for a $(k-1)$-step nilmanifold N/Γ, a group element $\theta \in N$, and a continuous function $F : N/\Gamma \to \mathbb{C}$; recall also that a bounded sequence $\sigma : G \to \mathbb{C}$ in a countable abelian group G is said to converge to 0 in uniform density if $\lim_{n\to\infty} \sup_{h\in G} \mathbb{E}_{g\in h+\Phi_n}|\sigma(g)| = 0$ for any Følner sequence $(\Phi_n)_{n=1}^\infty$. It was shown in [4] that the sequence $I_{c_0,\dots,c_k;f_0,\dots,f_k}$ can be decomposed as the sum of a $(k-1)$-step nilsequence and an error sequence $n \mapsto \sigma(n)$ which converges to 0 in uniform density.

1.2 New results. Having reviewed the preceding results, we now proceed to the description of new results in this paper, in which we focus on a family of countable abelian groups G at the opposite end of the spectrum to the integers $\mathbb{Z}$, namely the infinite-dimensional vector space $G := \mathbb{F}_p^\omega = \oplus\mathbb{F}_p$ over a finite field $\mathbb{F}_p$ of prime order p, with a countable basis $e_1, e_2, \dots$. This can be viewed as the direct limit[6] of the finite-dimensional subspaces $\mathbb{F}_p^n$, defined as the span of $e_1, \dots, e_n$; and a G-system can be viewed as a probability space with an infinite

[4]The negative signs here are artifacts of our sign conventions, and can be easily removed if desired.

[5]Strictly speaking, the results in [4] are only claimed in the case $f_0 = \cdots = f_k$ and $c_i = i$, but it is not difficult to see that the argument in fact applies in general.

[6]Note that this limit is distinct from the **inverse limit** $\prod_{i=1}^\infty \mathbb{F}_p$ of the $\mathbb{F}_p^n$; for instance, $\mathbb{F}_p^\omega$ is a countable vector space, whereas $\prod_{i=1}^\infty \mathbb{F}_p$ is an uncountable (but compact) group.

sequence $T_{e_n} : X \to X$ of commuting measure-preserving transformations, each of period p in the sense that $T_{e_n}^p = \mathrm{id}$. Observe that we can view these subspaces $\mathbb{F}_p^n$ as a Følner sequence for $\mathbb{F}_p^\omega$, but this is of course not the only such sequence (for instance, one can take the affine spaces $g_n + \mathbb{F}_p^n$, where $g_1, g_2, \ldots$ is an arbitrary sequence in $\mathbb{F}_p^\omega$). One can then ask for a formula for the limits in (1.3), (1.4), as well as a structure theorem for the correlation sequences (1.7) (now defined for $g \in G$ rather than $g \in \mathbb{Z}$).

To state the results, we need to introduce a variant of the concept of nilsystem that is suitable for $\mathbb{F}_p^\omega$-actions, which we refer to as a Weyl system. To define such systems, we first need the notion of a polynomial function on a G-system.

Definition 1.3 (Polynomials)**.** Let $G = (G, +)$ be a countable abelian group, let $U = (U, +)$ be an abelian group, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g \in G})$ be a measure preserving system. For any measurable function $\rho : X \to U$ and $g \in G$, let $\Delta_g \rho : X \to U$ denote the function $\Delta_g \rho(x) := \rho(T_g x) - \rho(x)$; thus Δ_g can be viewed as a difference operator on the measurable functions from X to U. If $k \geq 1$ is a natural number, we say that ρ is a **polynomial of degree less than** k if $\Delta_{g_1} \cdots \Delta_{g_k} \rho(x) = 0$ μ-almost everywhere for any $g_1, \ldots, g_k \in G$. We also adopt the convention that the zero function is the only polynomial of degree less than k if $k \leq 0$.

In a similar vein, a sequence $g : \mathbb{Z} \to U$ is said to be a **polynomial of degree less than** k if $\Delta_{h_1} \cdots \Delta_{h_k} g(n) = 0$ for all $h_1, \ldots, h_k, n \in \mathbb{Z}$, where $\Delta_h g(n) := g(n+h) - g(n)$, with the same convention as before if $k \leq 0$.

Note that a measurable function $\rho : X \to \mathbb{R}/\mathbb{Z}$ is a polynomial of degree less than 2 if and only if the function $e^{2\pi i \rho}$ is an eigenfunction of the system X. Thus we see that the polynomials of degree less than 2 are closely related to the Kronecker factor, which in turn controls the $k = 2$ averages (1.3), (1.4). More generally, we shall see (in the case $G = \mathbb{F}_p^\omega$) that the polynomials of degree less than k control the averages (1.3), (1.4). One can define polynomial maps between more general groups (not necessarily abelian); see [24]. However, we do not require this more general concept of a polynomial map here.

For future reference, we observe (by an easy induction using Pascal's triangle) that a sequence $g : \mathbb{Z} \to U$ is a polynomial of degree less than k if and only if it has a discrete Taylor expansion of the form $g(n) = \sum_{0 \leq j < k} \binom{n}{j} a_j$ for some coefficients $a_j \in U$, where $\binom{n}{j} := n(n-1)\cdots(n-j+1)/j!$. We remark that the top coefficient a_{k-1} of $g(n)$ can also be computed as $a_{k-1} = \Delta_1^{k-1} g(n)$ for any n.

Next, we recall the notion of a cocycle extension.

Definition 1.4 (Cocycle extension)**.** Let $G = (G, +)$ be a countable abelian group, let $U = (U, +)$ be a compact abelian group, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ be a measure preserving system. A (G, X, U)**-cocycle** is a measurable function $\rho : G \times X \to U$ that satisfies the **cocycle equation**

$$\rho(g + g', x) = \rho(g, T_{g'}x) + \rho(g', x) \tag{1.8}$$

for all $g, g' \in G$ and μ-almost all $x \in X$. Given such a cocycle, we define the **extension** $\mathrm{X} \times_\rho U$ of X by the cocycle ρ to be the G-system given by the product probability space $(X \times U, \mathcal{X} \times \mathcal{U}, \mu \times m_U)$, where $\mathcal{U}$ is the Borel σ-algebra on U, m_U is the Haar probability measure on U, and the action $(\tilde{T}_g)_{g\in G}$ on $X \times U$ is given by the formula $\tilde{T}_g(x, u) := (T_g x, u + \rho(g, x))$. Note that the cocycle equation (1.8) ensures that $\mathrm{X} \times_\rho U$ is indeed a G-system. If For a positive integer k, we say that the cocycle ρ is a **polynomial cocycle of degree less than** k if, for each $g \in G$, the function $x \mapsto \rho(g, x)$ is a polynomial of degree less than k.

Definition 1.5 (Weyl system)**.** Let $k \geq 0$ be an integer, and let $G = (G, +)$ be a countable abelian group. We define a k**-step Weyl** G**-system** recursively as follows:

- A 0-step Weyl G-system is a point.
- If $k \geq 1$, a k-step Weyl G-system is any system of the form $\mathrm{X} \times_{\rho_k} U_k$, where X is a Weyl G-system of order $k - 1$, U_k is a compact abelian group[7], and ρ_k is a polynomial (G, X, U_k)-cocycle of degree less than k.

We define the notion of a **continuous** k**-step Weyl system** similarly to a k-step Weyl system, except now that all the cocycles involved are also required to be continuous. (Note that a Weyl system is a Cartesian product of compact spaces and is thus also compact.)

Informally, a Weyl G-system of order k takes the form $U_1 \times_{\rho_2} U_2 \times_{\rho_3} \cdots \times_{\rho_k} U_k$ for some compact abelian groups $U_1, \ldots, U_k$ (which we refer to as the **structure groups** of the system) and polynomial cocycles $\rho_1, \ldots, \rho_k$ (the cocycle ρ_1 is essentially a homomorphism from G to U_1 and is not explicitly shown in the above notation). In the case $k = 1$, a Weyl G-system is simply a group rotation $T_g : u_0 \mapsto u_0 + \rho_1(g)$ on U_1.

Remark. In [6], we defined the notion of an Abramov $\mathbb{F}_p^\omega$-system $\mathrm{Abr}_{<k}(\mathrm{X})$. This is is a system where $P_{<k}(X)$ - the polynomials of degree less than k - span $L^2(X)$. We show that in the case where $k \leq \mathrm{char}(\mathbb{F})$, an Abramov system can be given the structure of a Weyl system.

[7] In this paper, "compact group" is understood to be short for "compact metrizable group".

Example. Let X_1 be the product space $\prod \mathbb{F}_p$ of sequences $(x_n)_{n=1}^\infty$ with $x_n \in \mathbb{F}_p$, with the product topology and the Haar probability measure. It becomes a 1-step Weyl G-system with $G := \mathbb{F}_p^\omega$ by using the shifts $T_g(x_n)_{n=1}^\infty := (x_n + g_n)_{n=1}^\infty$ when $g = \sum_{n=1}^\infty g_n e_n$ with $g_n \in \mathbb{F}_p$ (and with all but finitely many of the g_n vanishing).

The quadratic polynomial $Q : \prod \mathbb{F}_p \to \mathbb{F}_p$ defined formally by

$$Q((x_n)_{n=1}^\infty) = \sum_{n=1}^\infty x_n x_{n+1}$$

need not be well-defined, as the sum may contain infinitely many nonzero terms; however, the formal derivative for $g = \sum_{n=1}^\infty g_n e_n \in G$

$$\Delta_g Q = \sum_{n=1}^\infty g_n x_{n+1} + x_n g_{n+1} + g_n g_{n+1}$$

is a well-defined linear polynomial on $\prod \mathbb{F}_p$ since only finitely many of the g_n are nonzero. Setting $\rho_2(g,x) := \Delta_g Q(x)$, we see that $\rho_2(g,x)$ is a polynomial $(G, \mathrm{X}_1, \mathbb{F}_p)$-cocycle of degree less than 2. The cocycle extension $\mathrm{X}_2 := \mathrm{X}_1 \times_2 \mathbb{F}_p$ is then a 2-step Weyl system with structure groups $\prod \mathbb{F}_p$ and $\mathbb{F}_p$, with shift given by

$$T_g((x_n)_{n=1}^\infty, t) = ((x_n + g_n)_{n=1}^\infty, t + \sum_{n=1}^\infty g_n x_{n+1} + x_n g_{n+1} + g_n g_{n+1}).$$

This system can be viewed as a G-system analogue to a 2-step nilsystem arising from the Heisenberg group.

Our first main result, which is a corollary of our previous work in [6], establishes the existence of a Weyl system as a characteristic factor for the averages (1.3), (1.4):

Theorem 1.6 (Characteristic factor). *Let p be a prime, and let $1 \le k < p$ be an integer. Let $G := \mathbb{F}_p^\omega$, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g \in G})$ be an ergodic G-system. Then for each $0 \le k < p$, there exists a factor $(\mathrm{Z}_k, \pi_k) = (\mathrm{Z}_k(\mathrm{X}), \pi_k)$ of X, with Z_k an ergodic continuous k-step Weyl system having the following properties.*

(i) *(Recursive description) $\mathrm{Z}_k = \mathrm{Z}_{k-1} \times_{\rho_k} U_k$ for some compact abelian group U_k and some polynomial $(G, \mathrm{Z}_{k-1}, U_k)$-cocycle ρ_k of degree less than k. Furthermore, U_k is a p-torsion group (thus $pu_k = 0$ for all $u_k \in U_k$ or, equivalently, U_k is a vector space over $\mathbb{F}_p$).*

(ii) *(Connection with polynomials) The sub-σ-algebra of $\mathcal{X}$ generated by Z_k is generated by the polynomials $\phi : X \to \mathbb{R}/\mathbb{Z}$ of degree less than $k+1$. (Thus, for instance, Z_1 is the Kronecker factor.)*

(iii) *(Z_{k-1} characteristic for* (1.3)*) For distinct nonzero* $c_1, \dots, c_k \in \mathbb{F}_p \backslash \{0\}$, *the averages* (1.3) *converge strongly in* $L^2(X)$ *to* 0 *for any Følner sequence* $(\Phi_n)_{n=1}^\infty$ *of* G, *whenever* $f_1, \dots, f_k \in L^\infty(X)$ *is such that* $(\pi_{k-1})_* f_i = 0$ *for at least one* $i = 1, \dots, k$.

(iv) *(Z_{k-1} characteristic for* (1.4)*) For distinct* $c_0, \dots, c_k \in \mathbb{F}_p$, *the average* (1.4) *converges to* 0 *for any Følner sequence* $(\Phi_n)_{n=1}^\infty$ *of* G, *whenever* $f_0, \dots, f_k \in L^\infty(X)$ *is such that* $(\pi_{k-1})_* f_i = 0$ *for at least one* $i = 0, \dots, k$.

(v) *(Z_k characteristic for* (1.7)*) For distinct* $c_0, \dots, c_k \in \mathbb{F}_p$, *whenever* $f_0, \dots, f_k \in L^\infty(X)$ *is such that* $(\pi_k)_* f_i = 0$ *for some* $i = 0, \dots, k-1$, *the sequence* $I_{c_0,\dots,c_k; f_0,\dots,f_k} : G \to \mathbb{C}$ *converges to* 0 *in uniform density.*

We prove Theorem 1.6 in Section 3. We now turn to a discussion of some consequences of this result, starting with a limit formula for the average (1.4). We need the following construction.

Definition 1.7 (Hall-Petresco groups)**.** Let p be a prime, and let $U_1, \dots, U_m$ be compact p-torsion groups for some $0 \le m < p$. Let $1 \le k < p$, and suppose $c_0, \dots, c_k \in \mathbb{F}_p$ are distinct. The **Hall-Petresco group** $HP_{c_0,\dots,c_k}(U_1, \dots, U_m)$ is defined to be the closed subgroup of $(U_1 \times \dots \times U_m)^{k+1}$ consisting of tuples of the form $(P(c_i))_{i=0}^k$, where $P = (P_1, \dots, P_m) : \mathbb{Z} \to U_1 \times \cdots \times U_m$ and for each $1 \le j \le m$, $P_j : \mathbb{Z} \to U_j$ is a polynomial of degree less than $j+1$.

If $X = U_1 \times_{\rho_2} U_2 \times_{\rho_3} \cdots \times_{\rho_m} U_m$ is an ergodic m-step Weyl system, we abbreviate $HP_{c_0,\dots,c_k}(U_1, \dots, U_m)$ as $HP_{c_0,\dots,c_k}(X)$.

Thus, for instance

$$\begin{aligned} HP_{0,1}(U_1, U_2, U_3) &= \Big\{ ((a_1, a_2, a_3), (a_1 + b_1, a_2 + b_2, a_3 + b_3)) : \\ &\qquad\qquad a_1, b_1 \in U_1; a_2, b_2 \in U_2; a_3, b_3 \in U_3 \Big\} \\ &= (U_1 \times U_2 \times U_3)^2, \end{aligned}$$

and (for $p > 2$)

$$\begin{aligned} HP_{0,1,2}(U_1, U_2, U_3) &= \Big\{ \big((a_1, a_2, a_3), (a_1 + b_1, a_2 + b_2, a_3 + b_3), \\ &\qquad\qquad (a_1 + 2b_1, a_2 + 2b_2 + c_2, a_3 + 2b_3 + c_3)\big) : \\ &\qquad\qquad a_1, b_1 \in U_1; a_2, b_2, c_2 \in U_2; a_3, b_3, c_3 \in U_3 \Big\} \\ &= \Big\{ (h_0, h_1, h_2) \in U_1 \times U_2 \times U_3 : h_{01} - 2h_{11} + h_{21} = 0 \Big\} \end{aligned}$$

(with the convention $h_i = (h_{i1}, h_{i2}, h_{i3})$), and (for $p > 3$)

$$\begin{aligned} HP_{0,1,2,3}(U_1, U_2, U_3) = \Big\{ \big(&(a_1, a_2, a_3), (a_1 + b_1, a_2 + b_2, a_3 + b_3), \\ &(a_1 + 2b_1, a_2 + 2b_2 + c_2, a_3 + 2b_3 + c_3), \\ &(a_1 + 3b_1, a_2 + 3b_2 + 3c_2, a_3 + 3b_3 + 3c_3 + d_3) \big) : \\ &a_1, b_1 \in U_1; a_2, b_2, c_2 \in U_2; a_3, b_3, c_3, d_3 \in U_3 \Big\} \\ = \Big\{ &(h_0, h_1, h_2) \in U_1 \times U_2 \times U_3 : h_{01} - 2h_{11} + h_{21} \\ &= h_{11} - 2h_{21} + h_{31} = 0, h_{02} - 3h_{12} + 3h_{22} - h_{32} = 0 \Big\}. \end{aligned}$$

The following lemma, which we prove in Section 5, asserts that the Hall-Petresco group $HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})$ controls the equidistribution of progressions $(T_{c_0 g}x, \ldots, T_{c_k g}x)$ in X.

Lemma 1.8 (First limit formula)**.** *Let p be a prime, let $1 \leq k < p$ be an integer, and let $c_0, \ldots, c_k \in \mathbb{F}_p$ be distinct. Let $G := \mathbb{F}_p^\omega$, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ be an ergodic G-system. Let $f_0, \ldots, f_k \in L^\infty(\mathrm{X})$, and let $(\Phi_n)_{n=1}^\infty$ be a Følner sequence in G. Then*

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X (T_{c_0 g} f_0) \ldots (T_{c_k g} f_k) d\mu = \int_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})} (\pi_{k-1})_* f_0 \otimes \ldots \otimes (\pi_{k-1})_* f_k \, dm_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})}, \tag{1.9}$$

where $(\mathrm{Z}_{k-1}, \pi_{k-1})$ is the characteristic factor from Theorem 1.6, $m_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})}$ is the Haar probability measure on $HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})$, and

$$(\pi_{k-1})_* f_0 \otimes \ldots \otimes (\pi_{k-1})_* f_k : \mathrm{Z}_{k-1}^{k+1} \to \mathbb{C}$$

is the tensor product

$$(\pi_{k-1})_* f_0 \otimes \ldots \otimes (\pi_{k-1})_* f_k(x_0, \ldots, x_k) := (\pi_{k-1})_* f_0(x_0) \ldots (\pi_{k-1})_* f_k(x_k).$$

The right-hand side of (1.9) *can also be written more explicitly as*

$$\int_{U_1^2 \times \ldots \times U_{k-1}^k} \prod_{i=0}^k (\pi_{k-1})_* f_i \Big(\Big(\sum_{l=0}^j \binom{c_i}{l} a_{jl} \Big)_{j=1}^{k-1} \Big),$$

where the integral is over all tuples $(a_{jl})_{1\leq j\leq k-1; 0\leq l\leq j}$ with $a_{jl} \in U_j$, integrated using the product Haar measure on $U_1^2 \times \cdots \times U_{k-1}^k$, and $U_1, \ldots, U_{k-1}$ are the structure groups of Z_{k-1}.

$HP_{c_0,\dots,c_k}(\mathrm{Z}_{k-1})$ contains the diagonal group $\{(x,\dots,x) : x \in Z_{k-1}\}$ and so surjects onto each of the $k+1$ coordinates of $(Z_{k-1})^{k+1}$. In particular, the right-hand side of (1.9) is well-defined even though each of the f_i are only defined up to μ-almost everywhere equivalence.

As examples of the formula (1.9), we have (for $p > 2$)

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0(T_g f_1)(T_{2g} f_2) d\mu$$
$$= \int_{U_1^2} (\pi_1)_* f_0(x)(\pi_1)_* f_1(x+t)(\pi_1)_* f_2(x+2t) dm_{U_1}(x) dm_{U_1}(t)$$

and (for $p > 3$)

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0(T_g f_1)(T_{2g} f_2)(T_{3g} f_3) d\mu$$
$$= \int_{U_1^2\times U_2^3} (\pi_2)_* f_0(x_1,x_2)(\pi_2)_* f_1(x_1+t_1, x_2+t_2)(\pi_2)_* f_2(x_1+2t_1, x_2+2t_2+u_2)$$
$$(\pi_2)_* f_3(x_1+3t_1, x_2+3t_2+3u_2)$$
$$dm_{U_1}(x_1) dm_{U_1}(t_1) dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_2}(u_2).$$

We also remark that if for $G = \mathbb{Z}$ one considers nilsystems instead of Weyl systems, the analogue of the Hall-Petresco group is the group of Hall-Petresco sequences [19], [26], as can be seen from the equidistribution theory in [25].

By duality, the above limit formula (1.9) also gives a formula for limits (in $L^2(\mathrm{X})$) of the form

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} (T_{c_1 g} f_1)\dots(T_{c_k g} f_k) d\mu; \tag{1.10}$$

for instance, the limit $\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n}(T_g f_1)(T_{2g} f_2) d\mu$ in $L^2(\mathrm{X})$ is the function

$$x \mapsto \int_{U_1} (\pi_1)_* f_1(\pi_1(x)+t)(\pi_1)_* f_2(\pi_1(x)+2t) dm_{U_1}(t).$$

The formula in the general case has a similar (but messier) appearance, and is omitted here. In the above argument, we implicitly used the known result that the limit (1.10) in $L^2(\mathrm{X})$ exists; but, in fact, the arguments in this paper give an independent proof of this norm convergence result, see Remark 5 below.

We also have an analogous limit formula for the correlation functions $I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)$, which approximates these functions by a certain integral expression $J_{c_0,\dots,c_k;f_0,\dots,f_k}(g)$ up to a vanishingly small error in uniform density, in analogy to a similar result [4, Proposition 6.5] for $\mathbb{Z}$-systems. To state this formula, we need some additional notation. Let $G = \mathbb{F}_p^\omega$ for a prime p, let $1 \le k < p$, and let Z_k

be the characteristic factor from Theorem 1.6, with structure groups $U_1, \ldots, U_k$. For each $1 \leq i \leq k$, let $u_i : Z_k \to U_i$ be the coordinate function. Observe from the definition of a Weyl system that $\Delta_g u_i = \rho_i \circ \pi_{i-1}$ for all $1 \leq i \leq k$, where π_{i-1} is the projection from Z_{k-1} to Z_{i-1}. In particular, as ρ_i is a polynomial of degree less than i, u_i is a polynomial of degree less than $i+1$. For any $g \in G$, the derivative $\Delta_g^i u_i$ is a constant function, and can thus be identified with an element of U_i. Given $g \in G$, we define the tuple $\theta(g) = (\theta_1(g), \ldots, \theta_k(g)) \in U_1 \times \cdots \times U_k$ by the formula $\theta_i(g) := \Delta_g^i u_i$. We then define $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_{\theta(g)}$ to be the subset of $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)$ consisting of tuples $(P(c_0), \ldots, P(c_k))$ with $P = (P_1, \ldots, P_k)$, where each $P_i : \mathbb{Z} \to U_i$ is a polynomial of degree less than $i+1$ obeying the additional constraint $\Delta_1^i P_i = \theta_i(g)$ on the leading coefficient of each of the P_i. Note that $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_0$ is a closed subgroup of $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)$, and $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_\theta$ is a coset of $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_0$ for any $\theta \in U_1 \times \cdots \times U_k$. In particular, $HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_{\theta(g)}$ has a well-defined Haar measure $dm_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_{\theta(g)}}$.

Lemma 1.9 (Second limit formula)**.** *Let p be a prime, let $1 \leq k < p$ be an integer, and let $c_1, \ldots, c_k \in \mathbb{F}_p$ be distinct. Let $G := \mathbb{F}_p^\omega$ and $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g \in G})$ be an ergodic G-system. Let $f_0, \ldots, f_k \in L^\infty(\mathrm{X})$, and let $(\Phi_n)_{n=1}^\infty$ be a Følner sequence in G. Define the sequence $J_{c_0,\ldots,c_k;f_1,\ldots,f_k} : G \to \mathbb{C}$ by the formula*

$$J_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g) := \int_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_{\theta(g)}} (\pi_k)_* f_0 \otimes \ldots \otimes (\pi_k)_* f_k \, dm_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_k)_{\theta(g)}}.$$

Then the difference $I_{c_0,\ldots,c_k;f_0,\ldots,f_k} - J_{c_0,\ldots,c_k;f_0,\ldots,f_k}$ converges to 0 in uniform density.

Let us write $I(g) \approx_{UD} J(g)$ for the assertion that $I(g) - J(g)$ converges to 0 in uniform density. Then a simple special case of Lemma 1.9 is the approximation

$$\int_X f_0 T_g f_1 \, d\mu \approx_{UD} \int_{U_1} (\pi_1)_* f_0(x) (\pi_1)_* f_1(x + \Delta_g u_1) \, dm_{U_1}(x);$$

similarly, we have (for $p > 2$)

$$\int_X f_0 T_g f_1 T_{2g} f_2 \, d\mu \approx_{UD} \int_{U_1 \times U_2^2} (\pi_2)_* f_0(x_1, x_2) (\pi_2)_* f_1(x_1 + \Delta_g u_1, x_2 + t_2)$$
$$(\pi_2)_* f_2(x_1 + 2\Delta_g u_1, x_2 + 2t_2 + \Delta_g^2 u_2)$$
$$dm_{U_1}(x_1) dm_{U_2}(x_2) dm_{U_2}(t_2)$$

and (for $p > 3$)

$$\int_X f_0 T_g f_1 T_{2g} f_2 T_{3g} f_3 d\mu$$
$$\approx_{UD} \int_{U_1 \times U_2^2 \times U_3^3} (\pi_3)_* f_0(x_1, x_2, x_3)(\pi_3)_* f_1(x_1 + \Delta_g u_1, x_2 + t_2, x_3 + t_3)$$
$$(\pi_3)_* f_2(x_1 + 2\Delta_g u_1, x_2 + 2t_2 + \Delta_g^2 u_2, x_3 + 2t_3 + s_3)$$
$$(\pi_3)_* f_3(x_1 + 3\Delta_g u_1, x_2 + 3t_2 + 3\Delta_g^2 u_2, x_3 + 3t_3 + 3s_3 + \Delta_g^3 u_3)$$
$$dm_{U_1}(x_1) dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_3}(x_3) dm_{U_3}(t_3) dm_{U_3}(s_3).$$

We prove Lemma 1.9 in Section 6. The sequence $J_{c_0,\ldots,c_k;f_0,\ldots,f_k}$ can also be viewed as a "Weyl sequence" (analogous to the concept of a nilsequence, but with respect to a Weyl system rather than a nilsystem).

Proposition 1.10 (Structure theorem)**.** *Let the notation be as in Lemma 1.9. Then there exists a continuous k-step Weyl system $\mathrm{Y} = (Y, \mathcal{Y}, \nu, (S_g)_{g\in G})$, a continuous function $F \in L^\infty(\mathrm{Y})$, and a point $y_0 \in Y$ such that $J_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g) = F(S_g y_0)$ for all $g \in G$.*

We prove this in Section 7. Combining this proposition with Lemma 1.9, we see that $I_{c_0,\ldots,c_k;f_0,\ldots,f_k}$ is approximated by a k-step Weyl sequence up to an error that goes to 0 in uniform density (cf. [4, Theorem 1.9]).

In analogy to [4], we can use the limit formulae to obtain Khintchine type recurrence theorems.

Definition 1.11 (Khintchine property)**.** Let p be a prime, and let $c_0, \ldots, c_k$ be distinct elements of $\mathbb{F}_p$. We say that the tuple $(c_0, \ldots, c_k)$ has the **Khintchine property** (in characteristic p) if, whenever $G = \mathbb{F}_p^\omega$, $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ is an ergodic G-system, $A \in \mathcal{X}$, and $\varepsilon > 0$, the set

$$\{g \in G : \mu(T_{-c_0 g} A \cap \cdots \cap T_{-c_k g} A) \geq \mu(A)^{k+1} - \varepsilon\}$$

is a syndetic subset of G (i.e., G can be covered by finitely many translates of this set).

Of course, the negative signs in the subscripts here can be easily deleted if desired. It is trivial that any singleton tuple (c_0) has the Khintchine property, and the classical Khintchine recurrence theorem adapted to general abelian groups G implies that any pair (c_0, c_1) has the Khintchine property (in this case, we do not need to assume the ergodicity of our G-system). It is also clear that the Khintchine property is preserved if one applies an invertible affine tranformation

$x \mapsto ax + b$ to each element c_i of a tuple $(c_0, \dots, c_k)$, i.e., $(c_0, \dots, c_k)$ has the Khintchine property if and only if $(ac_0 + b, \dots, ac_k + b)$ has the Khintchine property. For longer tuples, we have the following positive results, which are finite characteristic analogues of results in [4].

Theorem 1.12 (Khintchine for double recurrence). *If $p > 2$ and c_0, c_1, c_2 are distinct elements of $\mathbb{F}_p$, then (c_0, c_1, c_2) has the Khintchine property.*

Theorem 1.13 (Khintchine for triple recurrence). *If $p > 3$ and c_0, c_1, c_2, c_3 are distinct elements of $\mathbb{F}_p$ which form a parallelogram in the sense that $c_i + c_j = c_k + c_l$ for some permutation $\{i, j, k, l\}$ of $\{1, 2, 3, 4\}$, then (c_0, c_1, c_2, c_3) has the Khintchine property.*

For comparison, a version of the classical Khintchine Recurrence Theorem for $G = \mathbb{F}_p^\omega$ implies that for distinct $c_0, c_1 \in \mathbb{F}_p$, a G-system $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g \in G})$, $A \in \mathcal{X}$, and $\varepsilon > 0$, the set

$$\{g \in G : \mu(T_{-c_0 g} A \cap T_{-c_1 g} A) \geq \mu(A)^2 - \varepsilon\}$$

is syndetic. In this classical setting of single recurrence, no ergodicity hypothesis is required; but by adapting the construction in [4, Theorem 2.1], one can show that ergodicity is needed for double or higher recurrence if p is sufficiently large (this hypothesis is needed to embed a version of the Behrend-type constructions used in [4]).

We prove these results in Section 8 and Section 9, respectively. We remark that a finitary analogue of Theorem 1.13, concerning dense subsets of finite-dimensional vector spaces $\mathbb{F}_p^n$ instead of subsets of $\mathbb{F}_p^\omega$-systems (and with the shifts g lying in a dense subset of $\mathbb{F}_p^n$, rather than a syndetic subset of $\mathbb{F}_p^\omega$), was established in [17, Theorem 4.1].

We conjecture that the above results exhaust all the possible tuples with the Khintchine property.

Conjecture 1.14. Let p be a prime, let $k < p$, and let $c_0, \dots, c_k$ be distinct elements of $\mathbb{F}_p$.

(i) If $k > 3$, then $(c_0, \dots, c_k)$ does not have the Khintchine property.

(ii) If $k = 3$, and (c_0, c_1, c_2, c_3) does not form a parallelogram, (c_0, c_1, c_2, c_3) does not have the Khintchine property.

In [4, Corollary 1.6], it was shown that the tuple $(0, 1, 2, 3, 4)$ do not have the analogous Khintchine property for $\mathbb{Z}$-systems; and it is not difficult to modify the construction there to show that $(0, 1, 2, 3, 4)$ does not have the Khintchine property in characteristic p if p is sufficiently large and similarly if $(0, 1, 2, 3, 4)$ is replaced by $(0, 1, \ldots, k)$ for any fixed $k \geq 4$, if p is sufficiently large depending on k.

While we were not able[8] establish the above conjecture in general, we can do so for "generic" tuples $(c_0, \ldots, c_k)$.

Theorem 1.15 (Khintchine property generically fails)**.** *Let $k \geq 3$. Then there exists a constant C_k depending only on k such that for any prime p, there exist at most $C_k p^k$ tuples $(c_0, \ldots, c_k) \in \mathbb{F}_p^{k+1}$ that have the Khintchine property.*

In other words, if one selects $c_0, \ldots, c_k \in \mathbb{F}_p$ uniformly at random, then the Khintchine property holds with probability at most C_k/p; so for large p, one has failure of the property for most tuples $(c_0, \ldots, c_k)$.

We prove this in Section 10. It remains open whether a weakened version of the Khintchine property can hold in which $\mu(A)^{k+1}$ is replaced by a larger power $\mu(A)^{C_k}$ of $\mu(A)$. In the case of $\mathbb{Z}$-systems, this was shown in [4, Corollary 1.6] not to be the case, at least for the model case $k = 4$ and $c_i = i$. However, that argument relies on the Behrend construction [2], and it remains an interesting open problem to adapt this construction to the finite field setting when the characteristic p is fixed. Note that it follows from the "syndetic" Szemerédi theorem for vector spaces over finite fields [15] that the Khintchine property does hold if $\mu(A)^{k+1}$ is replaced by some sufficiently small quantity $c(k, \mu(A)) > 0$ depending only on k and $\mu(A)$, if $\mu(A)$ is nonzero.

2 Continuity of polynomials

In this section, we establish a technical lemma which asserts, roughly speaking, that polynomials in an ergodic Weyl system are automatically continuous.

Lemma 2.1 (Polynomials are continuous)**.** *Let $G = (G, +)$ be a countable abelian group, let $k \geq 0$, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g \in G})$ be an ergodic k-step Weyl system.*

(i) *After modifying the cocycles used to define X on a set of measure zero, if necessary, X becomes a continuous k-step Weyl system.*
(ii) *If $\phi : X \to \mathbb{R}/\mathbb{Z}$ is a polynomial, then (after redefining ϕ on a a set of measure zero, if necessary), ϕ is continuous.*

[8] By some extremely lengthy computations involving a subdivision into a large number of subcases, and *ad hoc* constructions of counterexamples in each case, we have been able to verify this conjecture in the case when $k = 3$, c_0, c_1, c_2, c_3 are fixed integers that do not form a parallelogram, and p is sufficiently large depending on c_0, c_1, c_2, c_3 (or alternatively, if one considers $\mathbb{Z}$-systems rather than $\mathbb{F}_p^\omega$-systems). We plan to make details of these constructions available elsewhere.

Proof. Induction on k. The case $k = 0$ is trivial, so suppose that $k \geq 1$ and that claims (i), (ii) have already been proven for all smaller values of k. The claim (i) for k then follows by applying the induction hypothesis (ii) to all the cocycles used to construct X, so now we turn to claim (ii) for k. Write $\mathrm{X} = \mathrm{X}_{k-1} \times_{\rho_k} U_k$ for some compact abelian U_k and some polynomial $(G, \mathrm{X}_{k-1}, U_k)$ cocycle ρ_k of degree less than k. By claim (i) for k, we may assume without loss of generality that all cocycles involved in constructing X (including ρ_k) are continuous.

Let us first handle the case when the compact group U_k is finite (and thus discrete). For each $u_k \in U_k$, $\phi_{u_k} : X_{k-1} \to \mathbb{R}/\mathbb{Z}$ defined by $\phi_{u_k}(x_{k-1}) := \phi(x_{k-1}, u_k)$ is a polynomial on X_{k-1} (see [6, Lemma B.5(iii)]) and can thus be modified on a set of measure zero to become continuous. Applying this for each u_k and gluing, we obtain the claim.

Now we turn to the general case, in which U_k is not necessarily finite. For $t \in U_k$, define the **vertical derivative** $\Delta_t\phi : X \to \mathbb{R}/\mathbb{Z}$ of ϕ by the formula

$$\Delta_t\phi(x_{k-1}, u_k) := \phi(x_{k-1}, u_k + t) - \phi(x_{k-1}, u_k).$$

As ϕ is a polynomial, $\Delta_t\phi$ is a polynomial of uniformly bounded degree. On the other hand, as ϕ is measurable, $\Delta_t\phi$ converges to 0 in measure as $t \to 0$ in U_k; in particular, $\|e(\Delta_t\phi) - 1\|_{L^2(\mathrm{X})} \to 0$ as $t \to 0$, where $e(x) := e^{2\pi i x}$ is the standard character on $\mathbb{R}/\mathbb{Z}$. By [6, Lemma C.1] and the uniformly bounded degree of $\Delta_t\phi$, we conclude that $e(\Delta_t\phi)$ must be almost everywhere constant for t sufficiently close to 0. In particular, for t sufficiently close to 0, there exists $\chi(t) \in \mathbb{R}/\mathbb{Z}$ such that $\Delta_t\phi = \chi(t)$ almost everywhere. The set of all t with this property is easily seen to form a compact open subgroup U'_k of U_k, and χ is a homomorphism from U'_k to $\mathbb{R}/\mathbb{Z}$ which is measurable, and hence continuous (Steinhaus lemma). By Pontryagin duality, χ can then be extended to an additive character from U_k to $\mathbb{R}/\mathbb{Z}$. Defining the function $\psi : X \to \mathbb{R}/\mathbb{Z}$ by

$$\psi(x_{k-1}, u_k) := \phi(x_{k-1}, u_k) - \chi(u_k),$$

we see that ψ is also a polynomial on X, with $\Delta_t\psi = 0$ for $t \in U'_k$; thus, after modification on a set of measure zero, ψ is constant on all U'_k-orbits. We can use U'_k to form quotient spaces from U_k and ρ_k by U'_k and reduce to a Weyl system $\mathrm{X}_{k-1} \times_{\rho_k \bmod U'_k} U_k/U'_k$ in which the final group U_k/U'_k is finite. By the case already treated, the pushdown of ψ can be modified on a set of measure zero to become continuous on this system. Thus, on applying pullback, the same claim holds for ψ and hence for ϕ, giving the claim. □

3 Gowers-Host-Kra seminorms and characteristic factors

In this section, we derive Theorem 1.6 from the theory of Gowers-Host-Kra seminorms on $\mathbb{F}_p^\omega$-systems as developed in [6]. While the material here is standard for $\mathbb{Z}$-systems (see [23]) and the adaptation of that theory to $\mathbb{F}_p^\omega$-systems routine, we present it here for the sake of completeness.

We recall the definition of the Gowers-Host-Kra seminorms, introduced in [23] (and closely related to the combinatorial Gowers uniformity norms from [18]):

Definition 3.1 (Gowers-Host-Kra seminorms)**.** [23] Let $G = (G, +)$ be a countable abelian group, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ be a G-system. For $f \in L^\infty(\mathrm{X})$, we define the **Gowers-Host-Kra seminorms** $\|f\|_{U^k(\mathrm{X})}$ recursively for $k \geq 1$ by setting

$$\|f\|_{U^1(\mathrm{X})} := \lim_{n\to\infty} \|\mathbb{E}_{h\in\Phi_n} T_h f\|_{L^2(\mathrm{X})} \tag{3.1}$$

and

$$\|f\|_{U^k(\mathrm{X})} := \left(\lim_{n\to\infty} \|\mathbb{E}_{h\in\Phi_n} T_h f \overline{f}\|_{U^{k-1}(\mathrm{X})}^{2^{k-1}}\right)^{1/2^k} \tag{3.2}$$

for any $k \geq 2$ and any Følner sequence $(\Phi_n)_{n=1}^\infty$ of G.

It can be shown that the above definitions are, in fact, independent of the choice of the Følner sequence and define a sequence of seminorms on $L^\infty(\mathrm{X})$; see[9] [6, Lemma A.18]. From the Mean Ergodic Theorem, we observe that the U^1 seminorm can also be written as

$$\|f\|_{U^1(\mathrm{X})} = \|(\pi_0)_* f\|_{L^2(\mathrm{Z}_0)}, \tag{3.3}$$

where (Z_0, π_0) is the invariant factor.

The significance of these seminorms for us is that they control the convergence of expressions such as (1.3). More precisely, we have the following minor variant of [23, Theorem 12.1].

Lemma 3.2 (Generalized van der Corput lemma)**.** *Let $G = \mathbb{F}_p^\omega$ for a prime p, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ be a G-system. Let $1 \leq k < p$, and let $c_1, \ldots, c_k$ be distinct elements of $\mathbb{F}_p \backslash \{0\}$. Let $f_1, \ldots, f_k \in L^\infty(\mathrm{X})$. Let $(\Phi_n)_{n=1}^\infty$ be a Følner*

[9] Strictly speaking, in that lemma the additional hypothesis of nestedness $\Phi_1 \subset \Phi_2 \subset \ldots$ of the Følner sequence is imposed; but an inspection of the proof shows that this hypothesis is not needed (because the mean ergodic theorem holds for nonnested Følner sequences).

sequence of G. Then

$$\limsup_{n\to\infty} \|\mathbb{E}_{g\in\Phi_n}(T_{c_1g}f_1)(T_{c_2g}f_2)\dots(T_{c_kg}f_k)\|_{L^2(X)} \le \inf_{1\le i\le k} \|f_i\|_{U^k(X)} \prod_{1\le j\le k: j\ne i} \|f_j\|_{L^\infty(X)}.$$

Proof. Induction on k. When $k = 1$, we may rescale Φ_n by c_1 to normalize $c_1 = 1$, and then the claim follows from (3.1). Now suppose that $k > 1$, and that the claim has already been proved for $k-1$. By permuting indices, it suffices to show that

$$\limsup_{n\to\infty} \|\mathbb{E}_{g\in\Phi_n}F_g\|_{L^2(X)} \le \|f_k\|_{U^k(X)}, \tag{3.4}$$

where $\|f_i\|_{L^\infty(X)} \le 1$ for $i = 1,\dots,k-1$, and

$$F_g := (T_{c_1g}f_1)(T_{c_2g}f_2)\dots(T_{c_kg}f_k).$$

We may also normalize $c_k = 1$. Using the Følner property, we can rewrite the left-hand side of (3.4) as

$$\limsup_{m\to\infty}\limsup_{n\to\infty} \|\mathbb{E}_{g\in\Phi_n}\mathbb{E}_{h\in\Phi_m}F_{g+h}\|_{L^2(X)},$$

which we can then bound using the triangle inequality by

$$\limsup_{m\to\infty}\limsup_{n\to\infty} \mathbb{E}_{g\in\Phi_n}\|\mathbb{E}_{h\in\Phi_m}F_{g+h}\|_{L^2(X)}.$$

By Cauchy-Schwarz, this is bounded by

$$\left(\limsup_{m\to\infty}\limsup_{n\to\infty} \mathbb{E}_{g\in\Phi_n}\|\mathbb{E}_{h\in\Phi_m}F_{g+h}\|^2_{L^2(X)}\right)^{1/2},$$

which we may expand as

$$\left(\limsup_{m\to\infty}\limsup_{n\to\infty} \mathbb{E}_{h,h'\in\Phi_m}\int_X \mathbb{E}_{g\in\Phi_n}F_{g+h}\overline{F_{g+h'}}\,d\mu\right)^{1/2}.$$

This last expression is bounded above by

$$\left(\limsup_{m\to\infty}\mathbb{E}_{h,h'\in\Phi_m}\limsup_{n\to\infty} \left|\int_X \mathbb{E}_{g\in\Phi_n}F_{g+h}\overline{F_{g+h'}}\right|d\mu\right)^{1/2}.$$

Now, for each h, h', we may write

$$\int_X \mathbb{E}_{g\in\Phi_n}F_{g+h}\overline{F_{g+h'}} = \int_X (T_{c_1h}f_1)\overline{T_{c_1h'}f_1}\,\mathbb{E}_{g\in\Phi_n}\prod_{i=2}^k T_{(c_i-c_1)g}((T_{c_ih}f_i)\overline{T_{c_ih'}f_i})\,d\mu.$$

Applying Cauchy-Schwarz and the induction hypothesis (and the normalization $c_k = 1$), we conclude that

$$\limsup_{n\to\infty} |\int_X \mathbb{E}_{g\in\Phi_n} F_{g+h}\overline{F_{g+h'}}|d\mu \le \|(T_h f_k)\overline{T_{h'} f_k}\|_{U^{k-1}(\mathrm{X})}.$$

Note that $(T_h f_k)\overline{T_{h'} f_k}$ has the same $U^{k-1}(\mathrm{X})$ norm as $(T_{h-h'} f_k)\overline{f_k}$. Putting all this together, we can bound the left-hand side of (3.4) by

$$\left(\limsup_{m\to\infty} \mathbb{E}_{h,h'\in\Phi_m} \|T_{h-h'} f_k \overline{f_k}\|_{U^{k-1}(\mathrm{X})}\right)^{1/2}.$$

By the triangle inequality and the pigeonhole principle, we can bound this by

$$(\limsup_{m\to\infty} \mathbb{E}_{h\in\Phi_m - h'_m} \|T_h f_k \overline{f_k}\|_{U^{k-1}(\mathrm{X})})^{1/2}$$

for some sequence $h'_m \in \Phi_m$; and by Hölder's inequality, this is bounded by

$$(\limsup_{m\to\infty} \mathbb{E}_{h\in\Phi_m - h'_m} \|T_h f_k \overline{f_k}\|_{U^{k-1}(\mathrm{X})}^{2^{k-1}})^{1/2^k}.$$

But the $\Phi_m - h'_m$ form a Følner sequence of G, and the claim (3.4) then follows from (3.2). $\square$

We also need the following variant.

Lemma 3.3 (Generalized van der Corput lemma, II)**.** *Let $G = \mathbb{F}_p^\omega$ for a prime p, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ be a G-system. Let $1 \le k < p$, and let $c_0, c_1, \dots, c_k$ be distinct elements of $\mathbb{F}_p$. Let $f_0, \dots, f_k \in L^\infty(\mathrm{X})$, and let $(\Phi_n)_{n=1}^\infty$ be a Følner sequence of G. Then*

$$\limsup_{n\to\infty} \sup_{h\in G} \mathbb{E}_{g\in h+\Phi_n} |I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| \le \inf_{0\le i\le k} \|f_i\|_{U^{k+1}(\mathrm{X})} \prod_{0\le j\le k: j\ne i} \|f_j\|_{L^\infty(\mathrm{X})}.$$

Proof. As before, it suffices to show that

$$\limsup_{n\to\infty} \sup_{h\in G} \mathbb{E}_{g\in h+\Phi_n} |I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| \le \|f_k\|_{U^{k+1}(\mathrm{X})}$$

under the normalization $\|f_j\|_{L^\infty(\mathrm{X})} \le 1$ for $0 \le j < k$.

Next, we remove the supremum in the above estimate. Suppose that we have already shown that

$$\limsup_{n\to\infty} \mathbb{E}_{g\in\Phi_n} |I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| \le \|f_k\|_{U^{k+1}(\mathrm{X})}. \tag{3.5}$$

For any $\varepsilon > 0$, and any n, we can find $h_n = h_{n,\varepsilon} \in G$ such that

$$\sup_{h\in G} \mathbb{E}_{g\in h+\Phi_n} |I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| \le (1+\varepsilon)\mathbb{E}_{g\in h_n+\Phi_n} |I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)|.$$

Applying (3.5) to the Følner sequence $(h_n + \Phi_n)_{n=1}^{\infty}$, we conclude that

$$\limsup_{n\to\infty} \sup_{h\in G} \mathbb{E}_{g\in h+\Phi_n} |I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| \leq (1+\varepsilon)\|f_k\|_{U^{k+1}(\mathrm{X})}$$

and the claim then follows by sending ε to 0.

It remains to establish (3.5). Write $F_g := (T_{c_0 g} f_0)\cdots(T_{c_k g} f_k)$, so that $I_{c_0,\dots,c_k;f_0,\dots,f_k}(g) = \int_X F_g d\mu$. Then for any m, we have

$$|I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| = \int_X \mathbb{E}_{h\in\Phi_m} T_h F_g d\mu;$$

and hence, by Cauchy-Schwarz, the left-hand side of (3.5) is bounded by

$$(\limsup_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X |\mathbb{E}_{h\in\Phi_m} T_h F_g|^2 d\mu)^{1/2}.$$

We can expand this expression as

$$(\limsup_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}_{h,h'\in\Phi_m} \mathbb{E}_{g\in\Phi_n} \int_X (T_h F_g)(T_{h'}\overline{F_g}) d\mu)^{1/2},$$

which by the triangle inequality is bounded by

$$\left(\limsup_{m\to\infty} \mathbb{E}_{h,h'\in\Phi_m} \left| \limsup_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X (T_h F_g)(T_{h'}\overline{F_g}) d\mu \right|\right)^{1/2}.$$

Rewriting $\int_X (T_h F_g)(T_{h'}\overline{F_g}) d\mu$ as

$$\int_X \prod_{j=0}^{k} T_{c_j g}((T_h f_j)\overline{T_{h'} f_j}) d\mu$$

and applying Lemma 3.2, we may thus bound the left-hand side of (3.5) by

$$(\limsup_{m\to\infty} \mathbb{E}_{h,h'\in\Phi_m} \|(T_h f_k)\overline{T_{h'} f_k}\|_{U^k(\mathrm{X})})^{1/2}.$$

One can then argue as in the proof of Lemma 3.2 to bound this by $\|f_k\|_{U^{k+1}(\mathrm{X})}$, as required. □

Corollary 3.4. *Let $G = \mathbb{F}_p^{\omega}$ for a prime p, and let $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ be a G-system. Let $1 \leq k < p$. Let $f_0, f_1, \dots, f_k \in L^\infty(\mathrm{X})$. Let $(\Phi_n)_{n=1}^{\infty}$ be a Følner sequence of G.*

(i) *The sequence $\mathbb{E}_{g\in\Phi_n}(T_{c_1 g} f_1)\cdots(T_{c_k g} f_k)$ converges in $L^2(\mathrm{X})$ to 0 for distinct nonzero elements $c_1, \dots, c_k$ of $\mathbb{F}_p$ whenever $\|f_i\|_{U^k(\mathrm{X})} = 0$ for some $1 \leq i \leq k$.*

(ii) *The sequence* $\mathbb{E}_{g\in\Phi_n}\int_X (T_{c_0 g} f_0)\cdots(T_{c_k g} f_k)$ *converges to* 0 *for distinct elements* $c_0,\dots,c_k$ *of* $\mathbb{F}_p$ *whenever* $\|f_i\|_{U^k(\mathrm{X})}=0$ *for some* $0\le i\le k$.

(iii) *The sequence* $I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)$ *converges in uniform density to* 0 *for distinct elements* $c_0,\dots,c_k$ *of* $\mathbb{F}_p$ *whenever* $\|f_i\|_{U^{k+1}(\mathrm{X})}=0$ *for some* $0\le i\le k$.

Proof. The claim (i) is immediate from Lemma 3.2. To prove (ii), we may first permute so that $\|f_k\|_{U^k(\mathrm{X})}=0$, and then translate so that $c_0=0$. The claim then follows from (i) after using Cauchy-Schwarz to eliminate f_0. Finally, (iii) follows from Lemma 3.3. □

We remark that one can also prove (iii) using the structure of Host-Kra measures, after performing an ergodic decomposition; see [23, Corollary 4.5].

Theorem 1.6 is then immediate from Corollary 3.4 and the following result from [6].

Theorem 3.5 (Characteristic factor for the U^k norm)**.** *Let* $G=\mathbb{F}_p^\omega$ *for a prime* p*, and let* $\mathrm{X}=(X,\mathcal{X},\mu,(T_g)_{g\in G})$ *be an ergodic* G*-system. For each* $1\le k\le p$*, denote the sub-*σ*-algebra of* $\mathcal{X}$ *generated by the polynomials* $\phi: X\to\mathbb{R}/\mathbb{Z}$ *of degree less than* k *by* $\mathcal{B}_{<k}$*. Then there is a factor* $(\mathrm{Z}_{k-1},\pi_{k-1})$ *of* X *that is equivalent to* $\mathcal{B}_{<k}$*, and there is a continuous ergodic* $(k-1)$*-step Weyl system, with* $\mathrm{Z}_k=\mathrm{Z}_{k-1}\times_{\rho_k} U_k$ *for all* $1\le k<p$*, for some compact abelian* p*-torsion group* U_k*, and some polynomial* (G,Z_{k-1},U_k)*-cocycle* ρ_k *of degree less than* k*. Furthermore, if* $f\in L^\infty(\mathrm{X})$*, then* $\|f\|_{U^k(\mathrm{X})}=0$ *if and only if* $(\pi_{k-1})_*f=0$.

Proof. This follows from [6, Proposition 1.10], [6, Theorem 1.19] and [6, Corollary 8.7], using Lemma 2.1 to ensure that the Weyl system obtained is continuous. The ergodicity of the Weyl systems is automatic because any factor of an ergodic system is again ergodic.

Remark. The condition $k\le p$ was subsequently removed in [32] (but with the important caveat that the groups U_j need no longer be p-torsion, but are merely p^m-torsion for some $m\ge 1$); however, for our application ,we have $k\le p$, so we do not need the (more difficult) arguments from [32] here. It may also be possible to adapt the arguments in [28] to give an alternate proof of Theorem 3.5, but we do not pursue this here.

4 Some special cases of the limit formulae

In the next two sections, we prove the main limit formulae, Lemmas 1.8 1.9. To motivate the proof of these formulae in the general case, we discuss some model cases of these results here.

We begin with a special case of Lemma 1.8, when $p > 2$ and X is a 2-step Weyl system $\mathrm{X} = U_1 \times_{\rho_2} U_2$, where ρ_2 is a polynomial (G, U_1, U_2)-cocycle of degree less than 2. Furthermore, we assume (abusing notation slightly) that the base system U_1 is also the Kronecker factor $\mathrm{Z}_1(\mathrm{X})$; thus the only polynomials of degree less than 2 on X are those which are functions of the U_1 coordinate only. The special case of Lemma 1.8 we discuss is

$$\begin{aligned}(4.1)\quad \lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0(T_g f_1)(T_{2g} f_2)(T_{3g} f_3)\, d\mu &= \int_{U_1^2\times U_2^3} f_0(x_1, x_2) f_1(x_1 + t_1, x_2 + t_2) f_2(x_1 + 2t_1, x_2 + 2t_2 + u_2) \\ &\qquad \cdot f_3(x_1 + 3t_1, x_2 + 3t_2 + 3u_2) \\ &\qquad\qquad dm_{U_1}(x_1) dm_{U_1}(t_1) dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_2}(u_2).\end{aligned}$$

(The factor map π_2 is not needed in this special case, as it is the identity map.) To simplify things further, we assume that each $i = 0, 1, 2, 3$, the function f_i takes the special form

$$f_i(x_1, x_2) = e(\phi_{i,2}(x_2)) \tag{4.2}$$

for some additive character (i.e., continuous homomorphism) $\phi_{i,2} : U_2 \to \mathbb{R}/\mathbb{Z}$. One can (and should) also consider the slightly more general example of functions of the form

$$f_i(x_1, x_2) = e(\phi_{i,1}(x_1) + \phi_{i,2}(x_2)), \tag{4.3}$$

where $\phi_{i,1} : U_1 \to \mathbb{R}/\mathbb{Z}$ is an additive character of U_1, as these Fourier-analytic examples then span a dense subspace of $L^2(\mathrm{X})$. However, to keep the discussion here simple, we ignore the lower order terms $\phi_{i,1}$ and focus only on examples of the form (4.2).

The verification of (4.1) now splits into several cases, depending on the nature of the characters $\phi_{0,2}, \ldots, \phi_{3,2}$. One easy case is when $\phi_{0,2}, \ldots, \phi_{3,2}$ all vanish identically; then both sides of (4.1) are clearly equal to 1.

Next, suppose that $\phi_{3,2}$ vanishes identically, but that one of the other $\phi_{i,2}$, say $\phi_{2,2}$ is not identically 0. Then our task is to show that

$$\begin{aligned}(4.4)\quad \lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f_0(T_g f_1)(T_{2g} f_2)\, d\mu &= \int_{U_2^3} e(\phi_{0,2}(x_2) + \phi_{1,2}(x_2 + t_2) + \phi_{2,2}(x_2 + 2t_2 + u_2)) \\ &\qquad dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_2}(u_2).\end{aligned}$$

Observe that (x_2, t_2, u_2) varies over U_2^3; the tuple $(x_2, x_2+t_2, x_2+2t_2+u_2)$ is unconstrained in U_2^3. In particular, if (x_2, t_2, u_2) is drawn uniformly at random using the Haar measure on U_2^3, then $(x_2, x_2+t_2, x_2+2t_2+u_2)$ is also uniformly distributed with this Haar measure. Thus the right-hand side factors as

$$\left(\int_{U_2} e(\phi_{0,2}) d\mu_2\right)\left(\int_{U_2} e(\phi_{1,2}) d\mu_2\right)\left(\int_{U_2} e(\phi_{2,2}) d\mu_2\right).$$

By Fourier analysis, the third factor vanishes, since $\phi_{2,2}$ is assumed to not be identically 0; so the right-hand side of (4.4) vanishes. As for the left-hand side, observe that the function $f_2(x_1, x_2) = e(\phi_{2,2}(x_2))$ has mean 0 on every coset of U_2 in $U_1 \times U_2$; since we are assuming $U_1 = \mathrm{Z}_1$, this implies that $(\pi_1)_* f_2 = 0$. By Theorem 3.5, this implies that $\|f_2\|_{U^2(\mathrm{X})} = 0$. Applying Corollary 3.4, we conclude that the left-hand side of (4.4) vanishes also, so we are done in this case.

Finally, we consider the case in which $\phi_{3,2}$ does not vanish identically. We can then simplify the right-hand side of (4.1) by noting the extrapolation identity

$$x_2 + 3t_2 + 3u_2 = x_2 - 3(x_2+t_2) + 3(x_2+2t_2+u_2),$$

which allows us to write the right-hand side as

$$\int_{U_2^3} e\left(\phi'_{0,2}(x_2) + \phi'_{1,2}(x_2+t_2) + \phi'_{2,2}(x_2+2t_2+u_2)\right) dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_2}(u_2),$$

where $\phi'_{0,2} := \phi_{0,2} + \phi_{3,2}$, $\phi'_{1,2} := \phi_{1,2} - 3\phi_{3,2}$, and $\phi'_{2,2} := \phi_{2,2} + 3\phi_{3,2}$. Next, observe that $\Delta_g u_2 = \rho(g, \cdot)$ for all $g \in G$; since ρ is a polynomial cocycle of degree less than 2, we conclude that u_2 is a polynomial of degree less than 3 (i.e., a quadratic function). For any $g \in G$ and $x \in X$, the sequence $n \mapsto u_2(T_{ng}x)$ is then also a quadratic polynomial. In particular, we have the interpolation identity

$$u_2(T_{3g}x) = u_2(x) - 3u_2(T_g x) + 3u_2(T_{2g}x),$$

which allows us to write the left-hand side of (4.1) as

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X f'_0 (T_g f'_1)(T_{2g} f'_2) d\mu,$$

where $f'_i(x_1, x_2) := e(\phi'_{i,2}(x_2))$ for $i = 0, 1, 2$. Thus, we have reduced the task of verifying (4.1) when $\phi_{3,2}$ does not vanish identically to the task of verifying (4.1) when $\phi_{3,2}$ does vanish identically, which has already been covered by the preceding arguments. This concludes the demonstration of (4.1) for the model functions (4.2). The model cases (4.3) can be handled by similar arguments, exploiting the linear nature of $n \mapsto u_1(T_{ng}x)$ in addition to the quadratic nature of $n \mapsto u_2(T_{ng}x)$

eventually to reduce to the case in which $\phi_{2,1}, \phi_{3,1}, \phi_{3,2}$ vanish and $\phi_{2,2}$ does not vanish identically, which can then be treated by Theorem 3.5 and Corollary 3.4 as before; we leave the details to the interested reader (they are special cases of the argument in Section 5 below).

Now we consider an analogous example for the second limit formula, Lemma 1.9. Keeping the system $\mathrm{X} = U_1 \times_{\rho_2} U_2$ as before, we now turn to the task of showing that

$$\begin{aligned}(4.5) \quad \int_X f_0 T_g f_1 T_{2g} f_2 T_{3g} f_3 d\mu \approx_{UD} \int_{U_1 \times U_2^2} & f_0(x_1, x_2) f_1(x_1 + \Delta_g u_1, x_2 + t_2) \\ & f_2(x_1 + 2\Delta_g u_1, x_2 + 2t_2 + \Delta_g^2 u_2) f_3(x_1 + 3\Delta_g u_1, x_2 + 3t_2 + 3\Delta_g^2 u_2) \\ & dm_{U_1}(x_1) dm_{U_2}(x_2) dm_{U_2}(t_2),\end{aligned}$$

where the notation $\approx_{UD}$ is defined immediately following the statement of Lemma 1.9 in the Introduction.

Again, we restrict attention to the model case (4.2) for simplicity. If the $\phi_{i,2}$ all vanish identically, then the claim is trivial as before. Now suppose that both $\phi_{2,2}$ and $\phi_{3,2}$ vanish identically, but that $\phi_{1,2}$ does not. The right-hand side of (4.5) then simplifies to

$$\int_{U_2^2} e(\phi_{0,2}(x_2) + \phi_{1,2}(x_2 + t_2)) dm_{U_2}(x_2) dm_{U_2}(t_2),$$

which vanishes by a change of variables and Fourier analysis. Meanwhile, the left-hand side of (4.5) is $\int_X f_0 T_g f_1 d\mu$; the nonvanishing of $\phi_{1,2}$ guarantees that $\|f_1\|_{U^2(\mathrm{X})} = 0$ by Theorem 3.5, and so by Corollary 3.4, the left-hand side goes to 0 in uniform density, as required.

Now suppose that $\phi_{3,2}$ vanishes identically, but $\phi_{2,2}$ does not. For any $g \in G$ and $x \in X$, we consider the sequence $\psi_{2,g,x} : \mathbb{Z} \to U_2$ defined by $\psi_{2,g,x}(n) := u_2(T_{ng}x)$. As discussed earlier in this section, $\psi_{2,g,x}$ is a quadratic sequence. However, for fixed g, we can also compute the top order coefficient $\Delta_1^2 \psi_{2,g,x}$ of this sequence: $\Delta_1^2 \psi_{2,g,x} = \Delta_g^2 u_2$. Note that as $\psi_{2,g,x}$ and u_2 are both quadratic, the left and right-hand sides here are constants (i.e., elements of U_2). Thus, $\psi_{2,g,x}$ is not an arbitrary quadratic sequence, but is in fact the sum of the sequence $n \mapsto \binom{n}{2} \Delta_g^2 u_2$ and a linear sequence. Using the Lagrange interpolation formula $\psi_{2,g,x}(2) = -\psi_{2,g,x}(0) + 2\psi_{2,g,x}(1) + \Delta_g^2 u_2$, allows us to rewrite the left-hand side of (4.5) as $e(\Delta_g^2 u_2) \int_X f_0' T_g f_1' d\mu$, where $f_i' := e(\phi_{i,2}')$ for $i = 1, 2$, $\phi_{0,2}' := \phi_{0,2} - \phi_{2,2}$, and $\phi_{1,2}' := \phi_{1,2} + 2\phi_{2,2}$. Similarly, using the identity

$$x_2 + 2t_2 + \Delta_g^2 u_2 = -x_2 + 2(x_2 + t_2) + \Delta_g^2 u_2,$$

we may rewrite the right-hand side of (4.5) as

$$e(\Delta_g^2 u_2)\int_{U_1\times U_2^2} f_0'(x_1,x_2)f_1'(x_1+\Delta_g u_1, x_2+t_2)dm_{U_1}(x_1)dm_{U_2}(x_2)dm_{U_2}(t_2).$$

From the previously handled cases of (4.2), we already know that

$$\int_X f_0' T_g f_1' d\mu \approx_{UD} \int_{U_1\times U_2^2} f_0'(x_1,x_2)f_1'(x_1+\Delta_g u_1, x_2+t_2)\, dm_{U_1}(x_1)dm_{U_2}(x_2)dm_{U_2}(t_2).$$

Multiplying through by the phase $e(\Delta_g^2 u_2)$, we obtain (4.5) in the case that $\phi_{3,2}$ vanishes, but $\phi_{2,2}$ does not necessarily vanish. A similar calculation (which we omit) then allows us to extend the previous cases to cover the case when $\phi_{3,2}$ does not necessarily vanish either, giving (4.5) in all instances of the model case (4.2). Again, the addition of the lower order terms in (4.3) can be handled by a modification of these arguments, which are special cases of the argument in Section 6 below and which we leave to the reader.

5 Proof of the first limit formula

In this section, we prove Lemma 1.8. Let $p, G, \mathrm{X}, k, c_0, \ldots, c_k, (\Phi_n)_{n=1}^{\infty}$ be as in that lemma. If $(\pi_{k-1})_* f_i = 0$ for some $i = 0, \ldots, k$, then the claim is immediate from Theorem 1.6. By linearity, we may thus reduce to the case in which each f_i is a pullback by $(\pi_{k-1})^*$ from the associated function $\tilde{f}_i := (\pi_{k-1})_* f_i$. Our task is to show that for any $f_0, \ldots, f_k \in L^\infty(\mathrm{X})$,

$$\mathbb{E}_{g\in\Phi_n}\int_X f_0(T_{c_0 g}x)\ldots f_k(T_{c_k g}x)d\mu(x) \tag{5.1}$$

converges as $n \to \infty$ to the integral

$$\int_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})} \tilde{f}_0 \otimes \ldots \otimes \tilde{f}_k dm_{HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})}. \tag{5.2}$$

As noted after the statement of Lemma 1.8, $HP_{c_0,\ldots,c_k}(\mathrm{Z}_{k-1})$ surjects onto each of the $k+1$ coordinates of $(\mathrm{Z}_{k-1})^{k+1}$. Hence, we can bound (5.2) in magnitude by $\|f_i\|_{L^2(\mathrm{X})}$ for any $0 \le i \le k$, if we normalize so that $\|f_j\|_{L^\infty(\mathrm{X})} \le 1$ for $j \neq i$. Of course, a similar bound can also be obtained for (5.1). By combining these observations with Fourier analysis on the compact abelian group $U_1 \times \cdots \times U_k$ and a limiting[10] argument, it suffices to verify this claim under the assumption that

[10]Here we use the basic fact that an L^∞ function on a compact abelian group can be approximated to arbitrary accuracy in L^2 norm by a finite linear combination of multiplicative characters, while still staying uniformly bounded in L^∞. This can be established for instance by first approximating the function by a continuous function, then using the Stone-Weierstrass theorem.

each $\tilde{f}_i$, $i = 0, \dots, k$, is a tensor product of multiplicative characters,

$$\tilde{f}_i(u_1, \dots, u_{k-1}) = \prod_{j=1}^{k-1} e(\phi_{ij}(u_j))$$

for all $u_1 \in U_1, \dots, u_{k-1} \in U_{k-1}$ and some additive characters (i.e, continuous homomorphisms) $\phi_{ij} : U_j \to \mathbb{R}/\mathbb{Z}$ for $i = 0, \dots, k$ and $j = 1, \dots, k-1$. The expression (5.1) is then equal to

$$\mathbb{E}_{g \in \Phi_n} \int_X e\Big(\sum_{i=0}^{k} \sum_{j=1}^{k-1} \phi_{ij} \left(\psi_{j,g,x}(c_i) \right) \Big) d\mu(x), \tag{5.3}$$

where $\psi_{j,g,x} : \mathbb{Z} \to U_j$ is the (periodic) sequence

$$\psi_{j,g,x}(n) := u_j \left(\pi_{k-1}(T_{ng}x) \right) \tag{5.4}$$

and $u_j : Z_{k-1} \to U_j$, $j = 1, \dots, k-1$, are the coordinate functions. Also, by Fourier analysis, the expression (5.2) is equal to 1 when we have the identities

$$\sum_{i=0}^{k} \phi_{ij} \left(P_j(c_i) \right) = 0 \tag{5.5}$$

for all $j = 1, \dots, k-1$ and all polynomials $P_j : \mathbb{Z} \to U_j$ of degree less than $j+1$, and 0 otherwise.

The strategy is to use the polynomial structure of the Weyl system to place the additive characters ϕ_{ij} in a "normal form", at which point the convergence can be deduced from Lemma 3.2. This is achieved as follows. From construction of the Weyl system we have $\Delta_g u_j = \rho_j(g, \cdot)$ for all $g \in G$ and $j = 1, \dots, k-1$. Since ρ_j is a polynomial cocycle of degree less than j, we conclude that $u_j : Z_{k-1} \to U_j$ is a polynomial of degree less than $j+1$. This implies that for any $x \in X$, the sequence $\psi_{j,g,x}$ defined in (5.4) is a polynomial sequence of degree less than $j+1$, and thus has a Taylor expansion of the form

$$u_j(\pi_{k-1}(T_{ng}x)) = \sum_{l=0}^{j} \binom{n}{l} a_{j,g,x}$$

for some coefficients $a_{j,g,x} \in U_j$. As the $c_0, \dots, c_j$ are distinct elements of $\mathbb{F}_p$, we may then use Lagrange interpolation (using the p-torsion nature of U_j and the hypothesis $j < p$ to justify any division occurring in the interpolation formula). This implies that one can express $\psi_{j,g,x}(n)$ as a linear combination

$$\psi_{j,g,x}(n) = \sum_{i=0}^{j} b_{n,j,i} \psi_{j,g,x}(c_i) \tag{5.6}$$

for some integer coefficients $b_{n,j,i}$ that do not depend on g or x (but may depend on p and $c_0, \dots, c_j$). Indeed, the interpolation formula gives the more general identity $P_j(n) = \sum_{i=0}^{j} b_{n,j,i} P_j(c_i)$ for any polynomial $P_j : \mathbb{Z} \to U_j$ of degree less than $j+1$, with the same coefficients $b_{n,j,i}$. In particular, for any $j < i \leq k$, we can rewrite $\phi_{ij}(\psi_{j,g,x}(c_i))$ in (5.3) as a linear combination of the $\phi_{ij}(\psi_{j,g,x}(c_0)), \dots, \phi_{ij}(\psi_{j,g,x}(c_j))$, and similarly write $\phi_{ij}(P_j(c_i))$ in (5.5) as the same linear combination of the $\phi_{ij}(P_j(c_0)), \dots, \phi_{ij}(P_j(c_j))$. From this fact, we see that if the additive character ϕ_{ij} is not identically 0 for some $j < i \leq k$, we may delete that character (and adjust the characters $\phi_{0j}, \dots, \phi_{jj}$ by appropriate multiples of the deleted character) without affecting either (5.3) or (5.5). Using this observation repeatedly, we see that to prove the convergence of (5.3) to 1 when (5.5) holds and 0 otherwise, it suffices to do so under the normalization that $\phi_{ij} = 0$ for all $i > j$, which we now assume henceforth.

We now divide the argument into two cases. If the ϕ_{ij} are all identically 0, the claim is trivial. Otherwise, we may find $1 \leq j_* \leq k$ such that ϕ_{ij} all vanish for $j > j_*$, but $\phi_{i_* j_*}$ is not identically 0 for at least one $0 \leq i_* \leq j_*$. By permuting the i indices, and then readjusting the ϕ_{ij} characters for $j < j_*$ as before, we may assume without loss of generality that $i_* = j_*$.

Observe from Lagrange interpolation that if $P_{j_*} : \mathbb{Z} \to U_{j_*}$ is an arbitrary polynomial sequence of degree less than $j_* + 1$, then the tuple $(P_{j_*}(c_0), \dots, P_{j_*}(c_{j_*}))$ can take arbitrary values in $U_{j_*}^{j_*+1}$; in particular, as $\phi_{j_* j_*}$ is not identically 0, the identity (5.5) does *not* hold for $j = j_*$. Thus, the expression (5.2) vanishes in this case; and our task is now to show that (5.3) converges to 0. But from the vanishing of ϕ_{ij} when $i > j$ or $j > j_*$, we can write (5.3) in the form

$$\mathbb{E}_{g \in \Phi_n} \int_X \left(\prod_{i=0}^{j_*-1} F_i(T_{c_i g} x) \right) e\left(\phi_{j_* j_*}\left(u_{j_*}(\pi_{k-1}(T_{j_* g} x)) \right) \right) d\mu(x)$$

for some functions $F_0, \dots, F_{j_*-1} \in L^\infty(\mathrm{X})$ of unit magnitude which do not depend on g or x, and whose exact form is of no importance to us. Applying Lemma 3.2, we conclude that

$$\limsup_{n \to \infty} \left| \mathbb{E}_{g \in \Phi_n} \int_X e\left(\sum_{i=0}^{k} \sum_{j=1}^{m} \phi_{ij}\left(\psi_{j,g,x}(c_i) \right) \right) d\mu(x) \right| \leq \| e(\phi_{j_* j_*}(u_{j_*}(\pi_{k-1}))) \|_{U^{j_*}(\mathrm{X})}.$$

However, as the character $\phi_{j_* j_*}$ is not identically 0, the function $e(\phi_{j_* j_*}(u_{j_*})$ has mean 0 on every coset of U_{j_*} in $U_1 \times \dots \times U_{k-1}$, and thus

$$(\pi_{j_*-1})_* \left(e\left(\phi_{j_* j_*}\left(u_{j_*}(\pi_{k-1}) \right) \right) \right) = 0.$$

By Theorem 3.5, we conclude that $\left\| e\left(\phi_{j_* j_*}\left(u_{j_*}(\pi_{k-1}) \right) \right) \right\|_{U^{j_*}(\mathrm{X})} = 0$, giving the desired convergence of (5.3) to 0. This concludes the proof of Lemma 1.8.

Remark. The argument above gives a new proof of the convergence of the averages

$$\mathbb{E}_{g\in\Phi_n}\int_X (T_{c_0 g} f_0)\dots(T_{c_k g} f_k)d\mu$$

as $n\to\infty$ for arbitrary $k\geq 0$, $f_0,\dots,f_k\in L^\infty(\mathrm{X})$, and $c_0,\dots,c_k\in\mathbb{F}_p$ (since, after collecting like terms, we can reduce to the case where the $c_0,\dots,c_k\in\mathbb{F}_p$ are distinct, so that $k<p$). A modification of the argument also shows convergence in $L^2(\mathrm{X})$ of the averages

$$\mathbb{E}_{g\in\Phi_n}(T_{c_1 g} f_1)\dots(T_{c_k g} f_k) \tag{5.7}$$

for arbitrary $k\geq 0$, $c_1,\dots,c_k\in\mathbb{F}_p$ and $f_1,\dots,f_k\in L^\infty(\mathrm{X})$. We sketch the argument as follows. First, collecting like terms and factoring out those terms with $c_i=0$, we may assume that the $c_1,\dots,c_k$ are distinct and nonzero, so that $k<p$. By Theorem 1.6 (as in the proof of Lemma 1.8), we may assume that each f_i has the form $f_i=\pi^*_{k-1}\tilde{f}_i$ for some $\tilde{f}_i\in L^\infty(\mathrm{Z}_{k-1})$, and then use Fourier decomposition as before to assume that each $\tilde{f}_i$ is the tensor product of characters $e(\phi_{ij})$. We can then use identities of the form (5.6) (setting $c_0:=0$) to reduce to the case where the ϕ_{ij} vanish for $i>j$ and then adapt the preceding argument to show that the average (5.7) either is identically 1 or converges in norm to 0. We leave the details to the interested reader. We remark that the limit value of (5.7) does not depend on the Følner sequence (Φ_n).

6 Proof of the second limit formula

We now prove of Lemma 1.9, using a minor variant of the argument used to prove Lemma 1.8.

Let $p, G, \mathrm{X}, k, c_0,\dots,c_k, (\Phi_n)_{n=1}^\infty$ be as in Lemma 1.8. Using Theorem 1.6 as in the previous section (but with Z_k as the characteristic factor, instead of Z_{k-1}), we reduce to the case in which each f_i is a pullback by $(\pi_k)^*$ from the associated function $\tilde{f}_i:=(\pi_k)_*f_i$. Our task is to show that for any $f_0,\dots,f_k\in L^\infty(\mathrm{X})$,

$$|I_{c_0,\dots,c_k;f_0,\dots,f_k}(g)-J_{c_0,\dots,c_k;f_0,\dots,f_k}(g)| \tag{6.1}$$

converges in uniform density to 0.

Observe that the closed group $HP_{c_0,\dots,c_k}(\mathrm{Z}_k)_0$ contains the diagonal group $Z_k^\Delta:=\{(z,\dots,z):z\in Z_k\}\subset Z_k^{k+1}$ and thus surjects onto each factor Z_k. Translating, we see that the cosets $HP_{c_0,\dots,c_k}(\mathrm{Z}_k)_g$ also surject onto each factor Z_k. We can then repeat the limiting argument from the previous section and reduce to the case in which each $\tilde{f}_i$, $i = 0, \ldots, k$, is a tensor product of characters, viz.,

$$\tilde{f}_i(u_1, \ldots, u_k) = \prod_{j=1}^{k} e(\phi_{ij}(u_j))$$

for all $u_1 \in U_1, \ldots, u_k \in U_k$ and some characters (i.e continuous homomorphisms) $\phi_{ij} : U_j \to \mathbb{R}/\mathbb{Z}$ for $i = 0, \ldots, k$ and $j = 1, \ldots, k$. For $g \in G$, $I_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g)$ is then equal to

$$\int_X e\Big(\sum_{i=0}^{k}\sum_{j=1}^{k} \phi_{ij}(\psi_{j,g,x}(c_i))\Big)\,d\mu(x), \tag{6.2}$$

where

$$\psi_{j,g,x}(n) := u_j(\pi_k(T_{ng}x)) \tag{6.3}$$

and $u_j : Z_k \to U_j$, $j = 1, \ldots, k$ are the coordinate functions. Meanwhile, the value of $J_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g)$ depends on the behavior of the quantities

$$\sum_{i=0}^{k} \phi_{ij}(P_j(c_i)) \tag{6.4}$$

for $j = 1, \ldots, k$, where P_j ranges over all polynomials $P_j : \mathbb{Z} \to U_j$ of degree less than $j+1$ with leading coefficient $\Delta_1^j P_j = \Delta_g^j u_j$. If, for each j, the expression in (6.4) is equal to a constant $\theta_{j,g} \in \mathbb{R}/\mathbb{Z}$ independent of P_j, then $J_{c_0,\ldots,c_k;f_1,\ldots,f_k}(g)$ is equal to $e(\sum_{j=1}^{k} \theta_{j,g})$. In all other cases, $J_{c_0,\ldots,c_k;f_1,\ldots,f_k}(g)$ vanishes.

As in the previous section, we use the polynomial structure of the Weyl system to put the characters ϕ_{ij} in "normal form". Fix $g \in G$. As before, for each $1 \le j \le k$ and $x \in$ X, the sequence $\psi_{j,g,x} : \mathbb{Z} \to U_j$ is a polynomial sequence of degree less than $j + 1$ from $\mathbb{Z}$ to U_j. However, since g is now fixed, we see from (6.3) that there is an additional constraint on the top order coefficient of $\psi_{j,g,x}$:

$$\Delta_1^j \psi_{j,g,x} = \Delta_g^j u_j. \tag{6.5}$$

This additional (g-dependent) constraint on $\phi_{j,g,x}$ allows us to eliminate one further character ϕ_{ij} than was possible in the previous section. Indeed, from (6.5) we see that the modified sequence $n \mapsto \psi_{j,g,x}(n) - \binom{n}{j}\Delta_g^j u_j$ is now a polynomial sequence of degree less than j rather than $j + 1$. Applying Lagrange interpolation to this polynomial of one degree lower and then rewriting everything in terms of $\psi_{j,g,x}$, we obtain identities of the form

$$\psi_{j,g,x}(n) = \sum_{i=0}^{j-1} b'_{n,j,i,g}\psi_{j,g,x}(c_i) + a'_{n,j,i}$$

for all $n \in \mathbb{Z}$ and coefficients $b'_{n,j,i,g}, a'_{n,j,g} \in \mathbb{F}_p$ that do not depend on x. Furthermore, we have the same identity $P_j(n) = \sum_{i=0}^{j-1} b'_{n,j,i,g} P_j(c_i) + a'_{n,j,i}$ for any polynomial $P_j : \mathbb{Z} \to U_j$ of degree less than $j+1$ obeying the constraint $\Delta_1^j P_j = \Delta_g^j u_j$.

It follows from these identities that whenever ϕ_{ij} is not identically 0 for some $1 \leq j \leq i \leq k$, we can rewrite $\phi_{ij}(\psi_{j,g,x}(c_i))$ as a linear combination of $\phi_{ij}(\psi_{j,g,x}(c_0)), \ldots, \phi_{ij}(\psi_{j,g,x}(c_{j-1}))$ plus a constant independent of x, and, similarly, we can rewrite $\phi_{ij}(P_j(c_i))$ in (6.4) as the same linear combination of $\phi_{ij}(P_j(c_0)), \ldots, \phi_{ij}(P_j(c_{j-1}))$ plus the same constant. Because of this, we can delete ϕ_{ij} (and adjust the characters $\phi_{0j}, \ldots, \phi_{j-1,j}$ by appropriate multiples of the deleted character), resulting in $I_{c_0,\ldots,c_k;f_1,\ldots,f_k}(g)$ and $J_{c_0,\ldots,c_k;f_1,\ldots,f_k}(g)$ being rotated by the same (g-dependent) phase shift. In particular, (6.1) remains unchanged by these modifications of the characters ϕ_{ij}. Arguing as in the previous section, we can thus reduce to the case in which the ϕ_{ij} vanish for all $j \leq i \leq k$ (note that this is a slightly stronger vanishing condition than that of the previous section, where ϕ_{ij} vanishes only for $j < i \leq k$).

As in the preceding section, we now consider two cases. If the ϕ_{ij} are all identically 0, the claim is trivial. Otherwise, we can find $1 \leq j_* \leq k$ such that ϕ_{ij} all vanish for $j > j_*$, but $\phi_{i_* j_*}$ is not identically 0 for at least one $0 \leq i_* \leq j_* - 1$. By permuting the i indices, and then readjusting the ϕ_{ij} characters for $j < j_*$ as before, we may assume without loss of generality that $i_* = j_* - 1$.

Fix $g \in G$. Observe from Lagrange interpolation that if $P_{j_*} : \mathbb{Z} \to U_{j_*}$ is a polynomial sequence of degree less than $j_* + 1$ that is arbitrary save for obeying the constraint $\Delta_1^{j_*} P_{j_*} = \Delta_g^{j_*} u_{j_*}$, then the sequence $n \mapsto P_{j_*}(n) - \binom{n}{j_*} \Delta_g^{j_*} u_{j_*}$ is an arbitrary polynomial of degree less than j_*. In particular, the tuple $(P_{j_*}(c_0), \ldots, P_{j_*}(c_{j_*-1}))$ can take arbitrary values in $U_{j_*}^{j_*}$. Thus, as ϕ_{j_*-1,j_*} is not identically 0, the identity (5.5) does *not* hold for $j = j_*$; and so $J_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g)$ vanishes for all $g \in G$. Our task is now to show that $I_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g)$ converges in uniform density to 0. But from the vanishing of ϕ_{ij} when $i \geq j$ or $j > j_*$, we can write (5.3) in the form

$$\mathbb{E}_{g \in \Phi_n} \int_X \left(\prod_{i=0}^{j_*-2} F_i(T_{c_i g} x) \right) e\left(\phi_{j_*-1,j_*}\left(u_{j_*}\left(\pi_{k-1}(T_{c_{j_*-1} g} x)\right)\right)\right) d\mu(x)$$

for functions $F_0, \ldots, F_{j_*-2} \in L^\infty(\mathrm{X})$ of unit magnitude which do not depend on g or x. Arguing as in the previous section, we have

$$\left\| e\left(\phi_{j_*-1,j_*}\left(u_{j_*}(\pi_{k-1})\right)\right) \right\|_{U^{j_*}(\mathrm{X})} = 0;$$

and the claim now follows from Lemma 3.2.

7 Proof of structure theorem

We now prove Proposition 1.10. Let the notation be as in Lemma 1.9. Observe that the coset $HP_{c_0,\dots,c_k}(\mathrm{Z}_k)_g$ depends on g only through the quantities $\Delta_g^j u_j \in U_j$ for $j = 1, \dots, k$. Furthermore, the dependence of the integral

$$J_{c_0,\dots,c_k;f_0,\dots,f_k}(g) = \int_{HP_{c_0,\dots,c_k}(\mathrm{Z}_{k-1})} (\pi_k)_* f_0 \otimes \cdots \otimes (\pi_k)_* f_k dm_{HP_{c_0,\dots,c_k}(\mathrm{Z}_k)_{\theta(g)}}$$

on $\theta(g)$ is continuous (this is easiest to see by first approximating each $(\pi_i)_* f_i$ in L^2 norm with a continuous function on the compact group Z_k). Therefore, it suffices to represent each sequence $\theta_j : g \mapsto \Delta_g^j u_j$ for $j = 1, \dots, k$ in the form $\theta_j(g) = \Delta_g^j u_j = F_j(S_{j,g} y_j)$ for some continuous j-step Weyl system $\mathrm{Y}_j = (Y_j, \mathcal{Y}_j, \nu_j, (S_{j,g})_{g \in G})$, some continuous function $F_j : Y_j \to U_j$, and some point $y_j \in Y$. The claim then follows by composing the various continuous maps and noting that the product of finitely many continuous Weyl systems of step at most k is of step at most k.

For each j, we introduce the form $\Lambda_j : G^j \to U_j$ by $\Lambda_j(g_1, \dots, g_j) := \Delta_{g_1} \dots \Delta_{g_j} u_j$ (again, note that the right-hand side is a constant function and so can be identified with an element of U_j). From the identities $\Delta_g \Delta_h = \Delta_h \Delta_g$ and $\Delta_{g+h} = \Delta_g + \Delta_h + \Delta_g \Delta_h$ (and noting that any $j+1$-fold derivative of u_j vanishes), we see that Λ_j is a symmetric multilinear form. Our task is to establish a representation of the form

$$\Lambda_j(g, \dots, g) = F_j(S_{j,g} y_j) \tag{7.1}$$

for some continuous k-step Weyl system $\mathrm{Y}_j = (Y_j, \mathcal{Y}_j, \nu_j, (S_{j,g})_{g \in G})$, some continuous function $F_j : Y_j \to U_j$, and some point $y_j \in Y$.

Fix j. To achieve the above goal, we exploit a dynamical abstraction of the algebraic observation (essentially the binomial formula) that if we define

$$\Gamma_i(x)(h_1, \dots, h_{j-i}) := \Lambda_j(h_1, \dots, h_{j-i}, x^{(i)})$$

for $0 \le i \le j$ and $x, h_1, \dots, h_{j-i} \in G$, where $x^{(i)}$ denotes i copies of x, (so, in particular, $\Gamma_0(x) = \Lambda_j$), then the $\Gamma_i(x) : G^{j-i} \to U_j$ are symmetric multilinear forms (where the multilinearity is, of course, with respect to the field $\mathbb{F}_p$), and we have the shift identity

$$\Gamma_i(x+g)(h_1, \dots, h_{j-i}) = \sum_{l=0}^{i} \binom{l}{i} \Gamma_l(x)(h_1, \dots, h_{j-i}, g^{(i-l)})$$

for all $0 \le i \le j$ and $x, g, h_1, \dots, h_{j-i} \in G$.

Now we give the dynamical version of the above identity. For each $1 \leq i \leq j$, let V_i be the collection of all symmetric multilinear forms $\Gamma_i : G^{j-i} \to U_j$, where the multilinearity is with respect to the field $\mathbb{F}_p$. Then V_i can be viewed as a closed subgroup of the product space $U_i^{G^{j-i}}$ and is thus a compact abelian group. Set $Y_j := V_1 \times \ldots \times V_j$ with the product σ-algebra $\mathcal{Y}_j$ and Haar measure ν_j. Define the shift maps $S_{j,g}(\Gamma_1, \ldots, \Gamma_j) = (S_{j,g}(\Gamma_1, \ldots, \Gamma_j)_i)_{i=1}^j$ for $g \in G$ and $\Gamma_i \in V_i$ for $i = 1, \ldots, j$ by

$$S_{j,g}(\Gamma_1, \ldots, \Gamma_j)_i(h_1, \ldots, h_{j-i}) = \sum_{l=0}^{i} \binom{i}{l} \Gamma_l(h_1, \ldots, h_{j-i}, g^{(i-l)}) \tag{7.2}$$

with the convention that $\Gamma_0 := \Lambda_j$. We verify that this is an action:

$$\begin{aligned}
&S_{j,g'}(S_{j,g}(\Gamma_1, \ldots, \Gamma_j))_i(h_1, \ldots, h_{j-i}) \\
&\qquad = \sum_{l=0}^{i} \binom{i}{l} S_{j,g}(\Gamma_1, \ldots, \Gamma_j)_l(h_1, \ldots, h_{j-i}, (g')^{(i-l)}) \\
&\qquad = \sum_{l=0}^{i} \sum_{m=0}^{l} \binom{i}{l} \binom{l}{m} \Gamma_m(h_1, \ldots, h_{j-i}, (g')^{(i-l)}, g^{(l-m)}) \\
&\qquad = \sum_{m=0}^{i} \binom{i}{m} \sum_{a=0}^{i-m} \binom{i-m}{a} \Gamma_m(h_1, \ldots, h_{j-i}, (g')^{(a)}, g^{(i-m-a)}) \\
&\qquad = \sum_{m=0}^{i} \binom{i}{m} \Gamma_m(h_1, \ldots, h_{j-i}, (g+g')^{(i-m)}) \\
&\qquad = S_{j,g+g'}(\Gamma_1, \ldots, \Gamma_j)_i(h_1, \ldots, h_{j-i}),
\end{aligned}$$

where the penultimate equation follows from the symmetry and multilinearity of Γ_m, and we have implicitly used the fact that the formula (7.2) extends to the case $i = 0$ with the convention that $S_{j,g}(\Gamma_1, \ldots, \Gamma_j)_0$ and Γ_0 are both equal to Λ_j.

One easily verifies by induction that $\mathrm{Y}_j = (Y_j, \mathcal{Y}_j, \nu_j, (S_{j,g})_{g \in G})$ is a tower

$$\mathrm{Y}_j = V_1 \times_{\eta_2} V_2 \times_{\eta_3} \cdots \times_{\eta_j} V_j$$

of cocycle extensions, where the $(G, V_1 \times_{\eta_2} \cdots \times_{\eta_{i-1}} V_{i-1}, V_i)$-cocycle η_i is defined for $1 \leq i \leq j$ by the formula

$$\eta_i(g, (\Gamma_1, \ldots, \Gamma_{i-1}))(h_1, \ldots, h_{j-i}) = \sum_{l=0}^{i-1} \binom{i}{l} \Gamma_l(h_1, \ldots, h_{j-i}, g^{(i-l)}). \tag{7.3}$$

For each $1 \leq i \leq j$, we see from (7.2) that differentiation of the coordinate function $\upsilon_i : (\Gamma_1, \ldots, \Gamma_l) \mapsto \Gamma_i$ in some direction j yields an affine-linear combination of the previous coordinate functions $v_1, \ldots, v_{i-1}$. In particular, this implies that each coordinate function v_i is a polynomial of degree less than $i+1$, which implies from (7.3) that each cocycle η_i is a polynomial of degree less than i. Thus Y_j is a j-step Weyl system; an easy induction then shows that it is in fact a continuous j-step Weyl system.

From (7.2) we have $v_j(S_{j,g}(0, \ldots, 0)) = \Lambda_j(g, \ldots, g)$; as $v_j : Y_j \to V_j \equiv U_j$ is clearly a continuous function, we obtain the desired representation (7.1).

8 Khintchine for double recurrence

We now prove Theorem 1.12. Suppose for contradiction that the claim fails. Then we can find $p, \mathrm{X}, A, \varepsilon$ as in Definition 1.11 such that the set

$$\{\mu(T_{-c_0 g}A \cap T_{-c_1 g}A \cap T_{-c_2 g}A) \geq \mu(A)^3 - \varepsilon\}$$

fails to be syndetic. In particular, the complement of this set contains translates of any given finite set, and in particular must contain a Følner sequence $(\Phi_n)_{n=1}^\infty$. Thus we have $\mu(T_{-c_0 g}A \cap T_{-c_1 g}A \cap T_{-c_2 g}A) < \mu(A)^3 - \varepsilon$ for all $n = 1, 2, \ldots$ and $g \in \Phi_n$. We can rewrite this as

$$\int_X (T_{c_0 g}1_A)(T_{c_1 g}1_A)(T_{c_2 g}1_A)d\mu < \mu(A)^3 - \varepsilon, \tag{8.1}$$

where 1_A is the indicator function of A. If we could show that

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X (T_{c_0 g}1_A)(T_{c_1 g}1_A)(T_{c_2 g}1_A)d\mu > \mu(A)^3 - \varepsilon,$$

then we would obtain the desired contradiction. However, a direct application of Lemma 1.8 computes the left-hand side as

$$\int_{U_1}\int_{U_1} f(x+c_0 t)f(x+c_1 t)f(x+c_2 t)dm_{U_1}(x)dm_{U_1}(t),$$

where $U_1 = Z_1$ is the Kronecker factor and $f := \pi_1 1_A$; and it is possible[11] for this integral to be significantly smaller than $\mu(A)^3 = (\int_{U_1} f dm_{U_1})^3$. Fortunately, we can get around this difficulty by the following trick of Frantzikinakis [11] (see also [4]). Observe from Hölder's inequality that

$$\int_{U_1} f(x)f(x)f(x)dm_{U_1}(x) \geq (\int_{U_1} f dm_{U_1})^3 = \mu(A)^3.$$

[11] An explicit example of this phenomenon can be constructed, for large p at least, by adapting the Behrend construction [2], similar to the construction in [4, Section 2.1], which handled the case $c_i = i$ in which U_1 was replaced by $\mathbb{R}/\mathbb{Z}$ and f replaced by an indicator function 1_B; we omit the details.

As translations are continuous in (say) the L^2 norm, we conclude that

$$\int_{U_1} f(x+c_0t)f(x+c_1t)f(x+c_2t)dm_{U_1}(x) \geq \mu(A)^3 - \varepsilon/2$$

(say) for all $t \in U_1$ sufficiently close to the origin. In particular, by Urysohn's Lemma, we can find a nonnegative continuous function $\eta : U_1 \to \mathbb{R}^+$ with $\int_{U_1} \eta dm_{U_1} = 1$ such that

$$\int_{U_1}\int_{U_1} \eta(t)f(x+c_0t)f(x+c_1t)f(x+c_2t)dm_{U_1}(x)dm_{U_1}(t) \geq \mu(A)^3 - \varepsilon/2. \tag{8.2}$$

We now claim the weighted limit formula

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n}\eta(\rho_1(g)) \int_X (T_{c_0g}1_A)(T_{c_1g}1_A)(T_{c_2g}1_A)d\mu \\ = \int_{U_1}\int_{U_1} \eta(t)f(x+c_0t)f(x+c_1t)f(x+c_2t)dm_{U_1}(x)dm_{U_1}(t),\end{aligned} \tag{8.3}$$

which (on replacing A by all of X) gives $\lim_{n\to\infty}\mathbb{E}_{g\in\Phi_n}\eta(\rho_1(g)) = 1$ (this can also be established from the unique ergodicity of the Kronecker factor); and this gives a contradiction between (8.1) and (8.2).

It remains to establish (8.3). We prove this formula for arbitrary continuous functions $\eta : U_1 \to \mathbb{C}$. By the Stone-Weierstrass Theorem and Fourier analysis, it suffices to establish the claim when η is a multiplicative character, viz., $\eta = e(\phi)$ for some continuous homomorphism $\phi : U_1 \to \mathbb{R}/\mathbb{Z}$. But as c_0, c_1, c_2 are distinct elements of $\mathbb{F}_p$, there is a Lagrange interpolation identity of the form

$$t = a_0(x+c_0t) + a_1(x+c_1t) + a_2(x+c_2t)$$

for integers a_0, a_1, a_2 depending only on c_0, c_1, c_2. Thus the right-hand side of (8.3) can be rewritten as

$$\int_{U_1}\int_{U_1} f_0(x+c_0t)f_1(x+c_1t)f_2(x+c_2t)dm_{U_1}(x)dm_{U_1}(t),$$

where $f_i(x) := f(x)e(a_i\phi(x))$ for $i = 0, 1, 2$. Since the shift $(T_{1,g})_{g\in G}$ on the Kronecker system $Z_1 = U_1$ is given by $T_{1,g} : x \mapsto x+\rho_1(g)$ for each group element $g \in G$, we can write $\rho_1(g) = a_0T_{1,c_0g}(x)+a_1T_{1,c_1g}(x)+a_2T_{1,c_2g}(x)$ for all $g \in G$ and $x \in U_1$, which implies that

$$\eta(\rho_1(g)) \int_X (T_{c_0g}1_A)(T_{c_1g}1_A)(T_{c_2g}1_A)d\mu = \int_X (T_{c_0g}\tilde{f}_0)(T_{c_1g}\tilde{f}_1)(T_{c_2g}\tilde{f}_2)d\mu$$

for any $g \in G$, where $\tilde{f}_i(x) := 1_A(x)e(a_i\phi(\pi_1(x)))$ for $i = 0, 1, 2$ and $x \in X$. Since $f_i = (\pi_1)_*\tilde{f}_i$ for $i = 0, 1, 2$, the claim (8.3) now follows from Lemma 1.8.

9 Khintchine for triple recurrence

We now give the proof of Theorem 1.13, which follows the lines of Theorem 1.12 (and is, of course, also similar to the proof of the analogous claim for $\mathbb{Z}$-systems in [4]). We may permute indices so that $c_0 + c_3 = c_1 + c_2$. By translating, we may normalize $c_0 = 0$, and then by dilating normalize $c_1 = 1$, so that $c_3 = c_2 + 1$.

Again, we assume for contradiction that the claim fails. Then, arguing as before, we can find p, X, A, ε verifying the hypotheses of the theorem and a Følner sequence $(\Phi_n)_{n=1}^\infty$ such that

$$\int_X 1_A(T_g 1_A)(T_{c_2 g} 1_A)(T_{(c_2+1)g} 1_A)d\mu < \mu(A)^4 - \varepsilon$$

for all $n = 1, 2, \ldots$ and $g \in \Phi_n$. Arguing as before, we see that it suffices to locate a nonnegative continuous function $\eta : U_1 \to \mathbb{R}^+$ with $\int_{U_1} \eta dm_{U_1} = 1$ such that

$$\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \eta(\rho_1(g)) \int_X 1_A(T_g 1_A)(T_{c_2 g} 1_A)(T_{(c_2+1)g} 1_A)d\mu \geq \mu(A)^4 - \varepsilon/2.$$

A direct application of Lemma 1.8 gives

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \int_X 1_A(T_g 1_A)(T_{c_2 g} 1_A)(T_{(c_2+1)g} 1_A)d\mu \\
&= \int_{U_1^2\times U_2^3} f(x_1, x_2) f(x_1 + t_1, x_2 + t_2) f\left(x_1 + c_2 t_1, x_2 + c_2 t_2 + \binom{c_2}{2} u_2\right) \\
&\qquad f\left(x_1 + (c_2+1)t_1, x_2 + (c_2+1)t_2 + \binom{c_2+1}{2} u_2\right) \\
&\qquad\qquad dm_{U_1}(x_1) dm_{U_1}(t_1) dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_2}(u_2),
\end{aligned}$$

where we now set $f := (\pi_2)_* 1_A$. We can twist this identity by characters as in the previous section to conclude the weighted generalization

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}_{g\in\Phi_n} \eta(\rho_1(g)) \int_X 1_A(T_g 1_A)(T_{c_2 g} 1_A)(T_{(c_2+1)g} 1_A)d\mu \\
&= \int_{U_1^2\times U_2^3} \eta(t_1) \quad f(x_1, x_2) f(x_1 + t_1, x_2 + t_2) f\left(x_1 + c_2 t_1, x_2 + c_2 t_2 + \binom{c_2}{2} u_2\right) \\
&\qquad f\left(x_1 + (c_2+1)t_1, x_2 + (c_2+1)t_2 + \binom{c_2+1}{2} u_2\right) \\
&\qquad\qquad dm_{U_1}(x_1) dm_{U_1}(t_1) dm_{U_2}(x_2) dm_{U_2}(t_2) dm_{U_2}(u_2).
\end{aligned}$$

By Urysohn's Lemma and the Fubini-Tonelli Theorem, it thus suffices to show that

$$\int_{U_1\times U_2^3} f(x_1,x_2)f(x_1+t_1,x_2+t_2)f\left(x_1+c_2t_1,x_2+c_2t_2+\binom{c_2}{2}u_2\right)$$
$$f\left(x_1+(c_2+1)t_1,x_2+(c_2+1)t_2+\binom{c_2+1}{2}u_2\right)$$
$$dm_{U_1}(x_1)dm_{U_2}(x_2)dm_{U_2}(t_2)dm_{U_2}(u_2) \geq \mu(A)^4 - \varepsilon/2$$

for all t_1 sufficiently close to the origin. As before, this expression is continuous in t_1; so it suffices to show that

$$\int_{U_1\times U_2^3} f(x_1,x_2)f(x_1,x_2+t_2)f\left(x_1,x_2+c_2t_2+\binom{c_2}{2}u_2\right)$$
$$f\left(x_1,x_2+(c_2+1)t_2+\binom{c_2+1}{2}u_2\right)$$
$$dm_{U_1}(x_1)dm_{U_2}(x_2)dm_{U_2}(t_2)dm_{U_2}(u_2) \geq \mu(A)^4.$$

From Hölder's inequality and the Fubini-Tonelli theorem,

$$\int_{U_1}\left(\int_{U_2} f(x_1,x_2)dm_{U_2}(x_2)\right)^4 dm_{U_1}(x_1) \geq \left(\int_{U_1\times U_2} f dm_{U_1\times U_2}\right)^4 = \mu(A)^4,$$

so it suffices to prove

$$\int_{U_2^3} F(x_2)F(x_2+t_2)F\left(x_2+c_2t_2+\binom{c_2}{2}u_2\right)F\left(x_2+(c_2+1)t_2+\binom{c_2+1}{2}u_2\right)$$
$$dm_{U_2}(x_2)dm_{U_2}(t_2)dm_{U_2}(u_2) \geq \left(\int_{U_2} F dm_{U_2}\right)^4 \tag{9.1}$$

for any real-valued $F \in L^\infty(U_2)$.

This inequality can be established by Fourier analysis (cf. [4] or [17]), but one can also give a Cauchy-Schwarz-based proof as follows. The starting point is the identity

$$(c_2-1)x_2+(c_2+1)\left(x_2+c_2t_2+\binom{c_2}{2}u_2\right)$$
$$= (c_2-1)\left(x_2+(c_2+1)t_2+\binom{c_2+1}{2}u_2\right)+(c_2+1)(x_2+t_2).$$

From this and some routine computation, we see that for any $x, y, y' \in U_2$, there exists a unique triple (x_2, t_2, u_2) such that

$$\begin{aligned}(c_2-1)x_2 &= y,\\ (c_2+1)(x_2+c_2t_2+\binom{c_2}{2}u_2) &= x-y,\\ (c_2-1)x_2+(c_2+1)t_2+\binom{c_2+1}{2}u_2 &= y',\\ (c_2+1)x_2+t_2 &= x-y;\end{aligned}$$

and so we may rewrite the left-hand side of (9.1) after a linear change of variables as

$$\int_{U_2}\left(\int_{U_2} F\left((c_2-1)^{-1}y\right) F\left((c_2+1)^{-1}(x-y)\right) dm_{U_2}(y)\right)^2 dm_{U_2}(x)$$

(note that c_2-1, c_2+1 are invertible in $\mathbb{F}_p$). By Cauchy-Schwarz, this is greater than or equal to

$$\left(\int_{U_2}\left(\int_{U_2} F\left((c_2-1)^{-1}y\right) F\left((c_2+1)^{-1}(x-y)\right) dm_{U_2}(y)\right) dm_{U_2}(x)\right)^2,$$

which by a further linear change of variables is equal to $(\int_{U_2} F dm_{U_2})^4$, giving (9.1).

10 Counterexamples

We turn to the proof of Theorem 1.15. We begin by passing from sets A to functions f, which are more convenient from the perspective of building counterexamples. More precisely, we use the following “Bernoulli extension” construction.

Theorem 10.1 (Reduction to the function case)**.** *Let p be a prime, and let $c_0, \dots, c_k$ be distinct elements of $\mathbb{F}_p$. Let $G = \mathbb{F}_p^\omega$, and suppose that there exist an ergodic G-system $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$, a nonnegative function $f \in L^\infty(\mathrm{X})$, and $\varepsilon > 0$ such that the set*

$$\left\{g \in G : \int_X (T_{c_0 g} f)\cdots(T_{c_k g} f)d\mu \geq \left(\int_X f d\mu\right)^{k+1} - \varepsilon\right\}$$

is not syndetic. Then $(c_0, \dots, c_k)$ does not have the Khintchine property in characteristic p.

Proof. By rescaling, we may assume that f takes values in $[0, 1]$.

We construct a new system $\mathrm{Y} = (Y, \mathcal{Y}, \mu_Y, (S_g)_{g\in G})$ as follows. The underlying space is $X \times [0, 1]^G$, with the measure μ_Y given by the product of μ_X and the G-fold product of uniform measure on $[0, 1]$, and similarly for the σ-algebra $\mathcal{Y}$. The shift S_g is given by $S_g((x, (t_h)_{h\in G})) := (T_g x, (t_{h+g})_{h\in G})$. One easily verifies that Y is a G-system and that (X, π) is a factor of this system, where $\pi : Y \to X$ is the projection map onto X. Indeed, the system Y is the product of X and a Bernoulli system; as the latter system is weakly mixing, any product of such a system with an ergodic system is ergodic (see, e.g., [14, Proposition 4.5] or [10, Theorem 4.1]). In particular, as X is ergodic, Y is also ergodic.

Taking f as the function in the hypotheses of Theorem 10.1, we now define $A \subset Y$ by $A := \{(x, (t_h)_{h\in G}) : t_0 \le f(x)\}$. This is clearly a measurable set in Y; and from the Fubini-Tonelli Theorem, $\mu_Y(A) = \int_X f d\mu$. A further application of Fubini-Tonelli gives

$$\mu_Y(S_{-c_0 g}A \cap \cdots \cap S_{-c_k g}A) = \int_X (T_{c_0 g} f)\cdots(T_{c_k g} f) d\mu$$

for any $g \in G$, and so we obtain the desired counterexample to $(c_0, \dots, c_k)$ having the Khintchine property. $\square$

To verify the requirements of Theorem 10.1, we use the following "skew shift" construction, which reduces the task of demonstrating failure of the Khintchine property to the harmonic analysis task of finding a counterexample to a certain integral inequality.

Let $\mathbb{T} := \prod \mathbb{F}_p$ be the compact group formed as the product of countably many copies of $\mathbb{F}_p$. Following the notation of Lemma 1.9, we define the set $HP_{c_0,\dots,c_k}(\mathbb{T}^m)_\theta \subset (\mathbb{T}^m)^{k+1}$ for natural numbers $m, k \ge 1$ and elements $c_0, \dots, c_k$ of $\mathbb{F}_p$ to be the collection of all tuples $(P(c_0), \dots, P(c_k))$, where $P : \mathbb{Z} \to \mathbb{T}^m$ can be written in components as $P = (P_1, \dots, P_m)$, and for each $1 \le i \le j$, P_i is a polynomial of degree less than $i + 1$ with $\partial_1^i P_i = \theta_i$.

Theorem 10.2. *Let p be a prime, and let $c_0, \dots, c_k$ be distinct elements of $\mathbb{F}_p$. Suppose that there exist a natural number $1 \le m < p$ and a nonnegative function $f \in L^\infty(\mathbb{T}^m)$ such that*

$$\int_{HP_{c_0,\dots,c_k}(\mathbb{T}^m)_\theta} f \otimes \cdots \otimes f dm_{HP_{c_0,\dots,c_k}(\mathbb{T}^m)_\theta} < \left(\int_{\mathbb{T}^m} f dm_{\mathbb{T}^m}\right)^{k+1}$$

for all $\theta = (\theta_1, \dots, \theta_m) \in \mathbb{T}^m$. Then $(c_0, \dots, c_k)$ does not have the Khintchine property.

Proof. We use an explicit Weyl system, analogous to a skew-shift system on a torus. We identify the group G with the polynomial ring $\mathbb{F}_p[t]$ on one generator t, by identifying each generator e_n of G with t^{n-1}. We then embed $\mathbb{F}_p[t]$ in the field $\mathbb{F}_p[t]\big((1/t)\big)$ of half-infinite Laurent series $\sum_{n=-\infty}^{d} c_n t^n$, which is a locally compact space using the norm $\|\sum_{n=-\infty}^{d} c_n t^n\| := p^d$ when $c_d \neq 0$ (and with $\|0\| = 0$, of course). The quotient space $\mathbb{T} := \sum_{n=-\infty}^{d} c_n t^n / \mathbb{F}_p[t]$ is then a compact abelian group which can be identified with $\mathbb{F}_p^\infty$ and has a Haar probability measure $dm_{\mathbb{T}}$. (One should view $\mathbb{F}_p[t]$, $\mathbb{F}_p[t]\big((1/t)\big)$, $\mathbb{T}$ as being characteristic p analogues of $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{R}/\mathbb{Z}$ respectively.)

Let α be an element of $\mathbb{F}_p[t]\big((1/t)\big)$ that is irrational in the sense that it is not of the form f/g for any $f, g \in \mathbb{F}_p[t]$ with g nonzero. A simple cardinality (or category, or measure) argument shows that irrational α exist, and it is not hard to give concrete examples of irrational elements. We then construct a G-system $\mathrm{X} = (X, \mathcal{X}, \mu, (T_g)_{g\in G})$ by setting X to be the "torus" $X := \mathbb{T}^m$ with the product measure $dm_{\mathbb{T}^m}$ and product Borel σ-algebra, and shifts

$$T_g((x_i)_{i=1}^m) := \left(\sum_{j=0}^{i} \binom{g}{i-j} x_j \right)_{i=1}^{m} \tag{10.1}$$

for all $x_1, \dots, x_m \in \mathbb{T}$. Here, we adopt the convention that $x_0 := \alpha$, and $\binom{g}{i} := g(g-1)\dots(g-i+1)/i!$ is viewed as an element of $\mathbb{F}_p[t]$ (note that this is well-defined for any $i < p$, which is acceptable for us since $m < p$). This shift system is a dynamical abstraction of the binomial identity

$$\binom{h+g}{i} = \sum_{j=0}^{i} \binom{g}{i-j} \binom{h}{i}$$

for any $h, g \in G$.

It is easy to see that X is a G-system. Let us verify that it is ergodic. By the Ergodic Theorem, this is equivalent to the assertion that

$$\lim_{n\to\infty} \mathbb{E}_{g\in\mathbb{F}_p^n} T_g f = \int_X f d\mu \tag{10.2}$$

in L^2 norm for all $f \in L^2(\mathrm{X})$ and some Følner sequence $(\Phi_n)_{n=1}^\infty$ of G. By Fourier decomposition and a density argument, it suffices to achieve this for functions f of the form

$$f(x_1, \dots, x_m) = e_p(a_1 x_1 + \dots + a_m x_m) \tag{10.3}$$

for $a_1, \dots, a_m \in \mathbb{F}_p[t]$, where the standard character $e_p : \mathbb{T} \to \mathbb{C}$ is defined by

$$e_p\left(\sum_{n=-\infty}^{d} c_n t^n \bmod \mathbb{F}_p[t] \right) := e^{2\pi i c_{-1}/p}.$$

If all the $a_1, \dots, a_m$ vanish, both sides of (10.2) are clearly equal to 1, so the claim is trivial in this case. Now suppose that there is $1 \le i_* \le m$ such that a_{i_*} is nonvanishing, but $a_i = 0$ for all $i_* < i \le m$. We induct on i. If $i = 1$, then we have $\mathbb{E}_{g \in \mathbb{F}_p^n} T_g f = (\mathbb{E}_{g \in \mathbb{F}_p^n} e_p(a_1 \alpha g)) f$. As α is irrational, $a_1 \alpha$ does not lie in $\mathbb{F}_p[t]$; and a direct calculation shows that this expression converges to 0 as $n \to \infty$. If $i > 1$, we need to show that the left-hand side of (10.2) converges to 0. By the van der Corput Lemma (see[12], e.g., [5, Lemma 2.9]) it suffices to show that

$$\lim_{n \to \infty} \mathbb{E}_{g \in \mathbb{F}_p^n} \int_X T_g (T_h f \overline{f}) d\mu = 0$$

for all $h \in G \backslash \{0\}$. It is easy to verify that $T_h f \overline{f}$ takes the form (10.3) for some tuple $(a_1, \dots, a_m)$ which is not identically 0, but vanishes in the $a_{i_*}, \dots, a_m$ entries, so that the claim follows from the induction hypothesis.

For each $g \in G$ and $f_0, \dots, f_k \in L^\infty(\mathrm{X})$, we consider

$$I_{c_0,\dots,c_k; f_0,\dots,f_k}(g) := \int_X (T_{c_0 g} f_0) \dots (T_{c_k g} f_k) d\mu \tag{10.4}$$

and

$$J_{c_0,\dots,c_k; f_0,\dots,f_k}(g) := \int_{HP_{c_0,\dots,c_k}(\mathbb{T}^m)_\theta} f_0 \otimes \dots \otimes f_k dm_{HP_{c_0,\dots,c_k}(\mathbb{T}^m)_{\theta(g)}}$$

where $\theta_i(g) := \binom{g}{i}\alpha$. We shall establish the limit formula

$$\lim_{g \to \infty} |I_{c_0,\dots,c_k; f_0,\dots,f_k}(g) - J_{c_0,\dots,c_k; f_0,\dots,f_k}(g)| = 0; \tag{10.5}$$

this is very similar to Lemma 1.9, but here the limit is in the classical sense (using the one-point compactification $G \cup \{\infty\}$ of G, or, equivalently, using the Frechet filter on G) rather than in the sense of uniform density. Assume (10.5) for the moment. If f is the function in Theorem 10.2 (which we identify with an element of $L^\infty(\mathrm{X})$), we see from the hypotheses on f and compactness that there exists $\varepsilon > 0$ such that

$$\int_{HP_{c_0,\dots,c_k}(\mathbb{T}^m)_\theta} f \otimes \dots \otimes f dm_{HP_{c_0,\dots,c_k}(\mathbb{T}^m)_\theta} < \left(\int_X f d\mu \right)^{k+1} - \varepsilon$$

for all $\theta \in \mathbb{T}^m$. In particular,

$$J_{c_0,\dots,c_k; f,\dots,f}(g) < \left(\int_X f d\mu \right)^{k+1} - \varepsilon$$

[12]One could also use the case $k = 1$ of Lemma 3.3 here.

for all $g \in G$; and thus by (10.5),

$$I_{c_0,\ldots,c_k;f,\ldots,f}(g) < \left(\int_X f d\mu\right)^{k+1} - \varepsilon/2$$

for all but finitely many g. Applying Theorem 10.1, we conclude that $(c_0, \ldots, c_k)$ does not have the Khintchine property.

It remains to establish the limit formula (10.5). We can repeat large portions of the proof of Lemma 1.9 to do this. Indeed, by the same Fourier decomposition used to prove Lemma 1.9, we may assume that each f_i takes the form

$$f_i(x_1, \ldots, x_m) := e_p\Big(\sum_{j=1}^m a_{ij} x_j\Big) \tag{10.6}$$

for coefficients $a_{ij} \in \mathbb{F}_p[t]$. Using Lagrange interpolation identities exactly as in the proof of Lemma 1.9, we can reduce to the case where $a_{ij} = 0$ whenever $j \le i \le k$, and then reduce further to the case in which there exists $1 \le j_* \le m$ such that $a_{j_*-1,j_*} \neq 0$ and $a_{ij} = 0$ whenever $i \ge j$ or $j > j_*$. As in the proof of Lemma 1.9, $J_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g) = 0$ in this case; so it remains to show that

$$\lim_{g\to\infty} I_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g) = 0.$$

Using (10.4), (10.6), and (10.1), we can expand the expression $I_{c_0,\ldots,c_k;f_0,\ldots,f_k}(g)$ as

$$\int_{\mathbb{T}^m} e_p\Big(\sum_{i=0}^{j_*-1}\sum_{j=1}^{j_*} a_{ij}\Big(\sum_{l=0}^{j}\binom{c_i g}{j-l} x_l\Big)\Big)\, dm_{\mathbb{T}^m}(x_1, \ldots, x_m),$$

with the convention $x_0 = \alpha$. By Fourier analysis, this expression vanishes unless

$$\sum_{i=0}^{j_*-1}\sum_{j=l}^{j_*} a_{ij}\binom{c_i g}{j-l} = 0 \tag{10.7}$$

for all $l = 1, \ldots, j_*$. Thus, it suffices to show that for all but finitely many g, the identities (10.7) do not simultaneously hold for all $l = 1, \ldots, j_*$.

Suppose for contradiction that there are infinitely many $g \in G$ for which (10.7) holds for all $l = 1, \ldots, j_*$. As the left-hand side of (10.7) is a polynomial in g, these polynomials must vanish identically for each l. In particular, extracting the g^{j_*-l} coefficient of the left-hand side, we conclude that $\sum_{i=0}^{j_*-1} a_{ij_*} c_i^{j_*-l} = 0$ for all $l = 1, \ldots, j_*$. But as the Vandermonde determinant of the $c_0, \ldots, c_{j_*-1}$ is nonvanishing, this implies that $a_{ij_*} = 0$ for all $i = 0, \ldots, j_* - 1$, giving the desired contradiction. This establishes (10.5), and Theorem 10.2 follows. □

Now we can prove Theorem 1.15. Fix $k \geq 3$ and p. We say that a property holds for **generic tuples** $(c_0, \ldots, c_k) \in \mathbb{F}_p^{k+1}$ if the number of tuples which fail to have the property is at most $C_k p^k$ for some C_k depending only on k. Thus, for instance, a generic tuple $(c_0, \ldots, c_k)$ has all entries $c_0, \ldots, c_k$ distinct. Our task is to establish that a generic tuple $(c_0, \ldots, c_k)$ does not have the Khintchine property. In view of Theorem 10.2 (applied with $m = 2$), it suffices to exhibit, for each generic tuple, a nonnegative function $f \in L^2(\mathbb{T}^2)$ with the property that

$$\int_{HP_{c_0,\ldots,c_k}(\mathbb{T}^2)_\theta} f \otimes \cdots \otimes f dm_{HP_{c_0,\ldots,c_k}(\mathbb{T}^2)_\theta} < \left(\int_{\mathbb{T}^2} f dm_{\mathbb{T}^2} \right)^{k+1}$$

for all $\theta = (\theta_1, \theta_2) \in \mathbb{T}^2$. The left-hand side can be expanded as

$$\int_{\mathbb{T}^3} \prod_{i=0}^{k} f(x_1 + c_i\theta_1, x_2 + c_i t_2 + \binom{c_i}{2}\theta_2) dm_{\mathbb{T}^3}(x_1, x_2, t_2).$$

To create this counterexample, we use the perturbative ansatz

$$f(x_1, x_2) = 1 + \sum_{(a_1,a_2)\in A} \varepsilon_{a_1,a_2} e_p(a_1 x_1 + a_2 x_2)$$

where A is a set of nonzero elements of $\mathbb{F}_p[t]^2$ of bounded cardinality (in our example, $|A| = 8$) and ε_{a_1,a_2} are small complex coefficients to be chosen later. In order for f to be real-valued, we require A to be symmetric ($A = -A$) and the coefficients ε_{a_1,a_2} to satisfy the symmetry condition

$$\varepsilon_{-a_1,-a_2} = \overline{\varepsilon_{a_1,a_2}} \tag{10.8}$$

for all $(a_1, a_2) \in A$. If A is fixed and the ε_{a_1,a_2} are chosen sufficiently small, then f is a nonnegative element of $L^\infty(\mathbb{T}^2)$. As A is assumed to not contain $(0, 0)$, f has mean 1. It thus suffices to choose A and ε_{a_1,a_2} as above, for which

$$\int_{\mathbb{T}^3} \prod_{i=0}^{k} f(x_1 + c_i\theta_1, x_2 + c_i t_2 + \binom{c_i}{2}\theta_2) dm_{\mathbb{T}^3}(x_1, x_2, t_2) < 1 \tag{10.9}$$

for all $\theta_1, \theta_2 \in \mathbb{T}$. The left-hand side of (10.9) can be expanded as

$$\sum_{(a_{1,i},a_{2,i})\in A\cup\{(0,0)\} \text{ for } i=0,\ldots,k} \left(\prod_{i=0}^{k} \varepsilon_{a_{1,i},a_{2,i}} \right)$$
$$\int_{\mathbb{T}^3} e_p \left(\sum_{i=0}^{k} a_{1,i} \left(x_i + c_i\theta_1) + a_{2,i}(x_2 + c_i t_2 + \binom{c_i}{2}\theta_2 \right) \right) dm_{\mathbb{T}^3}(x_1, x_2, t_2),$$

with the convention that $\varepsilon_{0,0} := 1$. The term in which all the $(a_{1,i}, a_{2,i})$ vanish is 1. As for the other terms, they vanish unless

$$\sum_{i=0}^{k} a_{1,i} = 0, \quad \sum_{i=0}^{k} a_{2,i} = 0, \quad \sum_{i=0}^{k} a_{2,i} c_i = 0, \tag{10.10}$$

in which case that term is equal to the expression

$$\left(\prod_{i=0}^{k} \varepsilon_{a_{1,i},a_{2,i}}\right) e_p\left(\sum_{i=0}^{k} a_{1,i} c_i \theta_1 + a_{2,i}\binom{c_i}{2}\theta_2\right).$$

In order to establish (10.9), it thus suffices to select (for a generic choice of $(c_0, \dots, c_k)$) a finite symmetric set $A \subset \mathbb{F}_p^2 \backslash \{(0,0)\}$ and sufficiently small coefficients ε_{a_1,a_2} for $(a_1, a_2) \in A$ satisfying(10.8) and the following condition.

- There is at least one choice of tuple $(a_{1,i}, a_{2,i})_{i=0}^k \in (A \cup \{(0,0)\})^{k+1}$, not all vanishing, satisfying (10.10). Furthermore, for all such tuples, one has the additional constraints

$$\sum_{i=0}^{k} a_{1,i} c_i = 0, \tag{10.11}$$

$$\sum_{i=0}^{k} a_{1,i}\binom{c_i}{2} = 0 \tag{10.12}$$

and

$$\Re \prod_{i=0}^{k} \varepsilon_{a_{1,i},a_{2,i}} < 0 \tag{10.13}$$

Indeed, from the previous discussion, the left-hand side of (10.9) equals 1 plus the left-hand side (10.13) for all tuples $(a_{1,i}, a_{2,i})_{i=0}^k$, not all vanishing, satisfying (10.10).

It remains to exhibit A and $(\varepsilon_{a_1,a_2})_{(a_1,a_2)\in A}$ with the stated properties. We do this as follows. We first locate a nontrivial quadruple $(\alpha_0, \alpha_1, \alpha_2, \alpha_3) \in \mathbb{F}_p^4$ which satsfies

$$\sum_{i=0}^{3} \alpha_i = 0 \tag{10.14}$$

$$\sum_{i=0}^{3} \alpha_i c_i = 0 \tag{10.15}$$

$$\sum_{i=0}^{3} \alpha_i c_i^2 = 0 \tag{10.16}$$

(note that these sums are well-defined when $k \geq 3$). Indeed, one can use the Lagrange interpolation formula to take

$$\alpha_i := \prod_{0 \leq j \leq 3: j \neq i} \frac{1}{c_i - c_j}. \tag{10.17}$$

Note that generically, the c_i are all distinct, so that the α_i in (10.17) are well-defined and nonvanishing. We then set

$$A := \{\sigma(\alpha_i c_i, \alpha_i) : i = 0, 1, 2, 3; \sigma \in \{-1, +1\}\}.$$

This set is clearly symmetric. Because the c_i are all distinct, the α_i are all nonzero, and the characteristic p is not equal to two, A consists of eight distinct elements of $\mathbb{F}_p^2 \backslash \{(0, 0)\}$.

Now we classify all the tuples $(a_{1,i}, a_{2,i})_{i=0}^k \in (A \cup \{(0, 0)\})^{k+1}$ satisfying (10.10). Certainly the tuple when $(a_{1,i}, a_{2,i}) = (0, 0)$ for all i does this. By (10.14), (10.15), (10.16), we see that the tuples given by $(a_{1,i}, a_{2,i}) = \sigma 1_{i \leq 3}(\alpha_i c_i, \alpha_i)$ for $\sigma = +1, -1$ also satisfies (10.10), (10.11), and (10.12), and also satisfies (10.13) if the weights $\varepsilon_{a_{1,i}, a_{2,i}}$ are chosen so that $\Re \prod_{i=0}^3 \varepsilon_{\alpha_i c_i, \alpha_i} < 0$, which is easily accomplished. To conclude the construction, it suffices to show that for generic $(c_0, \ldots, c_k)$ there are no other tuples satisfying (10.10).

Suppose for contradiction that another tuple $(a_{1,i}, a_{2,i})_{i=0}^k \in (A \cup \{(0, 0)\})^{k+1}$ satisfying (10.10) exists. We can write $(a_{1,i}, a_{2,i}) = 1_{i \in B} \sigma_i (\alpha_{j_i} c_{j_i}, \alpha_{j_i})$ for some nonempty $B \subset \{0, \ldots, k\}$, with $\sigma_i \in \{-1, +1\}$ and $j_i \in \{0, 1, 2, 3\}$ for all $i \in B$. We can exclude the cases in which $B = \{0, 1, 2, 3\}$ and $\sigma_i = \sigma$ and $j_i = i$ for all $i \in B$ and some $\sigma = \{-1, +1\}$, since those tuples were already considered. As the number of possibilities for B, σ_i, j_i depend only on k, it suffices to show that for a *fixed* choice of B, σ_i, j_i not of the above form, the condition (10.10) fails for generic $(c_0, \ldots, c_k)$.

Fix B, σ_i, j_i as above. The conditions (10.10) can then be written as

$$\sum_{i \in B} \sigma_i \alpha_{j_i} c_{j_i} = 0 \tag{10.18}$$

$$\sum_{i \in B} \sigma_i \alpha_{j_i} = 0 \tag{10.19}$$

$$\sum_{i \in B} \sigma_i \alpha_{j_i} c_i = 0. \tag{10.20}$$

Suppose first that B contains an element i_* that lies outside of $\{0, 1, 2, 3\}$. Then $\sum_{i \in B} \sigma_i \alpha_{j_i} c_i$ can be written as $\sigma_{i_*} \alpha_{j_{i_*}} c_{i_*} + Q$, where Q does not depend on c_{i_*}. Since $\alpha_{j_{i_*}}$ is generically nonzero, we conclude (after first choosing all c_i for $i \neq i_*$, and then observing that generically the constraint (10.20) can hold for at most one c_{i_*}) that (10.20) fails for generic $(c_0, \ldots, c_k)$, and we are done in this case.

Thus we may assume that $B \subset \{0, 1, 2, 3\}$. We now focus on (10.19), which asserts that a certain linear combination of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ (with coefficients in $\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$) vanishes. From (10.17), we may write

$$\alpha_i = \pm\frac{1}{V} \prod_{0 \leq i' < i'' \leq 3 : i', i'' \neq i} (c_{i'} - c_{i''}), \tag{10.21}$$

where $V := \prod_{0 \leq i' < i'' \leq 3}(c_{i'} - c_{i''})$ is the Vandermonde determinant. Thus, (10.19) can be recast as the assertion that a certain linear combination of the polynomials $\prod_{0 \leq i' < i'' \leq 3 : i', i'' \neq i}(c_{i'} - c_{i''})$ for $i = 0, 1, 2, 3$ vanishes. But it is easy to see that these polynomials are linearly independent (indeed, they each contain a monomial term that is not present in any of the other three polynomials); and so, by the Schwarz-Zippel Lemma [27], any nontrivial linear combination of these polynomials is nonzero for generic $(c_0, \ldots, c_k)$. The only remaining case occurs when all the coefficients of $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ in (10.19) vanish. There are two ways this can happen: either $j_0 = j_1 = j_2 = j_3 = j$ for some j, or (up to permutation) one has $j_0 = j_1 = j$ and $j_2 = j_3 = j'$ and $\sigma_0, \sigma_2 = +1, \sigma_1, \sigma_3 = -1$ for some $j \neq j'$.

In the former case $j_0 = j_1 = j_2 = j_3 = j$, one can cancel α_j from (10.20) to obtain a nontrivial linear constraint between c_0, c_1, c_2, c_3 with coefficients in ± 1, which then fails for generic choices of $(c_0, \ldots, c_k)$. Thus we may assume that $j_0 = j_1 = j$ and $j_2 = j_3 = j'$ and $\sigma_0, \sigma_2 = +1, \sigma_1, \sigma_3 = -1$. We then turn to (10.20), which becomes $\alpha_j(c_0 - c_1) + \alpha_{j'}(c_2 - c_3) = 0$, which by (10.21) is a constraint of the form

$$(c_0 - c_1) \prod_{0 \leq i' < i'' \leq 3 : i', i'' \neq j} (c_{i'} - c_{i''}) = \pm(c_2 - c_3) \prod_{0 \leq i' < i'' \leq 3 : i', i'' \neq j'} (c_{i'} - c_{i''}).$$

By unique factorization, the two polynomials on the left and right-hand sides here are distinct; so by the Schwartz-Zippel Lemma, this identity fails for generic $(c_0, \ldots, c_k)$, and the claim follows.

Remark. The above arguments give an explicit description of the tuples $(c_0, \ldots, c_k)$ for which the Khintchine property is still possible. Further analysis of these exceptional cases (possibly involving modification of the set A and the weights $\varepsilon_{a_{1,i}, a_{2,i}}$ might well resolve Conjecture 1.14 stated in the Introduction, but this seems to require a rather large amount of combinatorial and algebraic case checking, and will not be pursued here.

Remark. Similar counterexamples can be constructed for $\mathbb{Z}$-systems; they are weaker than those based on the Behrend construction given in [4], although they have the advantage of applying to a wider class of coefficients $c_0, \ldots, c_k$. We leave the details to the interested reader.

Acknowledgements. The third author was on sabbatical at Stanford at the time this work was carried out and thanks the Stanford Math Department for its hospitality and support. We thank the anonymous referee for many useful corrections and suggestions.

References

[1] T. Austin, *On the norm convergence of non-conventional ergodic averages*, Ergodic Theory Dynam. Systems **30** (2010), 321–338.

[2] F. A. Behrend, *On sets of integers which contain no three in arithmetic progression*, Proc. Nat. Acad. Sci. **23** (1946), 331–332.

[3] V. Bergelson, *Ergodic theory and Diophantine problems*, Topics in Symbolic Dynamics and Applications ,, Cambridge Univ. Press, Cambridge, 2000, pp. 167–205.

[4] V. Bergelson, B. Host and B. Kra, *Multiple recurrence and nilsequences*, Invent. Math. **160** (2005), 261-303.

[5] V. Bergelson, A. Leibman, and R. McCutcheon *Polynomial Szemeredi theorem for countable modules over integral domains and finite fields*, J. Analyse Math. **95** (2005), 243–296.

[6] V. Bergelson, T. C. Tao, and T. Ziegler, *An inverse theorem for uniformity seminorms associated with the action of F^{ω}*, Geom. Funct. Anal. **19** (2010), 1539–1596.

[7] J. P. Conze and E. Lesigne, *Théorèmes ergodique pour des mesures diagonales*, Bull. Soc. Math. France **112** (1984), 143–175.

[8] J. P. Conze and E. Lesigne. *Sur un théorème ergodique pour des mesures diagonales*, Probabilités, Univ. Rennes I, Rennes, 1988, pp. 1–31.

[9] J. P. Conze and E. Lesigne, *Sur un théorème ergodique pour des mesures diagonales*, C. R. Acad. Sci. Paris, Série I, **306** (1988), 491–493.

[10] T. de la Rue, *An introduction to joinings in ergodic theory*, Discrete Contin. Dyn. Syst. **15** (2006), 121-142.

[11] N. Frantzikinakis, *Multiple ergodic averages for three polynomials and applications*, Trans. Amer. Math. Soc. **360** (2008), 5435–5475.

[12] E. Følner, *On groups with full Banach mean value*, Math Scand. **3** (1955), 243–254.

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

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

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

[16] H. Furstenberg and B. Weiss, *A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$*, Convergence in Ergodic Theory and Probability, de Gruyter, Berlin, 1996, pp. 193–227.

[17] B. Green, *Montréal notes on quadratic Fourier analysis*, Additive Combinatorics, Amer. Math. Soc., Providence, RI, 2007, pp. 69–102.

[18] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), 465–588.

[19] P. Hall, *A contribution to the theory of groups of prime-power order*, Proc. London Math. Soc. **36** (1933), 29–95.

[20] B. Host, *Ergodic seminorms for commuting transformations and applications*, Studia Math. **195** (2009), 31–49.

[21] B. Host and B. Kra. Personal communication, 2001.

[22] B. Host and B. Kra, *Averaging along cubes*, Modern Dynamical Systems and Applications, Cambridge University Press, Cambridge, 2004, pp. 123–144.

[23] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), 397–488.

[24] A. Leibman, *Polynomial mappings of groups*, Israel J. Math. **129** (2002), 29–60.

[25] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), 201–213.

[26] J. Petresco, *Sur les commutateurs*, Math. Z. **61** (1954), 348–356.

[27] J. Schwartz, *Fast probabilistic algorithms for verification of polynomial identities*, J. ACM **27** (1980), 701–717.

[28] B. Szegedy, *Structure of finite nilspaces and inverse theorems for the Gowers norms in bounded exponent groups*, arXiv:1011.1057v1[math.CO]

[29] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

[30] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Systems **28** (2008), 657–688.

[31] T. Tao and T. Ziegler, *The inverse conjecture for the Gowers norms over finite fields via the correspondence principle*, Anal. PDE **3** (2010), 1–20.

[32] T. Tao and T. Ziegler, *The inverse conjecture for the Gowers norm over finite fields in low characteristic*, Ann. Comb. **16** (2012), 121–188.

[33] H. Towsner, *Convergence of diagonal ergodic averages*, Ergodic Theory Dynam. Systems **29** (2009), 1309–1326.

[34] J. von Neumann, *Zur allgemeinen Theorie des Masses*, Fund. Math. **13** (1929), 73–116.

[35] V. S. Varadarajan, *Groups of automorphisms of Borel spaces*, Trans. Amer. Math. Soc. **109** (1963), 191–220.

[36] M. Walsh, *Norm convergence of nilpotent ergodic averages*, Ann. of Math. (2) **175** (2012), 1667–1688.

[37] Q. Zhang *On convergence of the averages $\frac{1}{N}\sum_{n=1}^{N} f_1(R^n x) f_2(S^n x) f_3(T^n x)$*, Monatsh. Math. **122** (1996), 275–300.

[38] T. Ziegler, *Non-conventional ergodic averages*, PhD Thesis, The Hebrew University, 2003.

[39] T. Ziegler, *A non-conventional ergodic theorem for a nilsystem*, Ergodic Theory Dynam. Systems **25** (2005), 1357–1370.

[40] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), 53–97.

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

[42] P. Zorin-Kranich, *Norm convergence of multiple ergodic averages on amenable groups*, J. Anal. Math., to appear.

Vitaly Bergelson
DEPARTMENT OF MATHEMATICS
THE OHIO-STATE UNIVERSITY
COLUMBUS, OH 43210 USA
email: vitaly@math.ohio-state.edu

Terence Tao
DEPARTMENT OF MATHEMATICS
UCLA
LOS ANGELES, CA 90095-1555 USA
email: tao@math.ucla.edu

Tamar Ziegler
DEPARTMENT OF MATHEMATICS
TECHNION
HAIFA, ISRAEL 32000
email: tamarzr@tx.technion.ac.il

(Received May 19, 2013 and in revised form November 4, 2013)