NON-CONVENTIONAL ERGODIC AVERAGES FOR SEVERAL COMMUTING ACTIONS OF AN AMENABLE GROUP

By

TIM AUSTIN*

Abstract. Let (X, μ) be a probability space, G a countable amenable group, and $(F_n)_n$ a left Følner sequence in G. This paper analyzes the non-conventional ergodic averages

$$\frac{1}{|F_n|}\sum_{g\in F_n}\prod_{i=1}^{d}(f_i \circ T_1^g \cdots T_i^g)$$

associated to a commuting tuple of μ-preserving actions $T_1, \ldots, T_d : G \curvearrowright X$ and $f_1, \ldots, f_d \in L^\infty(\mu)$. We prove that these averages always converge in $\|\cdot\|_2$, and that they witness a multiple recurrence phenomenon when $f_1 = \ldots = f_d = 1_A$ for a non-negligible set $A \subseteq X$. This proves a conjecture of Bergelson, McCutcheon and Zhang. The proof relies on an adaptation from earlier works of the machinery of sated extensions.

1 Introduction

Let (X, μ) be a probability space, G a countable amenable group, and $T_1, \ldots, T_d : G \curvearrowright (X, \mu)$ a tuple of μ-preserving actions of G which commute, meaning that $i \neq j$ implies $T_i^g T_j^h = T_j^h T_i^g$ for all $g, h \in G$. Also, let $(F_n)_n$ be a left Følner sequence of subsets of G; this is fixed for the rest of the paper.

In this context, Bergelson, McCutcheon and Zhang have proposed in [BMZ97] the study of the non-conventional ergodic averages

$$\Lambda_n(f_1, \ldots, f_d) := \frac{1}{|F_n|}\sum_{g\in F_n}\prod_{i=1}^{d}(f_i \circ T_1^g \cdots T_i^g) \tag{1.1}$$

for functions $f_1, \ldots, f_d \in L^\infty(\mu)$. These are an analog for commuting G-actions of the non-conventional averages for a commuting tuple of transformations, as introduced by Furstenberg and Katznelson [FK78] for their proof of the multi-dimensional generalization of Szemerédi's Theorem. Other analogs are possible,

*Research supported by a fellowship from the Clay Mathematics Institute.

JOURNAL D'ANALYSE MATHÉMATIQUE, Vol. 130 (2016)
DOI 10.1007/s11854-016-0036-6

but the averages above seem to show the most promise for building a theory: this is discussed in [BMZ97] and, for topological dynamics, in [BH92], where some relevant counterexamples are presented.

The main results of [BMZ97] are that these averages converge and that one has an associated multiple recurrence phenomenon, when $d = 2$. The first of these conclusions can be extended to arbitrary d along the lines of Walsh's recent proof of convergence for polynomial nilpotent non-conventional averages ([Wal12]).

Theorem A. *In the setting above, the functional averages $\Lambda_n(f_1, \dots, f_d)$ converge in the norm of $L^2(\mu)$ for all $f_1, \dots, f_d \in L^\infty(\mu)$.*

Zorin-Kranich has made the necessary extensions to Walsh's argument in [ZK]. However, his proof gives essentially no information about the limiting function, and in particular does not seem to enable a proof of multiple recurrence. The present paper gives both a new proof of Theorem A, and a proof of the following.

Theorem B. *If $\mu(A) > 0$, then*

$$\lim_{n\to\infty} \int_X \Lambda_n(1_A, \dots, 1_A)\mathrm{d}\mu = \lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu\big(T_1^{g^{-1}}A \cap \cdots \cap (T_1^{g^{-1}} \cdots T_d^{g^{-1}})A\big) > 0.$$

In particular, the set $\{g \in G : \mu\big(T_1^{g^{-1}}A \cap \cdots \cap (T_1^{g^{-1}} \cdots T_d^{g^{-1}})A\big) > 0\}$ has positive upper Banach density relative to $(F_n)_{n\geq 1}$.

As in the classical case of [FK78], this implies the following Szemerédi-type result for amenable groups.

Corollary. *Let G^d be the direct sum of d copies of G. If $E \subseteq G^d$ has positive upper Banach density relative to $(F_n^d)_{n\geq 1}$, then the set*

$$\{g \in G : \exists (x_1, \dots, x_d) \in G^d \text{ s. t. } \{(g^{-1}x_1, x_2, \dots, x_d), \dots, (g^{-1}x_1, \dots, g^{-1}x_d)\} \subseteq E\}$$

has positive upper Banach density relative to $(F_n)_{n\geq 1}$.

This deduction is quite standard, and can be found in [BMZ97].

Our proofs of Theorems A and B are descended from some work for commuting tuples of transformations: the proof of non-conventional-average convergence in [Aus09], and that of multiple recurrence in [Aus10a]. Both of those papers offered alternatives to earlier proofs, using new machinery for extending an initially-given probability-preserving action to another action under which the averages behave more simply. The present paper adapts to commuting tuples of G-actions the notion of a "sated extension", which forms the heart of the streamlined presentation of that machinery in [Aus10b]. Further discussion of this method may be found in that reference.

The generalization of the notion of satedness is nontrivial, but fairly straightforward: see Section 3 below. However, more serious difficulties appear in how it is applied. Heuristically, if a given system satisfies a satedness assumption, then, in any extension of that system, this constrains how some canonical σ-subalgebra "sits" relative to the σ-algebra lifted from the original system. An appeal to satedness always relies on constructing a particular extension for which this constraint implies some other desired consequence. The specific constructions of system extensions used in [Aus09, Aus10a, Aus10b] do not generalize to commuting actions of a non-abelian group G. This is because they rely on the commutativity of the diagonal actions $T_i \times \cdots \times T_i$ of G on X^d with the "off-diagonal" action generated by $T_1^g \times \cdots \times (T_1^g \cdots T_d^g)$, $g \in G$.

Thus, a key part of this paper is a new method of extending probability-preserving G^d-systems. It is based on a version of the Host-Kra self-joinings from [HK05] and [Hos09]. It also relies on a quite general result about probability-preserving systems, which may be of independent interest: Theorem 2.1 asserts that, given a probability-preserving action of a countable group and an extension of that action restricted to a subgroup, a compatible further extension may be found for the action of the whole group.

Developing ideas from [HK05], we find that the asymptotic behaviour of our non-conventional averages can be estimated by certain integrals over these Host-Kra-like extensions (Theorem 4.5). On the other hand, a suitable satedness assumption on a system gives extra information on the structure of those extensions, and combining these facts then implies simplified behaviour for the non-conventional averages for that system. Finally, the existence of sated extensions for all systems (Theorem 3.5) then enables proofs of convergence and multiple recurrence similar to those in [Aus09] and [Aus10a], respectively.

Bergelson and McCutcheon in [BM07] have suggested an interesting direction for further research. They studied multiple recurrence phenomena similar to Theorem B when $d = 3$, but without assuming that the group G is amenable, and proved that the set

$$\big\{g \in G : \mu\big(T_1^{g^{-1}}A \cap (T_1^{g^{-1}}T_2^{g^{-1}})A \cap (T_1^{g^{-1}}T_2^{g^{-1}}T_3^{g^{-1}})A\big) > 0\big\}$$

is "large" in a sense adapted to non-amenable groups, in terms of certain special ultrafilters in the Stone-Čech compactification of G. In particular, their result implies that this set is syndetic in G. Can their methods be combined with those below to extend this result to larger values of d?

2 Generalities on actions and extensions

2.1 Preliminaries. If $d \in \mathbb{N}$ then $[d] := \{1, 2, \ldots, d\}$; more generally, if $a, b \in \mathbb{Z}$ with $a \leq b$, then

$$(a; b] = [a+1; b] = [a+1; b+1) = (a; b+1) := \{a+1, \ldots, b\}.$$

The power set of $[d]$ is denoted $\mathcal{P}[d]$, and we let $\binom{[d]}{\geq p} := \{e \in \mathcal{P}[d] : |e| \geq p\}$.

Next, we call $\mathcal{A} \subseteq \mathcal{P}[d]$ an **up-set** if $a, b \in \mathcal{A}$ implies $a \cup b \in \mathcal{A}$. The set $\langle e \rangle := \{a \subseteq [d] : a \supseteq e\}$ is an up-set for every $e \subseteq [d]$, and every up-set is a union of such examples. On the other hand, we call $\mathcal{B} \subseteq \mathcal{P}[d]$ an **antichain** if $a, b \in \mathcal{B}$ and $a \subseteq b$ implies $a = b$. Every up-set contains a unique anti-chain of inclusion-minimal elements.

Standard notions from probability theory are assumed throughout this paper. If (X, μ) is a probability space with σ-algebra Σ, and if $\Phi, \Sigma_1, \Sigma_2 \subseteq \Sigma$ are σ-subalgebras with $\Phi \subseteq \Sigma_1 \cap \Sigma_2$, then Σ_1 and Σ_2 are **relatively independent** over Φ under μ if

$$\int_X fg \, \mathrm{d}\mu = \int_X \mathsf{E}_\mu(f \mid \Phi) \mathsf{E}_\mu(g \mid \Phi) \, \mathrm{d}\mu$$

whenever $f, g \in L^\infty(\mu)$ are Σ_1- and Σ_2-measurable, respectively. Relatedly, if (X, μ) is standard Borel, then on X^2 we may form the **relative product** measure $\mu \otimes_\Phi \mu$ over Φ by letting $x \mapsto \mu_x$ be a disintegration of μ over the σ-subalgebra Φ and then setting $\mu \otimes_\Phi \mu := \int_X \mu_x \otimes \mu_x \, \mu(\mathrm{d}x)$.

Let G be a countable group. Then a G**-space** is a triple (X, μ, T) consisting of a probability space (X, μ) and an action $T : G \curvearrowright X$ by measurable, μ-preserving transformations. Passing to an isomorphic model if necessary, we henceforth assume that (X, μ) is standard Borel. Often, a G-space is also denoted by a boldface letter such as $\mathbf{X}$.

If $\mathbf{X} = (X, \mu, T)$ is a G-space, then Σ_X or $\Sigma_{\mathbf{X}}$ denotes its σ-algebra of μ-measurable sets. A **factor** of such a G-space is a σ-subalgebra $\Phi \leq \Sigma_{\mathbf{X}}$ which is globally T-invariant, meaning that $A \in \Phi$ implies $T^g(A) \in \Phi$ for all $g \in G$. Relatedly, a **factor map** from one G-space $\mathbf{X} = (X, \mu, T)$ to another $\mathbf{Y} = (Y, \nu, S)$ is a measurable map $\pi : X \to Y$ such that $\pi_*\mu = \nu$ and $S^g \circ \pi = \pi \circ T^g$ for all $g \in G$, μ-a.e. In this case, $\pi^{-1}(\Sigma_{\mathbf{Y}})$ is a factor of $\mathbf{X}$. Such a factor map is also referred to as a G**-extension**, and $\mathbf{X}$ may be referred to as an **extension** of $\mathbf{Y}$.

On the other hand, if $\mathbf{X} = (X, \mu, T)$ is a G-space and $H \leq G$, then the H**-subaction** of $\mathbf{X}$, denoted $\mathbf{X}^{\upharpoonright H} = (X, \mu, T^{\upharpoonright H})$, is the H-space with probability space

(X, μ) and action given by the transformations $(T^h)_{h\in H}$. The associated σ-algebra of H-almost-invariant sets, $\{A \in \Sigma_X : \mu(T^h(A)\triangle A) = 0 \text{ for all } h \in H\}$, is denoted by either $\Sigma_{\mathbf{X}}^H$ or $\Sigma_{\mathbf{X}}^{T\restriction H}$, as seems appropriate.

In the sequel, we often consider a space (X, μ) endowed with a commuting tuple $T_1, \dots, T_d$ of G-actions. Slightly abusively, we simply refer to this as a "G^d-action" or "G^d-space" (leaving the distinguished G-subactions to the reader's understanding) and denote it by $(X, \mu, T_1, \dots, T_d)$. Also, for a G^d-space $(X, \mu, T_1, \dots, T_d)$ and $a, b \in [d]$ with $a \le b$, we frequently write

$$T^g_{[a;b]} = T^g_{(a-1;b]} = T^g_{[a,;b+1)} := T^g_a T^g_{a+1} \cdots T^g_b \quad \text{for all } g \in G.$$

Because the actions T_i commute, this defines another G-action for each a, b.

2.2 Actions of groups and their subgroups. Our approach to proving Theorems A and B descends from the notions of "pleasant" and "isotropized" extensions. These were introduced in [Aus09] and [Aus10a] respectively, where they were used to give new proofs of the analogs of Theorems A and B for commuting tuples of single transformations.

Subsequently, the more general notion of "sated"' extensions was introduced in [Aus10b]. It simplifies and clarifies those earlier ideas as special cases. In this paper, we show how "sated" extensions can be adapted to the non-abelian setting of Theorems A and B.

An important new difficulty is that we need to consider certain natural σ-subalgebras of a probability-preserving G-spaces which need not be factors in case G is not abelian. This subsection focuses on a key tool for handling this situation, which seems to be of interest in its own right. Given $H \le G$, it enables one to turn an extension of an H-subaction into an extension of a whole G-action. Satedness is introduced in the next subsection.

Theorem 2.1. *Suppose $H \le G$ is an inclusion of countable groups, that $\mathbf{X} = (X, \mu, T)$ is a G-space, and that*

$$\mathbf{Y} = (Y, \nu, S) \xrightarrow{\beta} \mathbf{X}^{\restriction H}$$

is an extension of H-spaces. Then there is an extension of G-spaces $\widetilde{\mathbf{X}} \xrightarrow{\pi} \mathbf{X}$ which admits a commutative diagram of H-spaces

$$\begin{tikzcd} \widetilde{\mathbf{X}}^{\restriction H} \arrow[rr, "\pi"] \arrow[dr, "\alpha"'] & & \mathbf{X}^{\restriction H} \\ & \mathbf{Y} \arrow[ur, "\beta"'] & \end{tikzcd}$$

This theorem was proved for abelian G and H in [Aus15, Subsection 3.2]. The non-abelian case is fairly similar.

Proof. We construct the new G-space $\widetilde{\mathbf{X}}$ by a kind of "relativized" co-induction of $\mathbf{Y}$ over $\mathbf{X}$ and then show that it has the necessary properties.

The construction of a suitable standard Borel dynamical system $(\widetilde{X}, \widetilde{T})$, deferring the construction of the measure, is easy. Let

$$\widetilde{X} := \{(y_g)_g \in Y^G : y_{gh} = S^{h^{-1}} y_g \text{ and } \beta(y_g) = T^{g^{-1}}\beta(y_e) \text{ for all } g \in G, h \in H\},$$

and let $\widetilde{T} : G \curvearrowright \widetilde{X}$ be the restriction to $\widetilde{X}$ of the left-regular representation:

$$\widetilde{T}^k((y_g)_{g\in G}) = (y_{k^{-1}g})_{g\in G}$$

(it is easily seen that this preserves $\widetilde{X} \subseteq Y^G$).

Also, let

$$\alpha : \widetilde{X} \to Y : (y_g)_g \mapsto y_e \quad \text{and} \quad \pi := \beta \circ \alpha : \widetilde{X} \to X : (y_g)_g \mapsto \beta(y_e).$$

These maps fit into a commutative diagram of the desired shape by construction. It remains to specify a suitable measure $\widetilde{\mu}$ on $\widetilde{X}$. It is be constructed as a measure on Y^G for which $\widetilde{\mu}(\widetilde{X}) = 1$.

Let $X \to \Pr Y : x \mapsto \nu_x$ be a disintegration of ν over the map $\beta : Y \to X$. Using this, define new probability measures for each $x \in X$ as follows. First, for each $g \in G$, define $\widetilde{\nu}_{g,x}$ on Y^{gH} by $\widetilde{\nu}_{g,x} := \int_Y \delta_{(S^{h^{-1}}y)_{gh\in gH}}\, \nu_x(\mathrm{d}y)$. Now let $C \subseteq G$ be a cross-section for the space G/H of left-cosets, identify $Y^G = \prod_{c\in C} Y^{cH}$; and on this product, define $\widetilde{\nu}_x := \bigotimes_{c\in C} \widetilde{\nu}_{c,T^{c^{-1}}x}$. One may easily write down the finite-dimensional marginals of $\widetilde{\nu}_x$ directly. If $c_1, \dots, c_m \in C$, and $h_{i,1}, \dots, h_{i,n_i} \in H$ for each $i \le m$, and also $A_{i,j} \in \Sigma_Y$ for all $i \le m$ and $j \le n_i$, then

$$\begin{aligned}
&\widetilde{\nu}_x\big\{(y_g)_g : y_{c_ih_{i,j}} \in A_{i,j}\ \forall i \le m,\ j \le n_i\big\} \\
&\qquad = \prod_{i=1}^m \widetilde{\nu}_{c_i, T^{c_i^{-1}}x}\big\{(y_{c_ih})_{h\in H} :\ y_{c_ih_{i,j}} \in A_{i,j}\ \forall j \le n_i\big\} \\
&\qquad = \prod_{i=1}^m \nu_{T^{c_i^{-1}}x}\big(S^{h_{i,1}}(A_{i,1}) \cap \dots \cap S^{h_{i,n_i}}(A_{i,n_i})\big).
\end{aligned}$$

The following basic properties of $\widetilde{\nu}_x$ are now easily checked.

(i) If $g_1H = g_2H$, say, with $g_1 = g_2h_1$, and $x \in X$, then

$$\begin{aligned}
\widetilde{\nu}_{g_1,T^{g_1^{-1}}x} &= \int_Y \delta_{(S^{h^{-1}}y)_{g_1h\in g_1H}}\ \nu_{T^{g_1^{-1}}x}(\mathrm{d}y) = \int_Y \delta_{(S^{h^{-1}}y)_{g_2h_1h\in g_2H}}\ \nu_{T^{h_1^{-1}}T^{g_2^{-1}}x}(\mathrm{d}y) \\
&= \int_Y \delta_{(S^{h^{-1}}y)_{g_2h_1h\in g_2H}}\ (S_*^{h_1^{-1}}\nu_{T^{g_2^{-1}}x})(\mathrm{d}y) = \int_Y \delta_{(S^{h^{-1}}S^{h_1^{-1}}y)_{g_2h_1h\in g_2H}}\ \nu_{T^{g_2^{-1}}x}(\mathrm{d}y) \\
&= \widetilde{\nu}_{g_2,T^{g_2^{-1}}x}.
\end{aligned}$$

It follows that $\widetilde{\nu}_x$ does not depend on the choice of cross-section C, and (2.1) holds with any choice of C.

(ii) For each $g \in G$, say $g = ch \in cH$, the marginal of $\widetilde{\nu}_x$ on coordinate g is

$$S_*^{h^{-1}} \nu_{T^{c^{-1}}x} = \nu_{T^{h^{-1}}T^{c^{-1}}x} = \nu_{T^{g^{-1}}x}.$$

(iii) If $(y_g)_g$ is sampled at random from $\widetilde{\nu}_x$ and $g \in cH$, then y_c a.s. determines the whole tuple $(y_{ch})_{ch \in cH}$. Specifically, $y_{ch} = S^{h^{-1}} y_c$ a.s.

Also, if $g_1, \dots, g_m$ lie in distinct left-cosets of H and $(y_g)_g \sim \widetilde{\nu}_x$, then the coordinates $y_{g_1}, \dots, y_{g_m}$ are independent, but we do not need this fact.

Finally, let $\widetilde{\mu} := \int_X \widetilde{\nu}_x \mu(\mathrm{d}x)$. Recalling the definition of $\widetilde{X}$, we see that properties (ii) and (iii) above imply that $\widetilde{\nu}_x(\widetilde{X}) = 1$ for all x; hence also $\widetilde{\mu}(\widetilde{X}) = 1$.

We have seen that the left-regular representation defines an action of G on $\widetilde{X}$, and the required triangular diagram commutes by the definition of π, so it remains to check the following.

- (The new G-space $(\widetilde{X}, \widetilde{\mu}, \widetilde{T})$ is probability-preserving.) Suppose that $k \in G$ and $x \in X$, that $c_1, \dots, c_m \in C$, that $h_{i,1}, \dots, h_{i,n_i} \in H$ for each $i \le m$, and that $A_{i,j} \in \Sigma_Y$ for all $i \le m$ and $j \le n_i$. Then

$$\begin{aligned}\widetilde{T}_*^k \widetilde{\nu}_x\big\{(y_g)_g : y_{c_i h_{i,j}} \in A_{i,j}\ \forall i \le m,\ j \le n_i\big\} \\ &= \widetilde{\nu}_x\big\{\widetilde{T}^{k^{-1}}(y_g)_g :\ y_{c_i h_{i,j}} \in A_{i,j}\ \forall i \le m,\ j \le n_i\big\} \\ &= \widetilde{\nu}_x\big\{(y_g)_g : y_{k^{-1}c_i h_{i,j}} \in A_{i,j}\ \forall i \le m,\ j \le n_i\big\}\end{aligned}$$

  Since C is a cross-section of G/H, so is $k^{-1}C$. We may therefore apply (2.1) with the cross-section $k^{-1}C$ to deduce that the above is equal to

$$\prod_{i=1}^{m} \nu_{T^{c_i^{-1}k}x}\big(S^{h_{i,1}}(A_{i,1}) \cap \cdots \cap S^{h_{i,n_i}}(A_{i,n_i})\big).$$

  On the other hand, (2.1) applied with the cross-section C gives that this is equal to

$$\widetilde{\nu}_{T^k x}\big\{(y_g)_g :\ y_{c_i h_{i,j}} \in A_{i,j}\ \forall i \le m,\ j \le n_i\big\}.$$

  Therefore $\widetilde{T}_*^k \widetilde{\nu}_x = \widetilde{\nu}_{T^k x}$, and integrating this over x gives $\widetilde{T}_*^k \widetilde{\mu} = \widetilde{\mu}$.
- (The map α defines a factor map of H-spaces.) If $h \in H$ and $(y_g)_g \in \widetilde{X}$, then

$$\alpha(\widetilde{T}^h((y_g)_g)) = \alpha((y_{h^{-1}g})_g) = y_{h^{-1}} = S^h y_e = S^h \alpha((y_g)_g),$$

  where the penultimate equality is given by property (iii) above. Also, property (ii) above gives $\alpha_* \widetilde{\mu} = \int_X \alpha_* \widetilde{\nu}_x\, \mu(\mathrm{d}x) = \int_X \nu_x\, \mu(\mathrm{d}x) = \nu$.

- (The map π defines a factor map of G-spaces.) If $k \in G$ and $(y_g)_g \in \widetilde{X}$, then

$$\pi(\widetilde{T}^k((y_g)_g)) = \beta(\alpha((y_{k^{-1}g})_g)) = \beta(y_{k^{-1}}).$$

If $x \in X$ and $(y_g)_g \sim \widetilde{\nu}_x$, then property (ii) above gives that $y_{k^{-1}} \sim \nu_{T^k x}$, and hence $\beta(y_{k^{-1}}) = T^k x = T^k \beta(y_e)$ a.s. Since this holds for every x, integrating over x gives $\pi(\widetilde{T}^k((y_g)_g)) = T^k \pi((y_g)_g)$ a.s. Another appeal to property (ii) above gives

$$\pi_* \widetilde{\mu} = \int_X \pi_* \widetilde{\nu}_x \, \mu(\mathrm{d}x) = \int_X \beta_* \nu_x \, \mu(\mathrm{d}x) = \int_X \delta_x \, \mu(\mathrm{d}x) = \mu. \qquad \square$$

3 Functorial σ-subalgebras and subspaces, and satedness

Definition 3.1 (Functorial σ-subalgebras and subspaces)**.** A **functorial σ-subalgebra of G-spaces** is a map F which to each G-space $\mathbf{X} = (X, \mu, T)$ assigns a μ-complete σ-subalgebra $\Sigma^{\mathsf{F}}_{\mathbf{X}} \subseteq \Sigma_{\mathbf{X}}$, and such that $\Sigma^{\mathsf{F}}_{\mathbf{X}} \supseteq \pi^{-1}(\Sigma^{\mathsf{F}}_{\mathbf{Y}})$ for every G-extension $\pi : \mathbf{X} \to \mathbf{Y}$. Similarly, a **functorial L^2-subspace of G-spaces** is a map V which to each G-space $\mathbf{X} = (X, \mu, T)$ assigns a closed subspace $\mathsf{V}_{\mathbf{X}} \leq L^2(\mu)$, and such that $\mathsf{V}_{\mathbf{X}} \geq \mathsf{V}_{\mathbf{Y}} \circ \pi := \{f \circ \pi : f \in \mathsf{V}_{\mathbf{Y}}\}$ for every G-extension $\pi : \mathbf{X} \to \mathbf{Y}$. In this setting, $P^{\mathsf{V}}_{\mathbf{X}} : L^2(\mu) \to \mathsf{V}_{\mathbf{X}}$ denotes the orthogonal projection onto $\mathsf{V}_{\mathbf{X}}$.

The above behaviour relative to factors is called the **functoriality** of F or V. Its first consequence is that F and V respect isomorphisms of G-spaces: if $\alpha : \mathbf{X} \xrightarrow{\cong} \mathbf{Y}$, then $\Sigma^{\mathsf{F}}_{\mathbf{X}} = \alpha^{-1}(\Sigma^{\mathsf{F}}_{\mathbf{Y}})$ (where strict equality holds owing to the assumption that these σ-algebras are both μ-complete) and $\mathsf{V}_{\mathbf{X}} = \mathsf{V}_{\mathbf{Y}} \circ \alpha$.

Example. If $H \leq G$ is a subgroup, the map $\mathbf{X} \mapsto \Sigma^H_{\mathbf{X}}$ (the σ-subalgebra of H-almost-invariant sets) defines a functorial σ-subalgebra of G-spaces. In case $H \trianglelefteq G$, this actually defines a factor of $\mathbf{X}$, but otherwise it may not: in general, $T^g(\Sigma^H_{\mathbf{X}}) = \Sigma^{gHg^{-1}}_{\mathbf{X}}$.

This class of examples provides the building blocks for all of the other functorial σ-subglebras that we encounter later.

If F is a functorial σ-subalgebra of G-spaces, then setting $\mathsf{V}_{\mathbf{X}} := L^2(\mu|\Sigma^{\mathsf{F}}_{\mathbf{X}})$ defines a functorial L^2-subspace of G-spaces, where this denotes the subspace of $L^2(\mu)$ generated by the $\Sigma^{\mathsf{F}}_{\mathbf{X}}$-measurable functions. In this case, $P^{\mathsf{V}}_{\mathbf{X}}$ is the operator of conditional expectation onto $\Sigma^{\mathsf{F}}_{\mathbf{X}}$. However, not all functorial L^2-subspaces arise in this way. For instance, given any two functorial L^2-subspaces V_1, V_2 of G-spaces, a new functorial L^2-subspace may be defined by $\mathsf{V}_{\mathbf{X}} := \overline{\mathsf{V}_{1,\mathbf{X}} + \mathsf{V}_{2,\mathbf{X}}}$. If $H_1, H_2 \leq G$, then this gives rise to the example $\mathsf{V}_{\mathbf{X}} := \overline{L^2(\mu\,|\,\Sigma_{\mathbf{X}}^{H_1}) + L^2(\mu\,|\,\Sigma_{\mathbf{X}}^{H_2})}$. The elements of this subspace generate the functorial σ-algebra $\Sigma_{\mathbf{X}}^{H_1} \vee \Sigma_{\mathbf{X}}^{H_2}$; but in general,

$$\overline{L^2(\mu\,|\,\Sigma_{\mathbf{X}}^{H_1}) + L^2(\mu\,|\,\Sigma_{\mathbf{X}}^{H_2})} \lneqq L^2(\mu\,|\,\Sigma_{\mathbf{X}}^{H_1} \vee \Sigma_{\mathbf{X}}^{H_2}).$$

In fact, the functorial L^2-subspaces that appear later in this work all correspond to functorial σ-subalgebras. However, the theory of satedness depends only on the subspace structure, so it seems appropriate to develop it in that generality.

To prepare for the next definition, recall that if $\mathfrak{K}_1, \mathfrak{K}_2 \leq \mathfrak{H}$ are two closed subspaces of a real Hilbert space, and $\mathfrak{L} \leq \mathfrak{K}_1 \cap \mathfrak{K}_2$ is a common further closed subspace, then $\mathfrak{K}_1$ and $\mathfrak{K}_2$ are **relatively orthogonal** over $\mathfrak{L}$ if $\langle u, v\rangle = \langle P_{\mathfrak{L}}u, P_{\mathfrak{L}}v\rangle$ for all $u \in \mathfrak{K}_1,\ v \in \mathfrak{K}_2$, where $P_{\mathfrak{L}}$ is the orthogonal projection onto $\mathfrak{L}$. This requires that $\mathfrak{L} = \mathfrak{K}_1 \cap \mathfrak{K}_2$, and is equivalent to asserting that $P_{\mathfrak{K}_2}u = P_{\mathfrak{L}}u$ for all $u \in \mathfrak{K}_1$ and $P_{\mathfrak{K}_1}u = P_{\mathfrak{L}}u$ for all $u \in \mathfrak{K}_2$. Clearly, it suffices to verify this for elements drawn from any dense subsets of $\mathfrak{K}_1$ and $\mathfrak{K}_2$.

Definition 3.2 (Satedness)**.** Let V be a functorial L^2-subspace of G-spaces. We say that a G-space $\mathbf{X} = (X, \mu, T)$ is **V-sated** if for every G-extension $\mathbf{Y} = (Y, \nu, S) \xrightarrow{\xi} (X, \mu, T)$, the subspaces $L^2(\mu) \circ \xi$ and $\mathsf{V}_{\mathbf{Y}}$ are relatively orthogonal over their common further subspace $\mathsf{V}_{\mathbf{X}} \circ \xi$. More generally, a G-extension $\widetilde{\mathbf{X}} \xrightarrow{\pi} \mathbf{X}$ is **relatively V-sated** if for any further G-extension $\mathbf{Y} \xrightarrow{\xi} \widetilde{\mathbf{X}}$, the subspaces $L^2(\mu) \circ (\pi \circ \xi)$ and $\mathsf{V}_{\mathbf{Y}}$ are relatively orthogonal over $\mathsf{V}_{\widetilde{\mathbf{X}}} \circ \pi$.

Clearly, a G-space $\mathbf{X}$ is V-sated if and only if $\mathbf{X} \xrightarrow{\mathrm{id}} \mathbf{X}$ is relatively V-sated. In case $\mathsf{V}_{\mathbf{X}} = L^2(\mu\,|\,\Sigma_{\mathbf{X}}^{\mathsf{F}})$ for some functorial σ-algebra F, we say that a G-space or G-extension is **F-sated**, rather than V-sated. For a G-space $\mathbf{X} = (X, \mu, T)$, this asserts that for each G-extension $\xi : \mathbf{Y} = (Y, \nu, S) \to \mathbf{X}$, the σ-subalgebras $\xi^{-1}(\Sigma_X)$ and $\Sigma_{\mathbf{Y}}^{\mathsf{F}}$ are relatively independent over $\xi^{-1}(\Sigma_{\mathbf{X}}^{\mathsf{F}})$.

The key feature of satedness is that all G-spaces have sated extensions. This generalizes the corresponding result for satedness relative to idempotent classes ([Aus10b, Theorem 2.3.2]). The proof here is a nearly a verbatim copy of that one, given the following auxiliary lemmas.

Lemma 3.3. *Suppose that $\widetilde{\mathbf{X}} \xrightarrow{\pi} \mathbf{X}$ is a relatively V-sated G-extension, and $\mathbf{Z} = (Z, \theta, R) \xrightarrow{\alpha} \widetilde{\mathbf{X}}$ is a further G-extension. Then $\mathbf{Z} \xrightarrow{\alpha\circ\pi} \mathbf{X}$ is also relatively V-sated.*

Proof. Suppose that $\mathbf{Y} = (Y, \nu, S) \xrightarrow{\xi} \mathbf{Z}$ is another G-extension, and that $f \in L^2(\mu)$ and $g \in \mathsf{V}_{\mathbf{Y}}$. Then applying the definition of relative satedness to the composed extension $\mathbf{Y} \stackrel{\alpha\circ\xi}{\to} \widetilde{\mathbf{X}}$ gives

$$\int_Y (f \circ \pi \circ \alpha \circ \xi) \cdot g \, \mathrm{d}\nu = \int_Y (P^{\mathsf{V}}_{\widetilde{\mathbf{X}}}(f \circ \pi) \circ \alpha \circ \xi) \cdot g \mathrm{d}\nu.$$

This turns into the required equality of inner products, once we show that

$$(P^{\mathsf{V}}_{\widetilde{\mathbf{X}}}(f \circ \pi)) \circ \alpha = P^{\mathsf{V}}_{\mathbf{Z}}(f \circ \pi \circ \alpha).$$

However, in light of the inclusion $\mathsf{V}_{\widetilde{\mathbf{X}}} \circ \alpha \subseteq \mathsf{V}_{\mathbf{Z}}$ and standard properties of orthogonal projection, this is equivalent to the equality

$$\int_Z (P^{\mathsf{V}}_{\widetilde{\mathbf{X}}}(f \circ \pi) \circ \alpha) \cdot h \, \mathrm{d}\theta = \int_Z (f \circ \pi \circ \alpha) \cdot h \, \mathrm{d}\theta \quad \forall h \in \mathsf{V}_{\mathbf{Z}},$$

and this is precisely the relative V-satedness of π applied to α. □

Lemma 3.4. *If* $\cdots \stackrel{\pi_2}{\to} \mathbf{X}_2 \stackrel{\pi_1}{\to} \mathbf{X}_1 \stackrel{\pi_0}{\to} \mathbf{X}_0$ *is an inverse sequence of G-spaces in which each π_i is relatively V-sated, and if $\mathbf{X}_\infty$, $(\psi_m)_m$ is the inverse limit of this sequence, then $\mathbf{X}_\infty$ is V-sated.*

Proof. All the resulting G-extensions $\mathbf{X}_\infty \stackrel{\psi_m}{\to} \mathbf{X}_m$ are relatively V-sated, since we may factorize $\psi_m = \pi_m \circ \psi_{m+1}$ and then apply Lemma 3.3. However, this now implies that for any further G-extension $\mathbf{Y} \stackrel{\xi}{\to} \mathbf{X}_\infty$ and for $\pi := \mathrm{id}_{\widetilde{X}}$,

$$\int_Y (f \circ \xi) \cdot g \, \mathrm{d}\nu = \int_Y ((P^{\mathsf{V}}_{\mathbf{X}_\infty} f) \circ \xi) \cdot g \mathrm{d}\nu$$

for all $g \in \mathsf{V}_{\mathbf{Y}}$ and all $f \in \bigcup_{m\geq 1}(L^2(\mu_m) \circ \psi_m)$. Since this last union is dense in $L^2(\mu_\infty)$, the result follows. □

Theorem 3.5. *If V is a functorial L^2-subspace of G-spaces, then every G-space has a V-sated extension.*

Proof. Let $\mathbf{X} = (X, \mu, T)$ be a G-space.

Step 1. We first show that $\mathbf{X}$ has a relatively V-sated extension. This uses the same "energy increment" argument used in [Aus10b].

Let $\{f_r : r \geq 1\}$ be a countable dense subset of the unit ball of $L^2(\mu)$, and let $(r_i)_{i\geq 1}$ be a member of $\mathbb{N}^{\mathbb{N}}$ in which every non-negative integer appears infinitely often.

We now construct an inverse sequence $(\mathbf{X}_m)_{m\geq 0}$, $(\psi^m_k)_{m\geq k\geq 0}$ by the following recursion. Set $\mathbf{X}_0 := \mathbf{X}$. Then, supposing that for some $m_1 \geq 0$ we have already obtained $(\mathbf{X}_m)_{m=0}^{m_1}$, $(\psi_k^m)_{m_1\geq m\geq k\geq 0}$, let $\psi_{m_1}^{m_1+1} : \mathbf{X}_{m_1+1} \to \mathbf{X}_{m_1}$ be an extension such that the difference

$$\|P^{\mathsf{V}}_{\mathbf{X}_{m_1+1}}(f_{r_{m_1}} \circ \psi_0^{m_1+1})\|_2 - \|P^{\mathsf{V}}_{\mathbf{X}_{m_1}}(f_{r_{m_1}} \circ \psi_0^{m_1})\|_2$$

is at least half its supremal possible value over all extensions of $\mathbf{X}_{m_1}$, where of course we let $\psi_0^{m_1+1} := \psi_0^{m_1} \circ \psi_{m_1}^{m_1+1}$.

Let $\mathbf{X}_\infty$, $(\psi_m)_{m\geq 0}$ be the inverse limit of this sequence. We show that $\mathbf{X}_\infty \stackrel{\psi_0}{\to} \mathbf{X}$ is relatively V-sated. Letting $\pi : \mathbf{Y} \to \mathbf{X}_\infty$ be an arbitrary further extension, we see that this is equivalent to showing that

$$P^{\mathsf{V}}_{\mathbf{Y}}(f \circ \psi_0 \circ \pi) = P^{\mathsf{V}}_{\mathbf{X}_\infty}(f \circ \psi_0) \circ \pi \quad \forall f \in L^2(\mu).$$

It suffices to prove this for every f_r in our previously chosen dense subset. Also, since $\mathsf{V}_{\mathbf{Y}} \supseteq \mathsf{V}_{\mathbf{X}_\infty} \circ \pi$, the result follows from

$$\|P^{\mathsf{V}}_{\mathbf{Y}}(f_r \circ \psi_0 \circ \pi)\|_2 \leq \|P^{\mathsf{V}}_{\mathbf{X}_\infty}(f_r \circ \psi_0)\|_2.$$

To prove this inequality, suppose, towards a contradiction, that the left-hand side is strictly larger than the right hand side. The sequence of norms $\|P^{\mathsf{V}}_{\mathbf{X}_m}(f_r \circ \psi_0^m)\|_2$ is non-decreasing as $m \to \infty$, and bounded above by $\|f_r\|_2$. Therefore, for some sufficiently large m, $r_m = r$, since each integer appears infinitely often as some r_m. But also

$$\begin{aligned}\|P^{\mathsf{V}}_{\mathbf{X}_{m+1}}(f_r \circ \psi_0^{m+1})\|_2 - \|P^{\mathsf{V}}_{\mathbf{X}_m}(f_r \circ \psi_0^m)\|_2 &\\ &< \frac{1}{2}\Big(\|P^{\mathsf{V}}_{\mathbf{Y}}(f_r \circ \psi_0 \circ \pi)\|_2 - \|P^{\mathsf{V}}_{\mathbf{X}_\infty}(f \circ \psi_0)\|_2\Big)\\ &\leq \frac{1}{2}\Big(\|P^{\mathsf{V}}_{\mathbf{Y}}(f_r \circ \psi_0 \circ \pi)\|_2 - \|P^{\mathsf{V}}_{\mathbf{X}_m}(f \circ \psi_0^m)\|_2\Big).\end{aligned}$$

This contradicts the choice of $\mathbf{X}_{m+1} \to \mathbf{X}_m$ in our construction above. Thus we must actually have equality of L^2-norms, as required.

Step 2. Iterating the construction of Step 1, we may let $\cdots \stackrel{\pi_2}{\to} \mathbf{X}_2 \stackrel{\pi_1}{\to} \mathbf{X}_1 \stackrel{\pi_0}{\to} \mathbf{X}$ be an inverse seqeuence in which each extension π_i is relatively V-sated. Letting $\mathbf{X}_\infty$, $(\pi_m)_{m\geq 0}$ be its inverse limit, and applying Lemma 3.4 complete the proof. □

Corollary 3.6. *Let* V_1, V_2, ... *be a countable family of functorial* L^2*-subspaces of* G*-spaces. Then every* G*-space has an extension which is simultaneously* V_r*-sated for every* r.

Proof. Let $(r_i)_i$ be an element of $\mathbb{N}^{\mathbb{N}}$ in which every positive integer appears infinitely often. Applying Theorem 3.5 repeatedly, we obtain an inverse sequence

$\cdots \xrightarrow{\pi_2} \mathbf{X}_2 \xrightarrow{\pi_1} \mathbf{X}_1 \xrightarrow{\pi_0} \mathbf{X}$ in which each $\mathbf{X}_i$ is V_{r_i}-sated. Also, let $\pi^n_m := \pi_m \circ \cdots \circ \pi_{n-1}$ whenever $m < n$. Finally, let $\mathbf{X}_\infty$ be the inverse limit of this sequence. Then for each $r \geq 1$, there exists an infinite subsequence $i_1(r) < i_2(r) < \ldots$ in $\mathbb{N}$ such that $r_{i_1(r)} = r_{i_2(r)} = \cdots = r$, and $\mathbf{X}_\infty$ may be identified with the inverse limit of the thinned-out inverse sequence $\cdots \xrightarrow{\pi^{i_3(r)}_{i_2(r)}} \mathbf{X}_{i_2(r)} \xrightarrow{\pi^{i_2(r)}_{i_1(r)}} \mathbf{X}_{i_1(r)} \xrightarrow{\pi^{i_1(r)}_0} \mathbf{X}$. Lemma 3.4 implies that $\mathbf{X}_\infty$ is V_r-sated. Since r is arbitrary, this completes the proof. □

4 Characteristic subspaces and proof of convergence

4.1 Subgroups associated to commuting tuples of actions. We now begin to work with commuting tuples of G-actions. We need to call on several different subgroups of G^d in the sequel, so the next step is to set up some bespoke notation for handling them. We sometimes use a boldface $\mathbf{g}$ to denote a tuple $(g_i)_{i=1}^d$ in G^d, and denote the identity element of G by 1_G.

Fix G and d, and let $e = \{i_1 < \ldots < i_r\} \subseteq [d]$ with $r \geq 2$ and $\{i < j\} \subseteq [d]$. Define

$$H_e := \{\mathbf{g} \in G^d : g_{i_s+1} = g_{i_s+2} = \ldots = g_{i_{s+1}} \text{ for each } s = 1, \ldots, r-1\},$$
$$K_{\{i,j\}} := \{\mathbf{g} \in H_{\{i,j\}} : g_\ell = 1_G \text{ for all } \ell \in (i;j]\},$$

and

$$L_e := \{\mathbf{g} \in H_e : g_i = 1_G \text{ for all } i \in [d] \setminus (i_1; i_r]\}.$$

Routine calculations give the following basic properties.

Lemma 4.1. (1) *The subgroups L_e and $K_{\{i_1,i_r\}}$ commute and generate H_e.*
(2) *If $a \subseteq e \subseteq [k]$ with $|a| \geq 2$, then $L_a \leq L_e$.*
(3) *If $a \subseteq e \subseteq [k]$ with $|a| \geq 2$ and $e \cap [\min a; \max a] = a$, then $L_a \trianglelefteq H_e$. In particular, $L_e \trianglelefteq H_e$.*

Part (3) of this lemma has the following immediate consequence.

Corollary 4.2. *If $a \subseteq e \subseteq [k]$ with $|a| \geq 2$, and $e \cap [\min a; \max a] = a$, then $\Sigma^{L_a}_{\mathbf{X}}$ is globally H_e-invariant.*

4.2 The Host-Kra inequality. In order to show that a suitably-sated G-space has some other desirable property, one must find an extension of it for which the relative independence given by satedness implies that other property. The key to such a proof is usually constructing the right extension.

Where satedness was used in the previous works [Aus09] and [Aus10a], that extension could be constructed directly from the Furstenberg self-joining arising from some non-conventional averages. However, this seems to be more problematic in the present setting, and we take a different approach. The construction below is a close analog of the construction by Host and Kra of certain "cubical" extensions of a $\mathbb{Z}$-space in [HK05]. That machinery has also been extended by Host to commuting tuples of $\mathbb{Z}$-actions in [Hos09].

Fix now a G^d-space $\mathbf{X} = (X, \mu, T_1, \dots, T_d)$, and let $\mathbf{Y}^{(0)} := \mathbf{X}$. Our next step is to construct recursively a height-$(d+1)$ tower of new probability-preserving G^d-spaces, which we denote by

$$\mathbf{Y}^{(d)} \stackrel{\xi^{(d)}}{\to} \mathbf{Y}^{(d-1)} \stackrel{\xi^{(d-1)}}{\to} \cdots \stackrel{\xi^{(2)}}{\to} \mathbf{Y}^{(1)} \stackrel{\xi^{(1)}}{\to} \mathbf{Y}^{(0)} = \mathbf{X}. \tag{4.1}$$

The construction also gives some other auxiliary G^d-spaces $\mathbf{Z}^{(j)}$, and they too are used later.

Supposing the tower has already been constructed up to some level $j \leq d-1$, the next extension is constructed in the following steps.

(i) From $\mathbf{Y}^{(j)} = (Y^{(j)}, \nu^{(j)}, S^{(j)})$, define a new $H_{\{d-j-1,d\}}$-action $\widetilde{S}^{(j)}$ on the same space by setting

$$(\widetilde{S}_i^{(j)})^g := (S_i^{(j)})^g \quad \forall g \in G,\ i < d-j-1, \tag{4.2a}$$

$$(\widetilde{S}_{d-j-1}^{(j)})^g := (S_{[d-j-1;d]}^{(j)})^g \quad \text{for all } g \in G, \tag{4.2b}$$

and

$$(\widetilde{S}_{(d-j-1;d]}^{(j)})^g := \mathrm{id} \quad \text{for all } g \in G \tag{4.2c}$$

(with the understanding that (4.2a) and (4.2b) are vacuous in case $j = d-1$).

(ii) Now consider the $H_{\{d-j-1,d\}}$-space

$$\begin{aligned}\mathbf{Z}^{(j+1)} &= (Z^{(j+1)}, \theta^{(j+1)}, R^{(j+1)}) \\ &:= \left(Y^{(j)} \times Y^{(j)},\ \nu^{(j)} \otimes_{\Sigma_{\mathbf{Y}^{(j)}}^{L_{\{d-j-1,d\}}}} \nu^{(j)},\ (S^{(j)})^{\upharpoonright H_{\{d-j-1,d\}}} \times \widetilde{S}^{(j)}\right).\end{aligned}$$

Let $\xi_0^{(j+1)}, \xi_1^{(j+1)} : Z^{(j+1)} \to Y^{(j)}$ be the two coordinate projections. They are both factor maps of $H_{\{d-j-1,d\}}$-spaces. Notice that $\theta^{(j+1)}$ is $R^{(j+1)}$-invariant because both of the actions $(S^{(j)})^{\upharpoonright H_{\{d-j-1,d\}}}$ and $\widetilde{S}^{(j)}$ preserve the σ-subalgebra $\Sigma_{\mathbf{Y}^{(j)}}^{L_{\{d-j-1,d\}}}$, by Corollary 4.2.

(iii) Finally, let $\mathbf{Y}^{(j+1)} \stackrel{\xi^{(j+1)}}{\to} \mathbf{Y}^{(j)}$ be an extension of G^d-spaces for which there exists a commutative diagram

$$\begin{tikzcd}[column sep=small] (\mathbf{Y}^{(j+1)})^{\upharpoonright H_{\{d-j-1,d\}}} \arrow[rr, "\xi^{(j+1)}"] \arrow[dr, "\alpha^{(j+1)}"'] & & (\mathbf{Y}^{(j)})^{\upharpoonright H_{\{d-j-1,d\}}} \\ & \mathbf{Z}^{(j+1)}, \arrow[ur, "\xi_0^{(j+1)}"'] & \end{tikzcd}$$

as provided by Theorem 2.1.

Having made this construction, for each $j \in \{1, 2, \ldots, d\}$ we also define a family of maps $\pi_\eta^{(j)} : Y^{(j)} \to X$ indexed by $\eta \in \{0, 1\}^j$, by setting

$$\pi^{(j)}_{(\eta_1,\ldots,\eta_j)} := \xi^{(1)}_{\eta_1} \circ \alpha^{(1)} \circ \xi^{(2)}_{\eta_2} \circ \alpha^{(2)} \circ \cdots \circ \xi^{(j)}_{\eta_j} \circ \alpha^{(j)}.$$

Clearly, $(\pi_\eta^{(j)})_* \nu^{(j)} = \mu$ for every η. Also, $\pi^{(j)}_{0^j} = \xi^{(1)} \circ \cdots \circ \xi^{(j)} : \mathbf{Y}^{(j)} \to \mathbf{X}$ is a factor map of G^d-spaces, where $0^j := (0, 0, \ldots, 0) \in \{0, 1\}^j$.

Lemma 4.3. *Let $r \in [d]$, let $\eta \in \{0, 1\}^r \setminus \{0^r\}$, and let $\ell \in [r]$ be maximal such that $\eta_\ell = 1$. Then $\pi_\eta^{(r)}$ satisfies the intertwining relations*

$$\pi_\eta^{(r)} \circ S_i^{(r)} = T_i \circ \pi_\eta^{(r)} \quad \forall i < d - \ell, \tag{4.3a}$$

$$\pi_\eta^{(r)} \circ S_{d-\ell}^{(r)} = T_{[d-\ell;d]} \circ \pi_\eta^{(r)} \tag{4.3b}$$

and

$$\pi_\eta^{(r)} \circ S_{(d-\ell;d]}^{(r)} = \pi_\eta^{(r)}. \tag{4.3c}$$

Remark. There are no such simple relations for the compositions $\pi_\eta^{(r)} \circ S_i^{(r)}$ when $i \geq d - \ell + 1$, but we do not need these.

Proof. By the definition of ℓ, for this η we may write $\pi_\eta^{(r)} = \pi' \circ \pi''$, where

$$\pi' := \pi^{(\ell)}_{(\eta_1,\ldots,\eta_\ell)} = \xi^{(1)}_{\eta_1} \circ \alpha^{(1)} \circ \xi^{(2)}_{\eta_2} \circ \alpha^{(2)} \circ \cdots \circ \xi^{(\ell)}_1 \circ \alpha^{(\ell)} \tag{4.4a}$$

and

$$\pi'' := \xi^{(\ell+1)} \circ \cdots \circ \xi^{(r)}. \tag{4.4b}$$

All three of the desired relations concern the actions of subgroups of $H_{\{d-\ell,d\}}$, and all the maps in the compositions in (4.4) are factor maps of $H_{\{d-\ell,d\}}$-spaces. We read off the desired results from the simpler relations (4.2a), (4.2b), and (4.2c).

First, observe that by construction. each $\xi^{(j)}$ appearing in the definition of π'' actually intertwines the whole G^d-actions, so $\pi'' \circ S_i^{(r)} = S_i^{(\ell)} \circ \pi''$ for all $i \in [d]$. It therefore suffices to prove that $\pi' \circ S_i^{(\ell)} = T_i \circ \pi'$ for all $i < d - \ell$, and similarly for the other two desired relations.

Step 1. Suppose that $i \le d-\ell$ and $j \le \ell$. Then the definitions of $\alpha^{(j)}$, $\xi_0^{(j)}$, and $\xi_1^{(j)}$ give

$$\xi_\eta^{(j)} \circ \alpha^{(j)} \circ S_i^{(j)} = \xi_\eta^{(j)} \circ R_i^{(j)} \circ \alpha^{(j)} = \begin{cases} S_i^{(j-1)} \circ \xi_\eta^{(j)} \circ \alpha^{(j)} & \text{if } \eta = 0, \\ \widetilde{S}_i^{(j-1)} \circ \xi_\eta^{(j)} \circ \alpha^{(j)} & \text{if } \eta = 1. \end{cases}$$

In case $i < d-\ell \le d-j$, this equals $S_i^{(j)} \circ \xi_\eta^{(j)} \circ \alpha^{(j)}$ for either value of η, by (4.2a). Applying this repeatedly for $j = \ell, \ell-1, \dots, 1$ in the composition that defines π', we obtain $\pi' \circ S_i^{(\ell)} = T_i \circ \pi'$. As explained above, this proves (4.3a).

Step 2. The same calculation as above gives

$$\xi_1^{(\ell)} \circ \alpha^{(\ell)} \circ S_{d-\ell}^{(\ell)} = \xi_1^{(\ell)} \circ R_{d-\ell}^{(\ell)} \circ \alpha^{(\ell)} = \widetilde{S}_{d-\ell}^{(\ell-1)} \circ \xi_1^{(\ell)} \circ \alpha^{(\ell)},$$

and now this is equal to $S_{[d-\ell,d]}^{(\ell-1)} \circ \xi_1^{(\ell)} \circ \alpha^{(\ell)}$, by (4.2b).

On the other hand, if $j \le \ell - 1$, then another call to the definitions of $\alpha^{(j)}$, $\xi_0^{(j)}$ and $\xi_1^{(j)}$ gives

$$\xi_\eta^{(j)} \circ \alpha^{(j)} \circ S_{[d-\ell,d]}^{(j)} = \xi_\eta^{(j)} \circ R_{[d-\ell,d]}^{(j)} \circ \alpha^{(j)} = \begin{cases} S_{[d-\ell;d]}^{(j-1)} \circ \xi_\eta^{(j)} \circ \alpha^{(j)} & \text{if } \eta = 0, \\ \widetilde{S}_{[d-\ell;d]}^{(j-1)} \circ \xi_\eta^{(j)} \circ \alpha^{(j)} & \text{if } \eta = 1. \end{cases}$$

This time, since $j \le \ell - 1$, (4.2a) and (4.2b) give

$$\begin{aligned} \widetilde{S}_{[d-\ell;d]}^{(j-1)} &= \widetilde{S}_{d-\ell}^{(j-1)} \circ \widetilde{S}_{d-\ell+1}^{(j-1)} \circ \cdots \circ \widetilde{S}_{d-j}^{(j-1)} \circ \widetilde{S}_{(d-j;d]}^{(j-1)} \\ &= S_{d-\ell}^{(j-1)} \circ S_{d-\ell+1}^{(j-1)} \circ \cdots \circ S_{[d-j;d]}^{(j-1)} \circ \mathrm{id} = S_{[d-\ell;d]}^{(j-1)}. \end{aligned}$$

Therefore, $\xi_\eta^{(j)} \circ \alpha^{(j)} \circ S_{[d-\ell,d]}^{(j)} = S_{[d-\ell,d]}^{(j-1)} \circ \xi_\eta^{(j)} \circ \alpha^{(j)}$ for all $j \le \ell - 1$ and either value of η. Combining these two calculations gives

$$\pi' \circ S_{d-\ell}^{(\ell)} = (\xi_{\eta_1}^{(1)} \circ \alpha^{(1)} \circ \cdots \circ \xi_{\eta_{\ell-1}}^{(\ell-1)} \circ \alpha^{(\ell-1)}) \circ S_{[d-\ell;d]}^{(\ell-1)} = T_{[d-\ell;d]} \circ \pi',$$

and hence (4.3b).

Step 3. Finally, (4.2c) gives

$$\xi_1^{(\ell)} \circ \alpha^{(1)} \circ S_{(d-\ell;d]}^{(\ell)} = \widetilde{S}_{(d-\ell;d]}^{(\ell-1)} \circ \xi_1^{(\ell)} \circ \alpha^{(1)} = \xi_1^{(\ell)} \circ \alpha^{(1)},$$

from which (4.3c) follows immediately. □

Corollary 4.4. *If $r \in [d]$, $\eta \in \{0,1\}^r$, and if $j \in [r]$ is such that $\eta_i = 0$ for all $i \ge j+1$, then $\pi_\eta^{(r)}$ satisfies the intertwining relations*

$$\pi_\eta^{(r)} \circ S_{[d-j;d]}^{(r)} = T_{[d-j;d]} \circ \pi_\eta^{(r)}. \tag{4.5}$$

We next prove an estimate relating the multi-linear forms Λ_n in (1.1) to certain integrals over these new G^d-spaces $\mathbf{Y}^{(j)}$. This is the key estimate which enables an appeal to satedness. The following theorem relies on an iterated application of the van der Corput estimate, and follows essentially the same lines as [HK05, Theorem 12.1].

Theorem 4.5. *Let $\mathbf{X} = (X, \mu, T_1, \dots, T_d)$ be a G^d-space, let $1 \le j \le d$, and let the tower* (4.1) *and the maps $\pi^{(j)}_\eta : Y^{(j)} \to X$ for $\eta \in \{0,1\}^j$ be constructed as above. For $f_{d-j+1}, \dots, f_d \in L^\infty(\mu)$, let*

$$\Lambda^{(j)}_n(f_{d-j+1}, \dots, f_d) := \frac{1}{|F_n|}\sum_{g\in F_n}\prod_{i=d-j+1}^{d}(f_i \circ T^g_{[d-j+1;i]}).$$

If $f_{d-j+1}, \dots, f_d$ are all uniformly bounded by 1*, then*

$$\limsup_{n\to\infty}\|\Lambda^{(j)}_n(f_{d-j+1}, \dots, f_d)\|_2 \le \Big(\int_{Y^{(j)}}\prod_{\eta\in\{0,1\}^j}(\mathcal{C}^{|\eta|}f_d \circ \pi^{(j)}_\eta)\,\mathrm{d}\nu^{(j)}\Big)^{2^{-j}},$$

where $|\eta| := \sum_i \eta_i \mod 1$ and $\mathcal{C}$ is the operator of complex conjugation.

Note that $\Lambda^{(d)}_n = \Lambda_n$, the averages in (1.1). The integral appearing on the right-hand side of the last inequality actually defines a seminorm of the function f_d: these are the adaptations of the Host-Kra seminorms to the present setting. However, our approach does not emphasize the seminorm axioms.

Proof. This is proved by induction on j.

Step 1: base case. When $j = 1$, the Norm Ergodic Theorem for amenable groups gives

$$\Lambda^{(1)}_n(f_d) \to \mathsf{E}_\mu(f_d \mid \Sigma^{T_d}_{\mathbf{X}}) = \mathsf{E}_\mu(f_d \mid \Sigma^{L_{\{d-1,d\}}}_{\mathbf{X}}) \quad \text{in } \|\cdot\|_2,$$

and the square of the norm of this limit equals

$$\begin{aligned}\int_{X\times X}(f_d\otimes\overline{f_d})\mathrm{d}\big(\mu\otimes_{\Sigma^{L_{\{d-1,d\}}}_{\mathbf{X}}}\mu\big) &= \int_{Z^{(1)}} f_d\circ\xi^{(1)}_0\cdot\overline{f_d\circ\xi^{(1)}_1}\mathrm{d}\theta^{(1)}\\ &= \int_{Y^{(1)}} f_d\circ\pi^{(1)}_0\cdot\overline{f_d\circ\pi^{(1)}_1}\mathrm{d}\nu^{(1)},\end{aligned}$$

by the definition of $Y^{(1)}$ and $\nu^{(1)}$.

Step 2: Van der Corput estimate. Now suppose the result is known up to some $j-1 \in \{1, 2, \dots, d-1\}$.

By the amenable-groups version of the van der Corput estimate ([BMZ97, Lemma 4.2]),

$$\limsup_{n\to\infty} \|\Lambda_n^{(j)}(f_{d-j+1},\dots,f_d)\|_2^2 \tag{4.6}$$
$$\leq \limsup_{m\to\infty} \frac{1}{|F_m|^2} \sum_{h,k\in F_m} \limsup_{n\to\infty} \Big|\frac{1}{|F_n|}\sum_{g\in F_n}\int_X \prod_{i=d-j+1}^{d} (f_i\circ T^{hg}_{[d-j+1;i]})\overline{(f_i\circ T^{kg}_{[d-j+1;i]})}\mathrm{d}\mu\Big|$$

For fixed h and k, we may use the T^g_{d-j+1}-invariance of μ to re-arrange the above integral as follows:

$$\Big|\frac{1}{|F_n|}\sum_{g\in F_n}\int_X \prod_{i=d-j+1}^{d} (f_i\circ T^{hg}_{[d-j+1;i]})\overline{(f_i\circ T^{kg}_{[d-j+1;i]})}\mathrm{d}\mu\Big|$$
$$= \Big|\frac{1}{|F_n|}\sum_{g\in F_n}\int_X (f_{d-j+1}\circ T^h_{d-j+1})\cdot\overline{(f_{d-j+1}\circ T^k_{d-j+1})}$$
$$\cdot\Big(\prod_{i=d-j+2}^{d}((f_i\circ T^h_{[d-j+1;i]})\cdot\overline{(f_i\circ T^k_{[d-j+1;i]})})\circ T^g_{(d-j+1;i]}\Big)\mathrm{d}\mu\Big|.$$

(At this point we have made crucial use of the commutativity of the different actions T_i.) By the Cauchy-Bunyakowski-Schwartz Inequality, this, in turn, is bounded above by

$$\|(f_{d-j+1}\circ T^h_{d-j+1})\cdot\overline{(f_{d-j+1}\circ T^k_{d-j+1})}\|_2$$
$$\cdot\Big\|\frac{1}{|F_n|}\sum_{g\in F_n}\prod_{i=d-j+2}^{d}((f_i\circ T^h_{[d-j+1;i]})\cdot\overline{(f_i\circ T^k_{[d-j+1;i]})})\circ T^g_{(d-j+1;i]}\Big\|_2$$
$$\leq \Big\|\Lambda_n^{(j-1)}\big((f_{d-j+2}\circ T^h_{[d-j+1;d-j+2]})\cdot\overline{(f_{d-j+2}\circ T^k_{[d-j+1;d-j+2]})},$$
$$\dots,(f_d\circ T^h_{[d-j+1;d]})\cdot\overline{(f_d\circ T^k_{[d-j+1;d]})}\big)\Big\|_2$$

(since $\|f_{d-j+1}\|_\infty \leq 1$).

Step 3: use of inductive hypothesis. Combining these inequalities and using the inductive hypothesis, we obtain

$$\limsup_{n\to\infty}\Big|\frac{1}{|F_n|}\sum_{g\in F_n}\int_X \prod_{i=d-j+1}^{d}(f_i\circ T^{hg}_{[d-j+1;i]})\overline{(f_i\circ T^{kg}_{[d-j+1;i]})}\,\mathrm{d}\mu\Big| \tag{4.7}$$
$$\leq \Big(\int_{Y^{(j-1)}}\prod_{\eta\in\{0,1\}^{j-1}}\big(\mathfrak{C}^{|\eta|}((f_d\circ T^h_{[d-j+1;d]})\cdot\overline{(f_d\circ T^k_{[d-j+1;d]})})\circ\pi^{(j-1)}_\eta\big)\,\mathrm{d}\nu^{(j-1)}\Big)^{2^{-(j-1)}}$$

for each h and k.

To simplify notation, now let $F := \prod_{\eta\in\{0,1\}^{j-1}}(\mathcal{C}^{|\eta|}f_d\circ\pi_\eta^{(j-1)})$. In terms of this function, Corollary 4.4 with $r := j-1$ allows us to write

$$\prod_{\eta\in\{0,1\}^{j-1}}(\mathcal{C}^{|\eta|}(f_d\circ T^h_{[d-j+1;d]})\circ\pi_\eta^{(j-1)}) \\ = \prod_{\eta\in\{0,1\}^{j-1}}(\mathcal{C}^{|\eta|}f_d\circ\pi_\eta^{(j-1)}\circ(S^{(j-1)}_{[d-j+1;d]})^h) = F\circ(S^{(j+1)}_{[d-j+1;d]})^h;$$

and similarly,

$$\prod_{\eta\in\{0,1\}^{j-1}}(\mathcal{C}^{|\eta|}\overline{(f_d\circ T^k_{[d-j+1;d]})}\circ\pi_\eta^{(j-1)}) = \overline{F}\circ(S^{(j-1)}_{[d-j+1;d]})^k.$$

Step 4: completion of the proof. Substituting the formulas obtained in Step 3 into the right-hand side of (4.7), one obtains

$$\limsup_{n\to\infty}\Big|\frac{1}{|F_n|}\sum_{g\in F_n}\int_X\prod_{i=d-j+1}^{d}(f_i\circ T^{hg}_{[d-j+1;i]})\overline{(f_i\circ T^{kg}_{[d-j+1;i]})}\,\mathrm{d}\mu\Big| \\ \leq\Big(\int_{Y^{(j-1)}}(F\circ(S^{(j+1)}_{[d-j+1;d]})^h)\cdot(\overline{F}\circ(S^{(j-1)}_{[d-j+1;d]})^k)\,\mathrm{d}\nu^{(j-1)}\Big)^{2^{-(j-1)}}.$$

Inserting this back into (4.6) and using Hölder's inequality for the average over (h,k) yield

$$\limsup_{n\to\infty}\|\Lambda_n^{(j)}(f_{d-j+1},\dots,f_d)\|_2^2 \\ \leq\limsup_{m\to\infty}\frac{1}{|F_m|^2}\sum_{h,k\in F_m}\Big(\int_{Y^{(j-1)}}(F\circ(S^{(j+1)}_{[d-j+1;d]})^h)\cdot(\overline{F}\circ(S^{(j-1)}_{[d-j+1;d]})^k)\,\mathrm{d}\nu^{(j-1)}\Big)^{2^{-(j-1)}} \\ \leq\limsup_{m\to\infty}\Big(\frac{1}{|F_m|^2}\sum_{h,k\in F_m}\int_{Y^{(j-1)}}(F\circ(S^{(j+1)}_{[d-j+1;d]})^h)\cdot(\overline{F}\circ(S^{(j-1)}_{[d-j+1;d]})^k)\,\mathrm{d}\nu^{(j-1)}\Big)^{2^{-(j-1)}}.$$

Finally, by the Norm Ergodic Theorem for amenable groups, the averages on the last line here converge as $m\to\infty$, giving

$$\limsup_{n\to\infty}\|\Lambda_n^{(j)}(f_{d-j+1},\dots,f_d)\|_2^2 \\ \leq\Big(\int_{Y^{(j-1)}}\mathsf{E}_{\nu^{(j-1)}}\big(F\bigm|\Sigma^{L_{\{d-j,d\}}}_{\mathbf{Y}^{(j-1)}}\big)\cdot\mathsf{E}_{\nu^{(j-1)}}\big(\overline{F}\bigm|\Sigma^{L_{\{d-j,d\}}}_{\mathbf{Y}^{(j-1)}}\big)\,\mathrm{d}\nu^{(j-1)}\Big)^{2^{-(j-1)}} \\ =\Big(\int_{Z^{(j)}}F\circ\xi_0^{(j)}\cdot\overline{F\circ\xi_1^{(j)}}\,\mathrm{d}\theta^{(j)}\Big)^{2^{-(j-1)}} \\ =\Big(\int_{Y^{(j)}}\Big(\prod_{\eta\in\{0,1\}^j}\mathcal{C}^{|\eta|}f_d\circ\pi_\eta^{(j)}\Big)\,\mathrm{d}\nu^{(j)}\Big)^{2^{-(j-1)}},$$

where $\mathbf{Z}^{(j)}$ is the auxiliary $H_{\{d-j,d\}}$-space constructed along with $\mathbf{Y}^{(j)}$. Taking square-roots, continues the induction. $\square$

4.3 Partially characteristic subspaces and the proof of convergence.

Definition 4.6. Consider a probability space (X, μ), and a sequence Ξ_n of multi-linear forms on $L^\infty(\mu)$ which are separately continuous for the norm $\|\cdot\|_2$ in each entry. A closed subspace $V \leq L^2(\mu)$ is **partially characteristic in position** i for the sequence Ξ_n if

$$\big\|\Xi_n(f_1, \dots, f_d) - \Xi_n(f_1, \dots, f_{i-1}, P^V f_i, f_{i+1}, \dots, f_d)\big\|_2 \to 0$$

as $n \to \infty$ for all $f_1, \dots, f_d \in L^\infty(\mu)$, where P^V is the orthogonal projection onto V.

The following proposition quickly leads to a proof of Theorem A. In fact, it gives rather more than one needs for the proof of Theorem A, but that extra strength is used during the proof of Theorem B.

Proposition 4.7. *For $1 \leq i \leq j \leq d$, let $\mathsf{F}_{i,j}$ be the functorial σ-algebra*

$$\Sigma_{\mathbf{X}}^{\mathsf{F}_{i,j}} := \bigvee_{\ell=0}^{i-1} \Sigma_{\mathbf{X}}^{T_{(\ell;i]}} \vee \bigvee_{\ell=i+1}^{j} \Sigma_{\mathbf{X}}^{T_{(i;\ell]}},$$

and let $\mathsf{V}_{i,j,\mathbf{X}} := L^2(\mu \,|\, \Sigma_{\mathbf{X}}^{\mathsf{F}_{i,j}})$ be the associated functorial L^2-subspace. Let

$$\widehat{\Lambda}_n^{(j)}(f_1, \dots, f_j) := \frac{1}{|F_n|} \sum_{g \in F_n} \prod_{i=1}^{j} (f_i \circ T_{[1;i]}^g).$$

If $\mathbf{X}$ is $\mathsf{V}_{i,j}$-sated whenever $1 \leq i \leq j \leq d$, then, for each $j \in [d]$, the subspaces

$$\mathsf{V}_{1,j,\mathbf{X}}, \; \dots, \; \mathsf{V}_{j,j,\mathbf{X}}$$

are partially characteristic in positions $1, \dots, j$ for the averages $\widehat{\Lambda}_n^{(j)}$.

Notice that we still have $\widehat{\Lambda}_n^{(d)} = \Lambda_n$ (the averages in (1.1)), but otherwise these averages differ from the averages $\Lambda_n^{(j)}$ considered in Theorem 4.5.

Proof. This is proved by induction on j. When $j = 1$, also $i = 1$, and $\Sigma_{\mathbf{X}}^{\mathsf{F}_{i,j}} = \Sigma_{\mathbf{X}}^{T_1}$. This is always partially characteristic because the Norm Ergodic Theorem gives

$$\widehat{\Lambda}_n^{(1)}(f_1) \to \mathsf{E}_\mu(f_1 \,|\, \Sigma_{\mathbf{X}}^{T_1}) \quad \text{in } \|\cdot\|_2$$

for any G-space. So now we focus on the recursion clause. For this it clearly suffices to assume $j = d - 1$, and prove the result for the averages $\widehat{\Lambda}_n^{(d)}$, which simplifies the notation.

Step 1. We first show that $\mathsf{V}_{d,d}$ is partially characteristic in position d.

Let $f_d \in L^\infty(\mu)$. Decomposing it as $P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d + (f_d - P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d)$ and using the multi-linearity of $\widehat{\Lambda}^{(d)}_n$, we see that it suffices to show that $P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d = 0$ implies $\|\widehat{\Lambda}^{(d)}_n(f_1, \dots, f_d)\|_2 \to 0$ for all $f_1, \dots, f_{d-1} \in L^\infty(\mu)$; equivalently, if there exist $f_1, \dots f_d \in L^\infty(\mu)$ such that $\|\widehat{\Lambda}^{(d)}_n(f_1, \dots, f_d)\|_2 \not\to 0$, then $P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d \neq 0$.

So suppose that $\limsup_{n\to\infty} \|\widehat{\Lambda}^{(d)}_n(f_1, \dots, f_d)\|_2 > 0$ for some $f_1, \dots, f_{d-1}$. Then Theorem 4.5 gives $\int_{Y^{(d)}} \Big(\prod_{\eta \in \{0,1\}^d} \mathcal{C}^{|\eta|} f_d \circ \pi^{(d)}_\eta \Big)\, d\nu^{(d)} \neq 0$. However, recalling relation (4.3c) from Lemma 4.3, we see that if $\eta \in \{0,1\}^d \setminus \{0^d\}$ and $\ell \in [d]$ is maximal such that $\eta_\ell \neq 0$, then $\mathcal{C}^{|\eta|} f_d \circ \pi^{(d)}_\eta \circ S^g_{(d-\ell;d]} = \mathcal{C}^{|\eta|} f_d \circ \pi^{(d)}_\eta$ for all $g \in G$, and so the function $\prod_{\eta \in \{0,1\}^d \setminus \{0^d\}} \mathcal{C}^{|\eta|} f_d \circ \pi^{(d)}_\eta$ is measurable with respect to $\bigvee_{\ell=1}^d \Sigma^{S^{(d)}_{(d-\ell;d]}}_{\mathbf{Y}^{(d)}} = \Sigma^{\mathsf{F}_{d,d}}_{\mathbf{Y}^{(d)}}$.

Therefore the non-vanishing of the above integral implies that $P^{\mathsf{V}_{d,d}}_{\mathbf{Y}^{(d)}}(f_d \circ \pi^{(d)}_{0^d}) \neq 0$. Since $\mathbf{X}$ is $\mathsf{V}_{d,d}$-sated, this implies that also $P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d \neq 0$, as required. This proves that the required subspace is partially characteristic for $i = j = d$.

Step 2. By Step 1, $\widehat{\Lambda}^{(d)}_n(f_1, \dots, f_d) - \widehat{\Lambda}^{(d)}_n(f_1, \dots, f_{d-1}, P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d) \to 0$ for all $f_1, \dots, f_d \in L^\infty(\mu)$. Also, $P^{\mathsf{V}_{d,d}}_{\mathbf{X}} f_d$ still lies in $L^\infty(\mu)$, because $P^{\mathsf{V}_{d,d}}_{\mathbf{X}}$ is actually a conditional expectation operator. It therefore suffices to check that the required factors are partially characteristic in the other positions under the additional assumption that f_d is $\Sigma^{\mathsf{F}_{d,d}}_{\mathbf{X}}$-measurable.

This assumption implies that f_d may be approximated in $\|\cdot\|_2$ by a finite sum of products of the form

$$h_0 \cdot \dots \cdot h_{d-1}, \tag{4.8}$$

where h_i is $\Sigma^{T_{(i;d]}}_{\mathbf{X}}$-measurable for each i. By multi-linearity, it therefore suffices to prove that the required factors are partially characteristic in the other positions when f_d is just one such product function. However, at this point, a simple rearrangement and the partial invariances of each of the h_is give

$$\begin{aligned}\widehat{\Lambda}^{(d)}_n(f_1, \dots, f_{d-1}, h_0 \cdots h_{d-1}) &= \frac{1}{|F_n|} \sum_{g \in F_n} \Big(\prod_{i=1}^{d-1} (f_i \circ T^g_{[1;i]}) \Big) \cdot ((h_0 \cdots h_{d-1}) \circ T^g_{[1;d]}) \\ &= h_0 \cdot \frac{1}{|F_n|} \sum_{g \in F_n} \prod_{i=1}^{d-1} ((f_i h_i) \circ T^g_{[1;i]}) \\ &= h_0 \cdot \widehat{\Lambda}^{(d-1)}_n(f_1 h_1, \dots, f_{d-1} h_{d-1}).\end{aligned}$$

Therefore, by the inductive hypothesis for $j = d-1$, if these averages do not vanish as $n \to \infty$, then $P^{\mathsf{V}_{i,d-1}}_{\mathbf{X}}(f_i h_i) = \mathsf{E}_\mu\big(f_i h_i \,\big|\, \Sigma^{\mathsf{F}_{i,d-1}}_{\mathbf{X}}\big) \neq 0$ for each $i \in [d-1]$. Since h_i is $\Sigma^{T_{(i;d]}}_{\mathbf{X}}$-measurable, this implies that

$$\mathsf{E}_\mu\big(f_i \,\big|\, \Sigma^{\mathsf{F}_{i,d-1}}_{\mathbf{X}} \vee \Sigma^{T_{(i;d]}}_{\mathbf{X}}\big) = \mathsf{E}_\mu(f_i \mid \Sigma^{\mathsf{F}_{i,d}}_{\mathbf{X}}) \neq 0.$$

By an argument as at the start of Step 1, this implies that for the averages $\widehat{\Lambda}_n^{(d)}$, the subspace $\mathsf{V}_{i,d,\mathbf{X}}$ is partially characteristic in position i for each $i \leq d-1$. Therefore the induction continues. □

Proof of Theorem A. The proof is by induction on d, and uses only the partially characteristic factor in position d. When $d = 1$, convergence is given by the Norm Ergodic Theorem for amenable groups. So suppose $d \geq 2$ and that convergence is known for all commuting tuples of fewer than d actions.

By Theorem 3.5, we may ascend from $\mathbf{X}$ to an extension which is $\mathsf{V}_{d,d}$-sated; and so we simply assume that $\mathbf{X}$ is itself $\mathsf{V}_{d,d}$-sated. This implies that

$$\big\|\Lambda_n(f_1, \dots, f_d) - \Lambda_n(f_1, \dots, f_{d-1}, P_{\mathbf{X}}^{\mathsf{V}_{d,d}} f_d)\big\|_2 \to 0 \quad \text{as } n \to \infty,$$

as in the proof of the Proposition 4.7. It therefore suffices to prove convergence for the right-hand averages inside these norms. However, again as in the proof of Proposition 4.7, for these, we may approximate $\mathsf{E}_\mu(f_d \mid \Sigma_{\mathbf{X}}^{\mathsf{F}_{d,d}})$ by a finite sum of products of the form in (4.8) and then re-arrange the resulting averages into the form

$$h_0 \cdot \frac{1}{|F_n|} \sum_{g \in F_n} \prod_{i=1}^{d-1} ((f_i h_i) \circ T_{[1;i]}^g).$$

Without the factor h_0, which is uniformly bounded and does not depend on n, this is now a system of non-conventional averages for a commuting $(d-1)$-tuple of G-actions, so convergence follows by the inductive hypothesis. □

5 Proof of multiple recurrence

Given a G^d-space, the convergence proved in Theorem A implies that the sequence of measures $\lambda_n := \frac{1}{|F_n|} \sum_{g \in F_n} \delta_{(T_1^g x, T_{[1;2]}^g x, \dots, T_{[1;d]}^g x)}$ converges in the usual topology on the convex set of d-fold couplings of μ. (This is the same as the topology on joinings when one gives all spaces the action of the trivial group; the joining topology is explained, e.g.,, in [Gla03, Section 6.1].)

Let $\lambda := \lim_{n\to\infty} \lambda_n$. It follows that

$$\frac{1}{|F_n|} \sum_{g \in F_n} \mu(T_1^g A \cap \cdots \cap T_{[1;d]}^g A) \to \lambda(A^d)$$

for every measurable $A \subseteq X$. To complete the proof of Theorem B, we show that

$$\lambda(A_1 \times \cdots \times A_d) = 0 \quad \text{implies} \quad \mu(A_1 \cap \cdots \cap A_d) = 0 \tag{5.1}$$

for any G^d-space $\mathbf{X} = (X, \mu, T)$ and measurable subsets $A_1, \dots, A_d \subseteq X$. This gives the desired conclusion by setting $A_1 := \dots := A_d := A$.

First, by replacing $\mathbf{X}$ with a suitable extension and lifting each A_i to that extension, we may reduce this task to the case in which $\mathbf{X}$ is sated with respect to any chosen family of functional σ-subalgebras. After doing this, we prove the result by making contact with a modification of Tao's Infinitary Removal Lemma from [Tao06]. This is the same strategy as in [Aus10a]. The modification of the Removal Lemma is essentially as in [Aus10a], but we use its more explicit formulation from [Aus10b]:

Proposition 5.1 ([Aus10b, Proposition 4.3.1])**.** *Let (X, μ) be a standard Borel probability space with σ-algebra Σ. Let $\pi_i : X^d \to X$ be the coordinate projection for each $i \leq d$. Let θ be a d-fold coupling of μ on X^d. Finally, suppose that $(\Psi_e)_e$ is a collection of σ-subalgebras of Σ, indexed by $e \in \binom{[d]}{\geq 2}$, such that*

(i) *if $a \subseteq e$, then $\Psi_a \supseteq \Psi_e$;*

(ii) *if $i, j \in e$ and $A \in \Psi_e$, then $\theta(\pi_i^{-1}(A) \triangle \pi_j^{-1}(A)) = 0$, so that we may let $\widehat{\Psi}_e$ denote the common θ-completion of the lifted σ-algebras $\pi_i^{-1}(\Psi_e)$ for $i \in e$;*

(iii) *the σ-algebras $\widehat{\Psi}_{\mathcal{I}}$ and $\widehat{\Psi}_{\mathcal{J}}$ are relatively independent under θ over $\widehat{\Psi}_{\mathcal{I} \cap \mathcal{J}}$, where $\widehat{\Psi}_{\mathcal{I}} := \bigvee_{e \in \mathcal{I}} \widehat{\Psi}_e$ for each up-set $\mathcal{I} \subseteq \binom{[d]}{\geq 2}$.*

Suppose also that $\mathcal{I}_{i,j}$ for $i = 1, 2, \ldots, d$ and $j = 1, 2, \ldots, k_i$ are up-sets in $\binom{[d]}{\geq 2}$ such that $[d] \in \mathcal{I}_{i,j} \subseteq \langle i \rangle$ for each i, j, and that $A_{i,j} \in \bigvee_{e \in \mathcal{I}_{i,j}} \Psi_e$ for each i, j. Then

$$\theta\Big(\prod_{i=1}^{d} \Big(\bigcap_{j=1}^{k_i} A_{i,j}\Big)\Big) = 0 \quad \textit{implies} \quad \mu\Big(\bigcap_{i=1}^{d} \bigcap_{j=1}^{k_i} A_{i,j}\Big) = 0.$$

This result is proved by a rather lengthy induction on the up-sets $\mathcal{I}_{i,j}$, which requires the full generality above: see [Aus10b, Subsection 4.3] for a proof and additional discussion. However, as in [Aus10a], we apply it only for $k_i = 1$ and $\mathcal{I}_{i,1} = \langle i \rangle$ for each i, in which case it asserts that, if $A_i \in \Psi_{\langle i \rangle}$ for each i, then

$$\theta(A_1 \times \cdots \times A_d) = 0 \quad \text{implies} \quad \mu(A_1 \cap \cdots \cap A_d) = 0. \tag{5.2}$$

We apply Proposition 5.1 with θ equal to the limit coupling λ, and with the following family of σ-subalgebras. Suppose that $e = \{i_1 < \ldots < i_\ell\} \subseteq [d]$ is non-empty, and define a new functorial σ-subalgebra of G^d-spaces $\mathbf{X}$ by

$$\Phi^e_{\mathbf{X}} := \Sigma^{L_e}_{\mathbf{X}} = \big\{A \in \Sigma_{\mathbf{X}} : \mu(T^g_{(i_1;i_2]} A \triangle A) = \mu(T^g_{(i_2;i_3]} A \triangle A) = \cdots = \mu(T^g_{(i_{\ell-1};i_\ell]} A \triangle A) = 0 \ \forall g \in G\big\},$$

where we interpret this as $\Sigma_{\mathbf{X}}$ in case $|e| = 1$. Observe that if $a \subseteq e$, then $L_a \leq L_e$, and so $\Phi^a_{\mathbf{X}} \supseteq \Phi^e_{\mathbf{X}}$. For any up-set $\mathcal{I} \subseteq \binom{[d]}{\geq 2}$, let $\Phi^{\mathcal{I}}_{\mathbf{X}} := \bigvee_{e \in \mathcal{I}} \Phi^e_{\mathbf{X}}$.

Most of our remaining work goes into checking properties (i)-(iii) above for the joint distribution of the lifted σ-algebras $\pi_i^{-1}(\Phi^e_{\mathbf{X}})$, subject to a certain satedness assumption on $\mathbf{X}$.

Definition 5.2. The G^d-space $\mathbf{X} = (X, \mu, T)$ is **fully sated** if it is sated for every functorial σ-subalgebra of the form $\bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s}$ for some $e_1, \dots, e_r \in \binom{[d]}{\geq 2}$.

Theorem 5.3. *Let* $\mathbf{X}$ *be fully sated, and let* λ *be its limit coupling as above.*

(1) *The coordinates factors* $\pi_i : X^d \to X$ *are relatively independent under* λ *over the further* σ*-subalgebras* $\pi_i^{-1}(\Phi_{\mathbf{X}}^{\langle i \rangle})$, $i = 1, 2, \dots, d$.

(2) *The collection* $(\Phi_{\mathbf{X}}^e)_e$ *satisfies properties (i)–(iii) of Proposition 5.1.*

Proof of Theorem B from Theorem 5.3. Passing to an extension as given by Corollary 3.6, we see that it suffices to prove Theorem B for fully sated G^d-spaces. Specifically, we prove the implication (5.1).

Let $A_1, \dots, A_d \subseteq X$ be measurable. Theorem 5.3(1) gives

$$\lambda(A_1 \times \cdots \times A_d) = \int_X f_1 \otimes \cdots \otimes f_d \, \mathrm{d}\lambda,$$

where $f_i := \mathsf{E}_\mu(1_{A_i} \mid \Phi_{\mathbf{X}}^{\langle i \rangle})$. Let $B_i := \{f_i > 0\}$ for each i. It follows that

$$\lambda(A_1 \times \cdots \times A_d) = 0 \quad \text{implies} \quad \lambda\{f_1 \otimes \cdots \otimes f_d > 0\} = \lambda(B_1 \times \cdots \times B_d) = 0.$$

On the other hand, each B_i is $\Phi_{\mathbf{X}}^{\langle i \rangle}$-measurable. By Theorem 5.3(2), we may therefore apply Proposition 5.1 in the form of the implication (5.2) to conclude that $\lambda(B_1 \times \cdots \times B_d) = 0$ implies $\mu(B_1 \cap \cdots \cap B_d) = 0$. Since $\mu(A_i \setminus B_i) = \int 1_{A_i} \cdot 1_{X \setminus B_i} \mathrm{d}\mu = \int f_i \cdot 1_{X \setminus B_i} \mathrm{d}\mu = 0$ for each i, this completes the proof. $\square$

The rest of this section is devoted to proving Theorem 5.3. Subsection 5.1 establishes various necessary joint-distribution properties of the σ-algebras $\Phi_{\mathbf{X}}^e$ in (X, μ) itself, and then Subsection 5.2 deduces the required properties of the lifts $\pi_i^{-1}(\Phi_{\mathbf{X}}^e)$ from these.

5.1 Joint distribution of some σ-algebras of invariant sets. The next proposition is the second major application of satedness in this paper. It shows that if a G^d-space $\mathbf{X}$ is sated relative to a suitable family of functorial σ-subalgebras constructed out of the collection $\Phi_{\mathbf{X}}^e$, $e \subseteq [k]$, then this forces some relative independence among those σ-subalgebras.

Proposition 5.4. *Suppose that* $\{i < j\} \subseteq [d]$, *suppose that* $e_1, \dots, e_r \in \binom{[d]}{\geq 2}$, *and let* $\mathbf{X} = (X, \mu, T_1, \dots, T_d)$ *be a* G^d*-space.*

(1) *If* $e_s \cap [i; j) = \{i\}$ *for every* $s \leq r$ *and* $\mathbf{X}$ *is* F*-sated for* $\Sigma_{\mathbf{X}}^{\mathsf{F}} := \bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s \cup \{j\}}$, *then* $\Phi_{\mathbf{X}}^{\{i,j\}}$ *and* $\bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s}$ *are relatively independent over* $\Sigma_{\mathbf{X}}^{\mathsf{F}}$.

(2) *If* $e_s \cap (i; j] = \{j\}$ *for every* $s \leq r$ *and* $\mathbf{X}$ *is* G*-sated for* $\Sigma_{\mathbf{X}}^{\mathsf{G}} := \bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s \cup \{i\}}$, *then* $\Phi_{\mathbf{X}}^{\{i,j\}}$ *and* $\bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s}$ *are relatively independent over* $\Sigma_{\mathbf{X}}^{\mathsf{G}}$.

Proof of Part (1). As always, the appeal to satedness depends on constructing the right extension. Let $e := e_1 \cup \ldots \cup e_r \cup \{j\}$, so that our assumptions give $e \cap [i; j] = \{i, j\}$.

Step 1. We first construct a suitable extension of the L_e-subaction $\mathbf{X}^{\upharpoonright L_e}$.

Since $e \cap [i; j] = \{i, j\}$, Corollary 4.2 assures that $\Phi_{\mathbf{X}}^{\{i,j\}}$ is globally L_e-invariant, so the relative product measure $\nu := \mu \otimes_{\Phi_{\mathbf{X}}^{\{i,j\}}} \mu$ is invariant under the diagonal action of L_e on $Y := X^2$. We make use of this by constructing a *non-diagonal* action of L_e on Y which still preserves ν.

Let e be enumerated as $\{i_1 < \ldots < i_m\}$. Our assumptions imply that $\{i, j\} = \{i_{\ell_0}, i_{\ell_0+1}\}$ for some $\ell_0 \leq m-1$. In these terms, L_e is generated by its $m-1$ commuting subgroups $L_{\{i_\ell, i_{\ell+1}\}} = \varphi_\ell(G)$, $\ell = 1, \ldots, m-1$, where $\varphi_\ell : G \longrightarrow G^d$ is the injective homomorphism defined by

$$\big(\varphi_\ell(g)\big)_i := \begin{cases} g & \text{if } i \in (i_\ell; i_{\ell+1}], \\ e & \text{otherwise.} \end{cases}$$

Specifying an action of L_e is equivalent to specifying commuting actions of its subgroups $\varphi_\ell(G)$. We define our new, non-diagonal action $S : L_e \curvearrowright (Y, \nu)$ by

$$S^{\varphi_\ell(g)} := \begin{cases} T^g_{(i_\ell; i_{\ell+1}]} \times T^g_{(i_\ell, i_{\ell+1}]} \quad \text{(i.e., diagonal)} & \text{if } \ell \notin \{\ell_0, \ell_0+1\}, \\ T^g_{(i_{\ell_0}; i_{\ell_0+1}]} \times \mathrm{id} & \text{if } \ell = \ell_0 \\ T^g_{(i_{\ell_0+1}; i_{\ell_0+2}]} \times T^g_{(i_{\ell_0}, i_{\ell_0+2}]} & \text{if } \ell = \ell_0 + 1. \end{cases}$$

(where the last option here is vacuous in case $\ell_0 = m-1$).

Each of these transformations leaves ν invariant. We have already remarked this for the diagonal transformations; and for the last two possibilities, we need only observe that ν is a relative product over a σ-algebra on which $T^g_{(i_{\ell_0}; i_{\ell_0+1}]} = T^g_{(i;j]}$ acts trivially for all $g \in G$.

Now let $\beta_1, \beta_2 : Y \longrightarrow X$ be the two coordinate projections. The above definition gives $\beta_1 \circ S^{\varphi_\ell(g)} = T^{\varphi_\ell(g)} \circ \beta_1$ for every ℓ and g. So, letting $\mathbf{Y}$ denote the L_e-space given by the above transformations S on (Y, ν), we have a factor map $\mathbf{Y} \xrightarrow{\beta_1} \mathbf{X}^{\upharpoonright L_e}$. On the other hand, recall that $\{i_{\ell_0}, i_{\ell_0+1}\} = \{i, j\}$. Thus the above definitions give that the subgroup $L_{\{i,j\}} \leq L_e$ acts trivially under S on the second coordinate in Y.

Step 2. Next, applying Theorem 2.1 enlarges this to an extension $\mathbf{X}_1 \xrightarrow{\pi} \mathbf{X}$ of G^d-spaces which factorizes through some L_e-extension $\mathbf{X}_1^{\upharpoonright L_e} \xrightarrow{\alpha} \mathbf{Y}$.

Step 3. Now suppose that $f \in L^\infty(\mu|_{\Phi_{\mathbf{X}}^{\{i,j\}}})$ and that $g_s \in L^\infty(\mu|_{\Phi_{\mathbf{X}}^{e_s}})$ for each $s \le r$. From the definition of ν and the fact that f is $\Phi_{\mathbf{X}}^{\{i,j\}}$-measurable, we have

$$\int_X f \cdot \prod_{s\le r} g_s \mathrm{d}\mu = \int_Y (f\circ\beta_2)\cdot\Big(\prod_{s\le r} g_s\circ\beta_2\Big)\mathrm{d}\nu = \int_Y (f\circ\beta_1)\cdot\Big(\prod_{s\le r} g_s\circ\beta_2\Big)\mathrm{d}\nu. \tag{5.3}$$

Our next step is to show that for $s \le r$, the function $g_s\circ\beta_2$ is invariant under the whole of $S^{\restriction L_{e_s\cup\{j\}}}$. Enumerating $e_s \cup \{j\} =: \{p_1 < \ldots < p_n\}$ shows that it suffices to prove invariance under each subgroup $L_{\{p_k,p_{k+1}\}}$ for $k \in \{1,2,\ldots,n-1\}$.

There are three cases to consider: $p_k \notin \{i,j\}$, $p_k = i$ and $p_k = j$. If $p_k \notin \{i,j\}$, then, since $e \cap [i;j) = \{i\}$, it follows that $(p_k;p_{k+1}]$ is disjoint from $(i_{\ell_0};i_{\ell_0+2}]$ (again we are using the notation from Step 1). In this case, the definition of S gives $\beta_2\circ S_{(p_k;p_{k+1}]} = T_{(p_k;p_{k+1}]}\circ\beta_2$, so the required invariance follows from the fact that g_s itself is L_{e_s}-invariant.

If $p_k = i$, then the assumption $e_s \cap [i;j) = \{i\}$ implies that $p_{k+1} = j$. Hence the definition of S implies that $\beta_2\circ S_{(p_k;p_{k+1}]} = \beta_2$, from which the $S_{(p_k;p_{k+1}]}$-invariance of $g_s\circ\beta_2$ is obvious.

Finally, if $p_k = j$, the definition of S gives $\beta_2\circ S_{(p_k;p_{k+1}]} = T_{(i;p_{k+1}]}\circ\beta_2$. Since $\{i,p_{k+1}\} \subseteq e_s$ (even if $j \notin e_s$), once again the L_{e_s}-invariance of g_s gives

$$g_s\circ\beta_2\circ S^g_{(p_k;p_{k+1}]} = g_s\circ T^g_{(i;p_{k+1}]}\circ\beta_2 = g_s\circ\beta_2,$$

as required.

Step 4. In light of Step 3, the function $\prod_{s\le r}(g_s\circ\beta_2\circ\alpha)$ is measurable with respect to $\bigvee_{s\le r}\Phi_{\mathbf{X}_1}^{e_s\cup\{j\}} = \Sigma_{\mathbf{X}_1}^{\mathsf{F}}$. By the assumed F-satedness, it follows that the right-hand integral in (5.3) equals $\int_Y(\mathsf{E}_\mu(f\mid\Sigma_{\mathbf{X}}^{\mathsf{F}})\circ\beta_1)\cdot\Big(\prod_{s\le r} g_s\circ\beta_2\Big)\mathrm{d}\nu$; and by the same reasoning that gave (5.3) itself, this equals

$$\int_X \mathsf{E}_\mu(f\mid\Sigma_{\mathbf{X}}^{\mathsf{F}})\cdot\prod_{s\le r} g_s d\mu = \int_X \mathsf{E}_\mu(f\mid\Sigma_{\mathbf{X}}^{\mathsf{F}})\cdot\mathsf{E}_\mu\Big(\prod_{s\le r} g_s\,\Big|\,\Sigma_{\mathbf{X}}^{\mathsf{F}}\Big)\mathrm{d}\mu.$$

Since f and each g_s were arbitrary subject to their measurability assumptions, this implies that $\Phi_{\mathbf{X}}^{\{i,j\}}$ and $\bigvee_{s=1}^r\Phi_{\mathbf{X}}^{e_s}$ are relatively independent over $\Sigma_{\mathbf{X}}^{\mathsf{F}}$. □

Proof of Part (2). This follows exactly the same steps as Part 1, except that now the new L_e-action S on $(Y,\nu) := (X^2, \mu\otimes_{\Phi_{\mathbf{X}}^{\{i,j\}}}\mu)$ is defined as

$$S^{\varphi_\ell(g)} := \begin{cases} T^g_{(i_\ell;i_{\ell+1}]}\times T^g_{(i_\ell,i_{\ell+1}]} & \text{(i.e., diagonal)} \quad \text{if } \ell\notin\{\ell_0-1,\ell_0\},\\ T^g_{(i_{\ell_0};i_{\ell_0+1}]}\times \mathrm{id} & \text{if } \ell=\ell_0,\\ T^g_{(i_{\ell_0-1};i_{\ell_0}]}\times T^g_{(i_{\ell_0-1},i_{\ell_0+1}]} & \text{if } \ell=\ell_0-1,\end{cases}$$

where now $e := e_1\cup\cdots\cup e_r\cup\{i\} = \{i_1 < \ldots < i_m\}$, and ℓ_0 is such that $\{i,j\} = \{i_{\ell_0}, i_{\ell_0+1}\}$, as before. □

The next two propositions contain the consequences of full satedness that we need. The first modifies the conclusion of Proposition 4.7 in the case of integrated averages, rather than functional averages.

Proposition 5.5. *Suppose that* $\mathbf{X}$ *is fully sated, that* $\{i_1 < \ldots < i_k\} \subseteq [d]$ *and that* $f_1, \ldots, f_k \in L^\infty(\mu)$. *Then*

$$\lim_{n\to\infty}\int_X \frac{1}{|F_n|}\sum_{g\in F_n}\prod_{j=1}^k (f_j\circ T^g_{[1;i_j]})\mathrm{d}\mu = \lim_{n\to\infty}\int_X \frac{1}{|F_n|}\sum_{g\in F_n}\prod_{j=1}^k (\mathsf{E}_\mu(f_j \mid \Delta_j)\circ T^g_{[1;i_j]})\mathrm{d}\mu,$$

where

$$\Delta_j := \bigvee_{\ell=1}^{j-1}\Sigma_{\mathbf{X}}^{T_{(i_\ell;i_j]}} \vee \bigvee_{\ell=j+1}^{k}\Sigma_{\mathbf{X}}^{T_{(i_j;i_\ell]}}.$$

Proof. This is proved by induction on k. When $k = 1$, the result is trivial, by the $T_{[1;i_1]}$-invariance of μ. So suppose $k \geq 2$.

Because μ is $T_{[1;i_1]}$-invariant, the desired conclusion is equivalent to

$$\int_X f_1\cdot\Big(\lim_{n\to\infty}\frac{1}{|F_n|}\sum_{g\in F_n}\prod_{j=2}^k (f_j\circ T^g_{(i_1;i_j]})\Big)\,\mathrm{d}\mu$$
$$= \int_X \mathsf{E}_\mu(f_1\mid\Delta_1)\cdot\Big(\lim_{n\to\infty}\frac{1}{|F_n|}\sum_{g\in F_n}\prod_{j=2}^k(\mathsf{E}_\mu(f_j\mid\Delta_j)\circ T^g_{(i_1;i_j]})\Big)\mathrm{d}\mu.$$

However, $\mathbf{X}$ being fully sated implies that the G^{k-1}-space $\mathbf{X}'$ defined by $T'_j := T_{(i_j,i_{j+1}]}$, $j = 1, 2, \ldots, k-1$, is also fully sated: otherwise, we could turn a G^{k-1}-extension witnessing the failure of satedness for $\mathbf{X}'$ back into a G^d-extension of $\mathbf{X}$ using Theorem 2.1. Therefore Proposition 4.7 applied to this G^d-space gives

$$\int_X f_1\cdot\Big(\lim_{n\to\infty}\frac{1}{|F_n|}\sum_{g\in F_n}\prod_{j=2}^k (f_j\circ T^g_{(i_1;i_j]})\Big)\mathrm{d}\mu$$
$$= \int_X f_1\cdot\Big(\lim_{n\to\infty}\frac{1}{|F_n|}\sum_{g\in F_n}\prod_{j=2}^k(\mathsf{E}_\mu(f_j\mid\Delta_j)\circ T^g_{(i_1;i_j]})\Big)\mathrm{d}\mu,$$

since the σ-algebras Δ_j for $j \geq 2$ are those that arise by applying the functorial σ-subalgebras $\mathsf{F}_{\bullet,\bullet}$ of that proposition to the G^{k-1}-space $\mathbf{X}'$.

This almost completes the proof. To finish, observe that, as in the proof of Theorem A, we may now approximate f_k (say) by a finite sum of finite products of functions measurable with respect to $\Sigma_{\mathbf{X}}^{T_{(i_\ell,i_k]}}$ for $\ell = 1, \ldots, k-1$; and having done so, we may re-arrange the above into an analogous system of averages with only $k-1$ transformations. At that point, the inductive hypothesis allows us to replace f_1 with $\mathsf{E}_\mu(f_1 \mid \Delta_1)$, completing the proof. $\square$

Proposition 5.6. *Suppose that* $\mathbf{X}$ *is fully sated. Let* $e_0, e_1, \ldots, e_r \in \binom{[d]}{\geq 2}$ *be sets such that* $\bigcap_{0\leq s\leq r} e_s \neq \varnothing$. *Let* g *be* $\Phi_{\mathbf{X}}^{e_0}$*-measurable and* f_s *be* $\Phi_{\mathbf{X}}^{e_s}$*-measurable for* $1 \leq s \leq r$. *Suppose* g *and every* f_s *are bounded. Then*

$$\int_X g\cdot f_1\cdot\cdots\cdot f_r\,\mathrm{d}\mu = \int_X \mathsf{E}_\mu(g \mid \Psi)\cdot f_1\cdot\cdots\cdot f_r\,\mathrm{d}\mu,$$

where $\Psi := \bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_0\cup e_s}$.

As is standard, this is equivalent to the assertion that

$$\mathsf{E}_\mu\Big(g\,\Big|\,\bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s}\Big) = \mathsf{E}_\mu(g \mid \Psi).$$

Proof. If $e_0 = e_1 = e_2 = \cdots = e_r$, then $\Phi_{\mathbf{X}}^{e_s} = \Psi$ for all s, so the result is trivial.

The general case is proved by an outer induction on r, and an inner induction on the cardinality of the up-set $\mathcal{A}_{e_0,e_1,\ldots,e_r} := \langle e_0\rangle \cup\cdots\cup\langle e_r\rangle$.

The base case corresponds to $r=1$ and $|\mathcal{A}_{e_0,e_1}| = 1$; hence $e_0 = e_1 = [k]$. This is among the trivial cases described above.

For the recursion, we may assume that there exist at least two distinct sets among the e_s for $0 \leq s \leq r$, so $\bigcap_{0\leq s\leq r} e_s \neq \bigcup_{0\leq s\leq r} e_s$ (else we would be in the trivial case treated above). We have also assumed that $\bigcap_{0\leq s\leq r} e_s \neq \varnothing$, so there must be $\{i<j\} \subseteq [k]$ such that $e_s \cap [i+1;j) = \varnothing$ for all $0 \leq s \leq r$, and

- either $i \in e_s$ for all $0 \leq s \leq r$, but j lies in some but not all e_s;
- or $j \in e_s$ for all $0 \leq s \leq r$, but i lies in some but not all e_s.

We complete the induction in the first of these cases, the second case being exactly analogous. In this first case, $e_s \cap [i;j) = \{i\}$ for all s. It now breaks into two further sub-cases: $j \in e_0$ and $j \notin e_0$.

Case 1. First assume that $j \in e_0$. Then, since $\Phi_{\mathbf{X}}^{e_0} \subseteq \Phi_{\mathbf{X}}^{\{i,j\}}$, Proposition 5.4(1) gives

$$\int_X g\cdot(f_1\cdot\cdots\cdot f_r)\mathrm{d}\mu = \int_X \mathsf{E}_\mu\Big(g\,\Big|\,\bigvee_{s=1}^r \Phi_{\mathbf{X}}^{e_s\cup\{j\}}\Big)\cdot(f_1\cdot\cdots\cdot f_r)\mathrm{d}\mu.$$

Now observe that $\mathcal{A}_{e_0,e_1\cup\{j\},\ldots,e_r\cup\{j\}} \subsetneq \mathcal{A}_{e_0,\ldots,e_r}$. The inclusion is strict because the right-hand family contains some set that does not contain j, whereas every element of the left-hand family contains j. Therefore we may apply the inductive hypothesis to e_0 and $e_1 \cup \{j\}, e_2 \cup \{j\}, \dots, e_r \cup \{j\}$, to conclude that

$$\mathsf{E}_\mu\Big(g\,\Big|\,\bigvee_{s=1}^{r}\Phi_{\mathbf{X}}^{e_s\cup\{j\}}\Big) = \mathsf{E}_\mu\Big(g\,\Big|\,\bigvee_{s=1}^{r}\Phi_{\mathbf{X}}^{e_0\cup e_s}\Big) = \mathsf{E}_\mu(g\,|\,\Psi).$$

Case 2. Now assume that $j \notin e_0$. Re-labeling the other sets if necessary, we may assume $j \in e_1$. Then the argument used in Case 1 gives

$$\int_X g\cdot f_1\cdot\dots\cdot f_r\,\mathrm{d}\mu = \int_X g\cdot\mathsf{E}_\mu(f_1\,|\,\Psi_1)\cdot f_2\cdot\dots\cdot f_r\,\mathrm{d}\mu,$$

where $\Psi_1 := \bigvee_{s\in\{0\}\cup[r]\setminus\{1\}}\Phi_{\mathbf{X}}^{e_1\cup e_s}$, and similarly,

$$\int_X \mathsf{E}_\mu(g\,|\,\Psi)\cdot f_1\cdot\dots\cdot f_r\,\mathrm{d}\mu = \int_X \mathsf{E}_\mu(g\,|\,\Psi)\cdot\mathsf{E}_\mu(f_1\,|\,\Psi_1)\cdot f_2\cdot\dots\cdot f_r\,\mathrm{d}\mu.$$

It therefore suffices to prove the desired equality when f_1 is Ψ_1-measurable. Since such an f_1 can be approximated in $\|\cdot\|_2$ by finite sums of products of $\Phi_{\mathbf{X}}^{e_1\cup e_s}$-measurable functions for $s \in \{0\}\cup[r]\setminus\{1\}$, it suffices furthermore to assume that $f_1 = f_{10}\cdot f_{12}\cdot\dots\cdot f_{1r}$ is one such product.

However, now the integral of interest may be written as

$$\int_X (gf_{10})\cdot(f_2f_{12})\cdot\dots\cdot(f_rf_{1r})\mathrm{d}\mu;$$

and, by the inductive hypothesis on r, this equals

$$\int_X \mathsf{E}_\mu(gf_{10}\,|\,\Psi_2)\cdot(f_2f_{12})\cdot\dots\cdot(f_rf_{1r})\mathrm{d}\mu$$

with $\Psi_2 := \bigvee_{s=2}^{r}\Phi_{\mathbf{X}}^{e_0\cup e_s}$.

Since $\Psi \geq \Psi_2$ and f_{10} is Ψ-measurable, the Law of Iterated Conditional Expectation gives

$$\mathsf{E}_\mu(gf_{10}\,|\,\Psi_2) = \mathsf{E}_\mu(\mathsf{E}_\mu(g\,|\,\Psi)f_{10}\,|\,\Psi_2).$$

Substituting this back into the integral completes the desired equality. □

Corollary 5.7. *Suppose that* $\mathbf{X}$ *is fully sated and that* $e_1, \dots, e_r \in \binom{[d]}{\geq 2}$ *are such that* $\bigcap_{s\leq r} e_s \neq \emptyset$. *Let* $\Psi_s := \bigvee_{t\in[r]\setminus s}\Phi_{\mathbf{X}}^{e_t\cup e_s}$ *for* $s = 1, 2, \dots, r$, *and let* f_s *be* $\Phi_{\mathbf{X}}^{e_s}$*-measurable for each* s. *Then*

$$\int_X \prod_{s=1}^{r} f_s\,\mathrm{d}\mu = \int_X \prod_{s=1}^{r}\mathsf{E}_\mu(f_s\,|\,\Psi_s)\mathrm{d}\mu.$$

Proof. Simply apply Proposition 5.6 to each factor of the integrand. □

5.2 Structure of the limit couplings. We now turn to the structure of the limit coupling λ, as discussed at the beginning of this section.

Lemma 5.8 ([Aus10a]). *If $i, j \in e \in \binom{[d]}{\geq 2}$ and $A \in \Phi_{\mathbf{X}}^{e}$, then*

$$\lambda(\pi_i^{-1}(A) \triangle \pi_j^{-1}(A)) = 0.$$

In particular, $\pi_i^{-1}(\Phi_{\mathbf{X}}^{e})$ and $\pi_j^{-1}(\Phi_{\mathbf{X}}^{e})$ agree up to λ-negligible sets.

Proof. If $i = j$, the result is trivial; so assume without loss of generality that $i < j$. By definition,

$$\begin{aligned}\lambda(\pi_i^{-1}(A) \cap \pi_j^{-1}(A)) &= \lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu(T_{[1;i]}^{g^{-1}} A \cap T_{[1;j]}^{g^{-1}} A) \\ &= \lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu(T_{[1;i]}^{g^{-1}} A \cap T_{[1;i]}^{g^{-1}} T_{(i;j]}^{g^{-1}} A) \\ &= \lim_{n\to\infty} \frac{1}{|F_n|} \sum_{g\in F_n} \mu(T_{[1;i]}^{g^{-1}} A \cap T_{[1;i]}^{g^{-1}} A) = \mu(A),\end{aligned}$$

since $\Phi_{\mathbf{X}}^{e} \leq \Phi_{\mathbf{X}}^{\{i,j\}}$; so A is $T_{(i;j]}$-invariant. Therefore,

$$\lambda(\pi_i^{-1}(A) \cap \pi_j^{-1}(A)) = \lambda(\pi_i^{-1}(A)) = \lambda(\pi_j^{-1}(A)),$$

and so $\lambda(\pi_i^{-1}(A) \triangle \pi_j^{-1}(A)) = 0$. □

Definition 5.9. In the setting above, the common λ-completion of the lifted σ-algebras $\pi_i^{-1}(\Phi_{\mathbf{X}}^{e})$ for $i \in e$ is called the **oblique copy** of $\Phi_{\mathbf{X}}^{e}$ and is denoted by $\widehat{\Phi}_{\mathbf{X}}^{e}$. If $\mathcal{I}$ is a non-empty up-set in $\binom{[d]}{\geq 2}$ and $\mathbf{X}$ is a G^d-space, then $\widehat{\Phi}_{\mathbf{X}}^{\mathcal{I}} := \bigvee_{e\in\mathcal{I}} \widehat{\Phi}_{\mathbf{X}}^{e}$ is called the **$\mathcal{I}$-oblique σ-algebra**.

The remainder of the proof of Theorem B follows almost exactly the same lines as [Aus10b, Subsection 4.2]; again we include the following lemma for completeness.

Lemma 5.10 (cf. [Aus10b, Proposition 4.2.6]). *If $\mathcal{I}$ and $\mathcal{J}$ are non-empty up-sets in $\binom{[d]}{\geq 2}$ and $\mathbf{X}$ is a fully sated G^d-space, then $\widehat{\Phi}_{\mathbf{X}}^{\mathcal{I}}$ and $\widehat{\Phi}_{\mathbf{X}}^{\mathcal{J}}$ are relatively independent over $\widehat{\Phi}_{\mathbf{X}}^{\mathcal{I}\cap\mathcal{J}}$ under λ.*

Proof. *Step 1.* Suppose first that $\mathcal{J} = \langle e \rangle$ where e is a maximal member of $\binom{[d]}{\geq 2} \setminus \mathcal{I}$. Let $\{a_1, a_2, \dots, a_m\}$ be the antichain of minimal elements of $\mathcal{I}$, so that $\widehat{\Phi}_{\mathbf{X}}^{\mathcal{I}} = \bigvee_{k\leq m} \widehat{\Phi}_{\mathbf{X}}^{a_k}$. The maximality assumption on e implies that $e \cup \{j\}$ contains some a_k for every $j \in [d] \setminus e$, and so $\mathcal{I} \cap \mathcal{J}$ is precisely the up-set generated by these sets $e \cup \{j\}$ for $j \in [d] \setminus e$. We must therefore show that $\widehat{\Phi}^{e}_{\mathbf{X}}$ is relatively independent from $\bigvee_{k \le m} \widehat{\Phi}^{a_k}_{\mathbf{X}}$ under λ over the σ-subalgebra $\bigvee_{j \in [d] \setminus e} \widehat{\Phi}^{e \cup \{j\}}_{\mathbf{X}}$.

Since $e \notin \mathcal{I}$ we can find some $j_k \in a_k \setminus e$ for each $k \le m$. Moreover, each $j \in [d] \setminus e$ must appear as some j_k in this list, since it appears for any k for which $a_k \subseteq e \cup \{j\}$.

Now Lemma 5.8 implies that $\widehat{\Phi}^{a_k}_{\mathbf{X}}$ agrees with $\pi_{j_k}^{-1}(\Phi^{a_k}_{\mathbf{X}})$ up to λ-negligible sets. On the other hand, we clearly have $\pi_{j_k}^{-1}(\Phi^{a_k}_{\mathbf{X}}) \le \pi_{j_k}^{-1}(\Sigma_{\mathbf{X}})$; so it suffices to show that $\widehat{\Phi}^{e}_{\mathbf{X}}$ is relatively independent from $\bigvee_{j \in [d] \setminus e} \pi_j^{-1}(\Sigma_{\mathbf{X}})$ over $\bigvee_{j \in [d] \setminus e} \widehat{\Phi}^{e \cup \{j\}}_{\mathbf{X}}$.

Choose $i \in e$. Then Lemma 5.8 also implies that $\widehat{\Phi}^{e}_{\mathbf{X}}$ agrees with $\pi_i^{-1}(\Phi^{e}_{\mathbf{X}})$ up to λ-negligible sets. On the other hand, for the sets $A_j \in \pi_j^{-1}(\Sigma_{\mathbf{X}})$ for $j \in [d] \setminus e$ and $B \in \Phi^{e}_{\mathbf{X}}$, the definition of λ gives

$$\lambda\Big(\Big(\bigcap_{j \in [d] \setminus e} \pi_j^{-1}(A_j)\Big) \cap \pi_i^{-1}(B)\Big) = \lim_{n \to \infty} \int_X \Lambda_n(f_1, \dots, f_d)\, \mathrm{d}\mu$$

with

$$f_\ell := \begin{cases} 1_{A_\ell} & \text{if } \ell \in [d] \setminus e, \\ 1_B & \text{if } \ell = i, \\ 1 & \text{otherwise.} \end{cases}$$

Proposition 5.5 gives the same limit from $\lim_{n \to \infty} \int_X \Lambda_n(f'_1, \dots, f'_d) \mathrm{d}\mu$, where

$$f'_\ell := \begin{cases} 1_{A_\ell} & \text{if } \ell \in [d] \setminus e, \\ \mathsf{E}_\mu(1_B \mid \Delta) & \text{if } \ell = i, \\ 1 & \text{otherwise,} \end{cases}$$

and

$$\Delta := \bigvee_{\ell \in [d] \setminus e,\, \ell < i} \Sigma^{T_{(\ell; i]}}_{\mathbf{X}} \vee \bigvee_{\ell \in [d] \setminus e,\, \ell > i} \Sigma^{T_{(i; \ell]}}_{\mathbf{X}} = \bigvee_{j \in [d] \setminus e} \Phi^{\{i,j\}}_{\mathbf{X}}.$$

This implies that $\pi_i^{-1}(\Phi^{e}_{\mathbf{X}})$ is relatively independent from $\bigvee_{j \in [d] \setminus e} \pi_j^{-1}(\Sigma_{\mathbf{X}})$ over $\pi_i^{-1}(\Delta)$ under λ. On the other hand, Corollary 5.7 gives that $\Phi^{e}_{\mathbf{X}}$ is relatively independent from Δ over $\bigvee_{j \in [d] \setminus e} \Phi^{e \cup \{j\}}_{\mathbf{X}}$. Combining these conclusions completes the proof in this case.

Step 2. The general case can now be treated for fixed $\mathcal{I}$ by induction on $\mathcal{J}$. If $\mathcal{J} \subseteq \mathcal{I}$, the result is clear; so let e be a minimal member of $\mathcal{J} \setminus \mathcal{I}$ of maximal size, and let $\mathcal{K} := \mathcal{J} \setminus \{e\}$. It suffices to prove that if $F \in L^\infty(\lambda)$ is $\widehat{\Phi}^{\mathcal{J}}_{\mathbf{X}}$-measurable, then $\mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}) = \mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{I} \cap \mathcal{J}}_{\mathbf{X}})$. Furthermore, by an approximation in $\|\cdot\|_2$ by finite sums of products, it suffices to prove this only for F that are of the form $F_1 \cdot F_2$ with F_1 and F_2 being bounded and respectively $\widehat{\Phi}^{\langle e \rangle}_{\mathbf{X}}$- and $\widehat{\Phi}^{\mathcal{K}}_{\mathbf{X}}$-measurable.

However, for such a product,

$$\mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}) = \mathsf{E}_\lambda\big(\mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{I}\cup\mathcal{K}}_{\mathbf{X}}) \,\big|\, \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}\big) = \mathsf{E}_\lambda\big(\mathsf{E}_\lambda(F_1 \mid \widehat{\Phi}^{\mathcal{I}\cup\mathcal{K}}_{\mathbf{X}}) \cdot F_2 \,\big|\, \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}\big).$$

By Step 1,

$$\mathsf{E}_\lambda(F_1 \mid \widehat{\Phi}^{\mathcal{I}\cup\mathcal{K}}_{\mathbf{X}}) = \mathsf{E}_\lambda(F_1 \mid \widehat{\Phi}^{(\mathcal{I}\cup\mathcal{K})\cap\langle e\rangle}_{\mathbf{X}}),$$

while on the other hand, $(\mathcal{I}\cup\mathcal{K})\cap\langle e\rangle \subseteq \mathcal{K}$ (since $\mathcal{K}$ contains every subset of $[d]$ that strictly includes e, since $\mathcal{J}$ is an up-set). Therefore $(\mathcal{I}\cup\mathcal{K})\cap\langle e\rangle = \mathcal{K}\cap\langle e\rangle$. Another appeal to Step 1 gives $\mathsf{E}_\lambda(F_1 \mid \widehat{\Phi}^{(\mathcal{I}\cup\mathcal{K})\cap\langle e\rangle}_{\mathbf{X}}) = \mathsf{E}_\lambda(F_1 \mid \widehat{\Phi}^{\mathcal{K}}_{\mathbf{X}})$, and hence the above expression for $\mathsf{E}_\lambda(F_1F_2 \mid \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}})$ simplifies to

$$\begin{aligned}\mathsf{E}_\lambda\big(\mathsf{E}_\lambda(F_1 \mid \widehat{\Phi}^{\mathcal{K}}_{\mathbf{X}}) \cdot F_2 \,\big|\, \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}\big) &= \mathsf{E}_\lambda\big(\mathsf{E}_\lambda(F_1 \cdot F_2 \mid \widehat{\Phi}^{\mathcal{K}}_{\mathbf{X}}) \,\big|\, \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}\big)\\ &= \mathsf{E}_\lambda\big(\mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{K}}_{\mathbf{X}}) \,\big|\, \widehat{\Phi}^{\mathcal{I}}_{\mathbf{X}}\big) = \mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{I}\cap\mathcal{K}}_{\mathbf{X}}) = \mathsf{E}_\lambda(F \mid \widehat{\Phi}^{\mathcal{I}\cap\mathcal{J}}_{\mathbf{X}}),\end{aligned}$$

where the third equality follows by the inductive hypothesis applied to $\mathcal{K}$ and $\mathcal{I}$. □

Proof of Theorem 5.3. *Part (1).* Since $\mathbf{X}$ is fully sated, in particular it is sated with respect to the functorial σ-subalgebra

$$\bigvee_{\ell=1}^{i-1} \Sigma_{\mathbf{X}}^{T_{(\ell;i]}} \vee \bigvee_{\ell=i+1}^{d} \Sigma_{\mathbf{X}}^{T_{(i;\ell]}} = \bigvee_{\ell=1}^{i-1} \Phi_{\mathbf{X}}^{\{\ell,i\}} \vee \bigvee_{\ell=i+1}^{d} \Phi_{\mathbf{X}}^{\{i,\ell\}} = \Phi_{\mathbf{X}}^{\langle i\rangle}$$

for each $1 \le i \le d$. Therefore, Proposition 5.5 gives

$$\begin{aligned}\int_{X^d} f_1\otimes\cdots\otimes f_d\,\mathrm{d}\lambda &= \lim_{n\to\infty}\frac{1}{|F_n|}\sum_{g\in F_n}\int_X \prod_{i=1}^d (f_i\circ T^g_{[1;i]})\mathrm{d}\mu\\ &= \lim_{n\to\infty}\frac{1}{|F_n|}\sum_{g\in F_n}\int_X \prod_{i=1}^d (\mathsf{E}_\mu(f_i \mid \Phi_{\mathbf{X}}^{\langle i\rangle})\circ T^g_{[1;i]})\mathrm{d}\mu\\ &= \int_{X^d} \mathsf{E}_\mu(f_1 \mid \Phi_{\mathbf{X}}^{\langle 1\rangle})\otimes\cdots\otimes \mathsf{E}_\mu(f_d \mid \Phi_{\mathbf{X}}^{\langle d\rangle})\,\mathrm{d}\lambda\end{aligned}$$

for all $f_1, \dots, f_d \in L^\infty(X,\mu)$.

Part (2). Proposition 5.1(i) holds by construction of the σ-algebras $\Phi^e_{\mathbf{X}}$; property (ii) is given by Lemma 5.8; and property (iii) is given by Lemma 5.10. □

References

[Aus09] T. Austin, *On the norm convergence of nonconventional ergodic averages*, Ergodic Theory Dynam. Systems **30** (2009), 321–338.

[Aus10a] T. Austin, *Deducing the multidimensional Szemerédi theorem from an infinitary removal lemma*, J. Anal. Math. **111** (2010), 131–150.

[Aus10b] T. Austin, *Multiple recurrence and the structure of probability-preserving systems*, Ph.D. thesis, University of California, Los Angeles, 2010.

[Aus15] T. Austin, *Pleasant extensions retaining algebraic structure, I*, J. Anal. Math. **125** (2015), 1–36.

[BH92] V. Bergelson and N. Hindman, *Some topological semicommutative van der Waerden type theorems and their combinatorial consequences*, J. London Math. Soc. (2) **45** (1992), 385–403.

[BM07] V. Bergelson and R. McCutcheon, *Central sets and a non-commutative Roth theorem*, Amer. J. Math. **129** (2007), 1251–1275.

[BMZ97] V. Bergelson, R, McCutcheon, and Q. Zhang, *A Roth theorem for amenable groups*, Amer. J. Math. **119** (1997), 1173–1211.

[FK78] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi Theorem for commuting transformations*, J. Anal. Math. **34** (1978), 275–291.

[Gla03] E. Glasner, *Ergodic Theory via Joinings*, American Mathematical Society, Providence, RI, 2003.

[Hos09] B. Host, *Ergodic seminorms for commuting transformations and applications*, Studia Math. **195** (2009), 31–49.

[HK05] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. **161** (2005), 397–488.

[Tao06] T. Tao, *Szemerédi's regularity lemma revisited*, Contrib. Discrete Math. **1** (2006), 8–28.

[Tao07] T. Tao, *A correspondence principle between (hyper)graph theory and probability theory, and the (hyper)graph removal lemma*, J. Anal. Math. **103** (2007), 1–45.

[Wal12] M. Walsh, *Norm convergence of nilpotent ergodic averages*, Ann. of Math. (2) **175** (2012), 1667–1688.

[ZK] P. Zorin-Kranich, *Norm convergence of multiple ergodic averages on amenable groups* J. Anal. Math. **130** (2016), 219–241.

Tim Austin
COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
NEW YORK NY 10012, USA
email: tim@cims.nyu.edu

(Received October 25, 2013)