

AVERAGING ALONG CUBES

BERNARD HOST AND BRYNA KRA

ABSTRACT. We study the convergence of an average of eight functions, where the average is taken along cubes whose sizes tend to $+\infty$ and show that this reduces to proving an ergodic theorem for translations on a 2-step nilsystem. We derive a combinatorial interpretation for the arithmetic structure inside a set of integers of positive upper density.

1. INTRODUCTION

Nonconventional ergodic averages along arithmetic progressions have been studied by numerous authors. However, averaging along progressions turns out not to be the natural setting. Rather, averaging along *combinatorial cubes* is a more natural context and here we give a specific example in dimension three. We generalize our results for higher dimensions in a forthcoming paper [HK02b].

1.1. Background. A subset A of a discrete abelian group G is said to be *syndetic* if finitely many translates of A cover G . In particular, if $G = \mathbb{Z}^d$, this means that there exists some L so that A intersects every d -dimensional cube of size L . Khintchine proved the following result (originally for \mathbb{R} actions):

Theorem 1 (Khintchine [K34]). *Let (X, \mathcal{B}, μ, T) be an invertible measure-preserving probability system and let $A \in \mathcal{B}$. Then for any $\epsilon > 0$, the set*

$$\{n \in \mathbb{Z} : \mu(A \cap T^n A) \geq \mu(A)^2 - \epsilon\}$$

is syndetic.

Bergelson [B00] showed that an analogous result holds for second iteration. More precisely, Bergelson showed that for any $\epsilon > 0$, the set of $(n, m) \in \mathbb{Z}^2$ so that $\mu(A \cap T^n A \cap T^m(A \cap T^n A)) \geq \mu(A)^4 - \epsilon$ is syndetic in \mathbb{Z}^2 . Although Khintchine's Theorem can be proven without use of an ergodic theorem, such elementary methods do not seem to suffice for proving higher dimensional results, such as Bergelson's Theorem.

1.2. Statements of Theorems. We generalize Khintchine's Theorem to three dimensions. We show:

Theorem 2. *Let (X, \mathcal{B}, μ, T) be an invertible measure-preserving probability system and let $A \in \mathcal{B}$. Then for any $\epsilon > 0$, the set of $(m, n, p) \in \mathbb{Z}^3$ so that*

$$\mu\left(A \cap T^m A \cap T^n A \cap T^{m+n} A \cap T^p A \cap T^{m+p} A \cap T^{n+p} A \cap T^{m+n+p} A\right) \geq \mu(A)^8 - \epsilon$$

is syndetic.

The first author thanks the Shapiro fund of Pennsylvania State University for its support and the second author thanks the Université de Marne la Vallée for its hospitality while this work was completed.

This lower bound is optimal if one only considers bounds dependent on $\mu(A)$. Theorem 2 follows from the following convergence theorems:

Theorem 3. *Let (X, \mathcal{B}, μ, T) be an invertible measure-preserving probability system and let $f_1, f_2, \dots, f_7 \in L^\infty(X)$. Then the average over $m \in [M, M']$, $n \in [N, N']$ and $p \in [P, P']$ of*

$$(1) \quad f_1(T^m x) f_2(T^n x) f_3(T^{m+n} x) f_4(T^p x) f_5(T^{m+p} x) f_6(T^{n+p} x) f_7(T^{m+n+p} x)$$

converges in L^2 as $M' - M, N' - N$ and $P' - P$ tend to $+\infty$.

Theorem 4. *Let (X, \mathcal{B}, μ, T) be an invertible measure-preserving probability system. For any $A \in \mathcal{B}$, the average over $m \in [M, M']$, $n \in [N, N']$ and $p \in [P, P']$ of*

$$\mu\left(A \cap T^m A \cap T^n A \cap T^{m+n} A \cap T^p A \cap T^{m+p} A \cap T^{n+p} A \cap T^{m+n+p} A\right)$$

converges to a limit which is greater than or equal to $\mu(A)^8$, when $M' - M, N' - N$ and $P' - P$ tend to $+\infty$.

We view this as an average taken over the *combinatorial cubes* $(0, m, n, m + n, p, m + p, n + p, m + n + p)$.

1.3. Generalization of Khintchine's Theorem. We now prove Theorem 2, assuming Theorems 3 and 4:

Proof of Theorem 2. Let E be the subset of \mathbb{Z}^3 appearing in Theorem 4. If E is not syndetic, there exist intervals $[M_i, M'_i], [N_i, N'_i], [P_i, P'_i]$ with the lengths of the intervals tending to $+\infty$, such that

$$E \cap ([M_i, M'_i] \times [N_i, N'_i] \times [P_i, P'_i]) = \emptyset.$$

Taking the average along these three-dimensional cubes in Theorem 4, we have a contradiction. \square

1.4. Combinatorial Interpretation. The *upper density* \bar{d} of a set $A \subset \mathbb{N}$ is defined to be

$$\bar{d}(A) = \limsup_{N \rightarrow \infty} \frac{|A \cap \{1, 2, \dots, N\}|}{N}.$$

Using Furstenberg's correspondence principle [F77], we obtain the following combinatorial statement as a corollary of Theorem 2:

Theorem 5. *Let $A \subset \mathbb{Z}$ with $\bar{d}(A) > \delta > 0$. The set of $(m, n, p) \in \mathbb{N}^3$ such that*

$$\bar{d}\left(A \cap (A + m) \cap (A + n) \cap (A + m + n) \cap (A + p) \cap (A + m + p) \cap (A + n + p) \cap (A + m + n + p)\right) \geq \delta^8$$

is syndetic.

This theorem is closely related to other combinatorial statements, in particular Szemerédi's Theorem. Both theorems are concerned with demonstrating the existence of some arithmetic structure inside a set of positive upper density. Moreover, an arithmetic progression of length four can be seen as a cube with $m = n = p$. However, the end result is rather different. In our theorem, we have an explicit lower bound that is optimal, but it is impossible to have any control over the size of the syndeticity constant. This means that this result does not have a finite version.

On the other hand, Szemerédi’s Theorem can be expressed in purely finite terms, but the problem of finding the optimal bound is completely open.

1.5. Outline of the paper. In order to prove a convergence result, we find a factor of the system that controls the limit behavior and then prove convergence for this factor. In the terminology of Furstenberg and Weiss [FW96], this factor is known as a *characteristic factor*.

For Bergelson’s Theorem, the factor used is the classical Kronecker factor. However, for convergence of a three-dimensional combinatorial cube, we need the more complicated and less classical *Conze-Lesigne algebra*.

In Section 2, we start by introducing the background needed on 2-step nilmanifolds. In Sections 3 and 4, we present a construction of the Conze-Lesigne algebra using several measures on certain Cartesian powers of the system. One of the most important properties of the Conze-Lesigne algebra is that it can be represented as an inverse limit of 2-step nilmanifolds and we outline how this result is obtained.

Along the way, we have the tools needed to prove Bergelson’s Theorem and we do so in Section 5. The three-dimensional convergence requires more work, as it relies on an L^2 convergence theorem for seven terms and a weak convergence theorem for eight terms. In Section 6, we prove a convergence theorem for the integral of seven terms. In Section 7, we show that the Conze-Lesigne algebra is a characteristic factor for eight terms and thus reduce the convergence for a three-dimensional combinatorial cube to a problem of convergence in a 2-step nilmanifold. We conclude by proving this convergence in Section 8.

1.6. Notation and conventions. By ergodic decomposition, it suffices to prove the results for ergodic dynamical systems. Throughout, we assume that (X, \mathcal{X}, μ, T) is an ergodic dynamical system.

1.6.1. Ergodic systems. We generally omit the σ -algebra from our notation: the system is written (X, μ, T) and we assume implicitly that all sets and functions that we write are measurable. When needed, the σ -algebra of (X, μ, T) is written \mathcal{X} . Each Cartesian power of X is endowed with the product σ -algebra.

In a slight abuse of vocabulary, we use the word factor with two different meanings. First we call a *factor* any T -invariant sub- σ -algebra \mathcal{Z} of \mathcal{X} . For $f \in L^1(\mu)$, we write $\mathbb{E}(f \mid \mathcal{Z})$ for the conditional expectation relative to the measure μ . On the other hand, if (Y, ν, S) is an ergodic system and $p: X \rightarrow Y$ is a measurable map with $S \circ p = p \circ T$ that maps μ to ν , then we say that (Y, ν, S) is a *factor of (X, μ, T)* with *factor map p* . The σ -algebra $p^{-1}(\mathcal{Y})$ is then a factor and every factor can be built in this way. For $f \in L^1(\mu)$, we write $\mathbb{E}(f \mid Y)$ for the function on Y defined by

$$\mathbb{E}(f \mid p^{-1}(\mathcal{Y})) = \mathbb{E}(f \mid Y) \circ p .$$

We often identify the σ -algebras \mathcal{Y} and $p^{-1}(\mathcal{Y})$, writing $\mathbb{E}(f \mid Y)$ instead of $\mathbb{E}(f \mid p^{-1}(\mathcal{Y}))$.

1.6.2. Cartesian products and binary indexing. We consider Cartesian products, X^2 , X^4 , X^7 , and X^8 of some space X and so we introduce some notation to shorten the expressions.

When T is a transformation of X , we write T_k for $T \times T \times \dots \times T$ (k times). In particular, Id_k is the identity on X^k .

Let $E = \{0, 1\}^3$. When $\epsilon \in E$ appears as a subscript, we write it without parentheses and without commas. For a point in X^8 , we write $\mathbf{x} = (x_\epsilon; \epsilon \in E)$. Given eight functions f_ϵ , $\epsilon \in E$, on X , we write $\bigotimes_{\epsilon \in E} f_\epsilon$ for the function $\mathbf{x} \mapsto \prod_{\epsilon \in E} f_\epsilon(x_\epsilon)$ on X^8 .

We also use the notation $E^* = \{0, 1\}^3 \setminus \{(0, 0, 0)\}$, $D = \{0, 1\}^2$ and $D^* = D \setminus \{(0, 0)\}$ with the same conventions as in E for points of X^7 , X^4 and X^3 and for functions on these spaces.

For $n = (n_1, n_2, n_3) \in \mathbb{Z}^3$ and $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3) \in E$ we write $n \cdot \epsilon = \epsilon_1 n_1 + \epsilon_2 n_2 + \epsilon_3 n_3$. We use analogous notations for $n \cdot \eta$ when $n \in \mathbb{Z}^2$ and $\eta \in D$.

For $\epsilon \in E$, we write $\sigma(\epsilon) = (-1)^{\epsilon_1 + \epsilon_2 + \epsilon_3}$.

With this notation, the average of the set in Theorem 2 becomes

$$\left\{ n \in \mathbb{Z}^3 : \mu \left(\bigcap_{\epsilon \in E} T^{n \cdot \epsilon} A \right) \geq \mu(A)^8 - \epsilon \right\}.$$

To remain in the realm of only relatively disagreeable notations, we restrict ourselves to real-valued functions, except when explicitly noted. All the theorems can be easily extended to apply to complex-valued functions.

2. 2-STEP NILSYSTEMS

None of the results in this section is new, but some of them have been rephrased for use in our set up. We do not include any proofs and refer to Malcev [Ma], Auslander, Green and Hahn [AGH63], Parry ([P69], [P70]), Lesigne ([L89], [L91]) and Leibman [Le02] for references.

The notation introduced here is used throughout the sequel.

2.1. 2-step nilmanifolds. A group G is said to be *2-step nilpotent* if its commutator subgroup $[G, G]$ is included in its center. Throughout this section, G is a 2-step nilpotent Lie group. We let U denote the commutator subgroup of G .

Let Λ be a discrete, cocompact subgroup of G . We call the compact manifold $X = G/\Lambda$ a *2-step nilmanifold*. We use (see Malcev [Ma]) :

- *The subgroups U and $U\Lambda$ are closed in G and are Lie subgroups of G .*

The group G acts on X by left translation and we write this action as $(g, x) \mapsto g \cdot x$ or $x \mapsto T_g x$. For $u \in U$, the left translation T_u by u is called a *vertical translation*.

There exists a unique probability measure on X that is invariant under the action of G , called the *Haar measure* of X . We denote it by μ .

2.2. 2-step nilsystems. We maintain the notation of the previous section. Let α be a fixed element of G and $T = T_\alpha$. The topological system (X, T) is called a *topological 2-step nilsystem* and the system (X, μ, T) a *2-step nilsystem* or a *translation on a nilmanifold*.

In the remainder of this Section, (X, μ, T) is a 2-step nilsystem and G, Λ, α are as above. X can be represented as a nilmanifold in several ways and we reduce to a particular representation.

Assume first that (X, μ, T) is ergodic. Let G_0 be the connected component of the identity in G , G' the subgroup of G spanned by G_0 and α and $\Lambda' = G' \cap \Lambda$. Then G' is open in G and Λ' is a discrete cocompact subgroup of G' . By ergodicity, the projection of G' on X is onto and thus we can identify X with G'/Λ' . Therefore, for an ergodic 2-step nilsystem we can assume:

(H1) G is spanned by the connected component of the identity and α .

The second reduction is possible for arbitrary 2-step nilsystems. Let Λ'' be the largest normal subgroup of G contained in Λ . By substituting G/Λ'' for G and Λ/Λ'' for Λ , we can assume that

(H2) Λ does not contain any nontrivial normal subgroup of G .

We notice that if property (H1) was satisfied in the initial representation of X , then it is also satisfied in the new one. Thus for any ergodic 2-step nilsystem we can assume that properties (H1) and (H2) are satisfied.

Property (H2) means that the group G acts faithfully on X . That is, the unique $g \in G$ with $g \cdot x = x$ for every x is the identity element. This implies that $\Lambda \cap U$ is the trivial group and it follows that U is compact. It can be shown that U is connected and thus is a finite-dimensional torus. Moreover Λ is abelian.

Let us assume that the two properties (H1) and (H2) are satisfied. We fix some more notation. Let λ denote the Haar measure of U , π be the natural projection of $X = G/\Lambda$ on the compact abelian group $K = G/\Lambda U$, m be the Haar measure of K , $q: G \rightarrow K$ be the natural projection, $\beta = q(\alpha)$ and let R denote the rotation by β on K .

The action of U on X commutes with the transformation T . Furthermore, this action is free, meaning that for $x \in X$ and $u \in U$, if $u \cdot x = x$ then u is the identity element 1 of U . The quotient of X under this action is K . Thus the topological dynamical system (X, T) is *distal* in the sense of Furstenberg [F63]. This observation leads to the following result:

Proposition 6 (Parry [P69]). *If (X, μ, T) is ergodic, then (X, T) is uniquely ergodic and minimal.*

Unique ergodicity means that μ is the unique measure invariant under T . Minimality means that X does not admit any nonempty closed, invariant proper subset for this transformation.

The next result is also due to Parry:

Proposition 7 (Parry [P69]). *Assume that property (H1) is satisfied.*

- (a) (X, μ, T) is ergodic if and only if the rotation (K, m, R) is ergodic.
- (b) In this case, the Kronecker factor (see Section 3) of (X, μ, T) is (K, m, R) , with factor map $\pi: X \rightarrow K$.

2.3. Several commuting transformations. We also need to consider the case of a \mathbb{Z}^k -action spanned by k commuting elements of the group. Proposition 6 and the first part of Proposition 7 generalize in this case.

Proposition 8 (Parry [P69], Leibman [Le02]). *Let $X = G/\Lambda$ be a nilmanifold with Haar measure μ and let $\alpha_1, \dots, \alpha_k$ be commuting elements of G . Assume that*

(H3) G is spanned by the connected component of the identity and $\alpha_1, \dots, \alpha_k$.

Then

- (a) The joint action of $T_{\alpha_1}, \dots, T_{\alpha_k}$ on X is ergodic if and only if the action induced on $K = G/\Lambda U$ is ergodic.
- (b) In this case, X is uniquely ergodic and minimal for this action.

Finally we make use of a general convergence result of Leibman.

Theorem 9 (Leibman [Le02]). *Let $X = G/\Lambda$ be a nilmanifold and assume that $\alpha_1, \dots, \alpha_k$ are commuting elements of G . Then for any continuous function f on X , the average over $n_1 \in [M_1, N_1]$, $n_2 \in [M_2, N_2], \dots, n_k \in [M_k, N_k]$ of*

$$f(T_{\alpha_1}^{n_1} T_{\alpha_2}^{n_2} \dots T_{\alpha_k}^{n_k} x)$$

converges for all $x \in X$ when $N_1 - M_1, N_2 - M_2, \dots, N_k - M_k$ tend to $+\infty$.

The next Proposition says that the Theorem 3 holds for 2-step nilsystems; it follows easily from Theorem 9.

Proposition 10. *Let (X, μ, T) be a 2-step ergodic nilsystem and let $f_\epsilon, \epsilon \in E^*$, be seven bounded functions on X . Then the averages over $n = (n_1, n_2, n_3) \in [M_1, N_1] \times [M_2, N_2] \times [M_3, N_3]$ of*

$$(2) \quad \prod_{\epsilon \in E^*} f_\epsilon(T^{n \cdot \epsilon} x)$$

converge in $L^2(\mu)$ when $N_1 - M_1, N_2 - M_2$ and $N_3 - M_3$ tend to $+\infty$.

Proof. We assume first that the functions f_ϵ are continuous on X . X^7 can be given the structure of a nilmanifold, as the quotient of the 2-step nilpotent Lie group G^7 by the discrete cocompact subgroup Λ^7 . Let $\alpha_{7,1}, \alpha_{7,2}$ and $\alpha_{7,3}$ be the three elements of G^7 defined by:

$$\alpha_{7,1} = (\alpha, 1, \alpha, 1, \alpha, 1, \alpha) ; \alpha_{7,2} = (1, \alpha, \alpha, 1, 1, \alpha, \alpha) ; \alpha_{7,3} = (1, 1, 1, \alpha, \alpha, \alpha, \alpha) .$$

These elements clearly commute. The translations by these elements are the transformations

$$T_{7,1} = T \times \mathbf{Id} \times T \times \mathbf{Id} \times T \times \mathbf{Id} \times T ; T_{7,2} = \mathbf{Id} \times T \times T \times \mathbf{Id} \times \mathbf{Id} \times T \times T \\ \text{and } T_{7,3} = \mathbf{Id} \times \mathbf{Id} \times \mathbf{Id} \times T \times T \times T \times T$$

of X^7 . Let F be the continuous function on X^7

$$F(\tilde{x}) = \prod_{\epsilon \in E^*} f_\epsilon(x_\epsilon) .$$

For every $n = (n_1, n_2, n_3) \in \mathbb{Z}^3$ and $x \in X$, writing $\tilde{\mathbf{x}} = (x, x, x, x, x, x, x) \in X^7$, we have

$$\prod_{\epsilon \in E^*} f_\epsilon(T^{n \cdot \epsilon} x) = F(T_{7,1}^{n_1} T_{7,2}^{n_2} T_{7,3}^{n_3} \tilde{\mathbf{x}}) .$$

By using Theorem 9 at the point $\tilde{\mathbf{x}}$ for the continuous function F , we have that the average of the functions (2) converges at the point x . The convergence in $L^2(\mu)$ follows.

The general case of bounded functions follows from the density of continuous functions in $L^2(\mu)$. \square

As this Proposition is crucial, in order to make this paper a bit more self-contained, in Section 8 we sketch a direct proof of the particular case of Theorem 9 used in its proof.

3. THE KRONECKER FACTOR

Throughout, we assume that (X, μ, T) is an ergodic system.

3.1. Several constructions of the Kronecker factor. The Kronecker factor of (X, μ, T) can be defined in several equivalent ways; we present two classical methods and some others better adapted for explaining the construction of the Conze-Lesigne algebra.

(A) Define

\mathcal{K} to be the sub- σ -algebra of \mathcal{X} spanned by the eigenfunctions of (X, μ, T) .

By definition, \mathcal{K} is a factor and thus there is an associated factor map $\pi: (X, \mu, T) \rightarrow (K, m, R)$. The system (K, m, R) is called the *Kronecker factor* of X . From this definition, it is classical to deduce that K can be given the structure of a compact abelian group endowed with Haar measure m and that the transformation R is a *rotation*, meaning that $R(z) = \beta z$ for some fixed element β of K .

(B) Let \mathcal{I}_2 be the σ -algebra of T_2 -invariant subsets of $X \times X$. Every set A belonging to this σ -algebra is equal up to a set of $\mu \times \mu$ -measure 0 to a set of the form

$$\{(x, y) \in X \times X : \pi(y)\pi(x)^{-1} \in B\}$$

for some $B \subset K$. It follows that:

- \mathcal{K} is the smallest sub- σ -algebra of \mathcal{X} so that every set of \mathcal{I}_2 is measurable with respect to $\mathcal{K} \otimes \mathcal{K}$.

(C) Before giving other characterizations of the Kronecker factor, we introduce some objects defined in [HK02b] that we use throughout the sequel.

Define the measure μ_4 on X^4 to be the relatively independent self-joining of $\mu \times \mu$ over \mathcal{I}_2 . This means that μ_4 is the probability measure on X^4 defined by

$$(3) \quad \int_{X^4} f_{00} \otimes f_{01} \otimes f_{10} \otimes f_{11} d\mu_4 = \int_{X^2} \mathbb{E}(f_{00} \otimes f_{01} | \mathcal{I}_2) \cdot \mathbb{E}(f_{10} \otimes f_{11} | \mathcal{I}_2) d(\mu \times \mu),$$

for $f_{00}, f_{01}, f_{10}, f_{11} \in L^\infty(\mu)$. Note that μ_4 is invariant under the transformation $T_4 = T \times T \times T \times T$.

From this definition and characterization (B) of the Kronecker factor we deduce another characterization of the Kronecker factor:

- The measure μ_4 is relatively independent with respect to \mathcal{K}^4 and the factor \mathcal{K} of X is minimal with this property.

This means that the integral (3) remains unchanged if $\mathbb{E}(f_\epsilon | \mathcal{K})$ is substituted for f_ϵ for each $\epsilon \in D$ and that the factor \mathcal{K} is the smallest one with this property.

Define K_4 to be the closed subgroup

$$K_4 := \{(z, sz, tz, stz) : z, s, t \in K\}$$

of K^4 , let m_4 its Haar measure and let $\pi_4 = \pi \times \pi \times \pi \times \pi: X^4 \rightarrow K^4$.

- When $g_\epsilon, \epsilon \in D$, are four bounded functions on K ,

$$\int_{X^4} (g_{00} \circ \pi) \otimes (g_{01} \circ \pi) \otimes (g_{10} \circ \pi) \otimes (g_{11} \circ \pi) d\mu_4 = \int_{K^4} g_{00} \otimes g_{01} \otimes g_{10} \otimes g_{11} dm_4.$$

This formula is easy to check using the definition (3) of μ_4 when the functions g_ϵ are characters of K . The general case follows by density. From the last two properties we deduce:

- m_4 is the image of μ_4 on K^4 under π_4 .

- When $f_\epsilon, \epsilon \in D$, are four bounded functions on X ,

$$(4) \quad \int_{X^4} f_{00} \otimes f_{01} \otimes f_{10} \otimes f_{11} d\mu_4 = \int_{K^4} \mathbb{E}(f_{00} | K) \otimes \mathbb{E}(f_{01} | K) \otimes \mathbb{E}(f_{10} | K) \otimes \mathbb{E}(f_{11} | K) dm_4 .$$

We identify K^4 with $K \times K^3$ and X^4 with $X \times X^3$ in the natural way. Note that the projection of m_4 on K^3 is equal to $m \times m \times m$. Since μ_4 is relatively independent with respect to \mathcal{K}^4 , we deduce:

- The projection of μ_4 on X^3 is $\mu \times \mu \times \mu$.

Let $\psi: K^3 \rightarrow K$ be defined by $\psi(z_{01}, z_{10}, z_{11}) = z_{01}z_{10}z_{11}^{-1}$. Then if $(z_{00}, z_{01}, z_{10}, z_{11}) \in K_4$, we have that $z_{00} = \psi(z_{01}, z_{10}, z_{11})$ and this relation holds m_4 -almost everywhere. Since m_4 is the projection of μ_4 on K^4 , we have that for any function f on X measurable with respect to \mathcal{K} there exists a function F on X^3 with

$$(5) \quad f(x_{00}) = F(x_{01}, x_{10}, x_{11}) \text{ for } \mu_4\text{-almost every } \mathbf{x} = (x_{00}, x_{01}, x_{10}, x_{11}) \in X^4.$$

Furthermore, F is measurable with respect to $\mathcal{K}^3 := \mathcal{K} \otimes \mathcal{K} \otimes \mathcal{K}$.

Conversely, if f is a bounded function on X and F is a bounded function on X^3 satisfying Equation (5), we have

$$\begin{aligned} \|f\|_{L^2(\mu)}^2 &= \int_{X^4} f(x_{00})F(x_{01}, x_{10}, x_{11}) d\mu_4(\mathbf{x}) \\ &= \int_{X^4} \mathbb{E}(f | \mathcal{K})(x_{00})F(x_{01}, x_{10}, x_{11}) d\mu_4(\mathbf{x}) = \int_X \mathbb{E}(f | \mathcal{K})(x)f(x) d\mu(x) \end{aligned}$$

and it follows that f is measurable with respect to \mathcal{K} . We summarize:

- A subset A of X is measurable with respect to \mathcal{K} if and only if there exists a subset B of X^3 such that $A \times X^3 = X \times B$ up to a set of μ_4 -measure zero.

This characterization of the Kronecker factor serves as a model for the characterization of the Conze-Lesigne algebra.

3.2. More properties of the measure μ_4 . We establish a few more results about the measure μ_4 which were not needed in the preceding discussion but are used in the sequel.

The rotations $R_{4,1} := \mathbf{Id} \times R \times \mathbf{Id} \times R$ and $R_{4,2} := \mathbf{Id} \times \mathbf{Id} \times R \times R$ of K^4 clearly leave the subgroup K_4 invariant and thus also leave the Haar measure m_4 invariant. Since μ_4 is relatively independent with respect to m_4 , we have:

- The measure μ_4 is invariant under the transformations

$$T_{4,1} = \mathbf{Id} \times T \times \mathbf{Id} \times T \text{ and } T_{4,2} = \mathbf{Id} \times \mathbf{Id} \times T \times T .$$

From the symmetries of the group K_4 , in a similar way we can show:

- The measure μ_4 is invariant under the group of isometries of the unit Euclidean square, acting on X^4 by permutation of the coordinates.

From the definition (3) of μ_4 , we have that for any bounded function f on X

$$(6) \quad \begin{aligned} \int_{X^4} f \otimes f \otimes f \otimes f d\mu_4 &= \int_{X^2} (\mathbb{E}(f \otimes f | \mathcal{I}_2))^2 d(\mu \times \mu) \\ &\geq \left(\int_{X^2} \mathbb{E}(f \otimes f | \mathcal{I}_2) d(\mu \times \mu) \right)^2 = \left(\int f d\mu \right)^4 . \end{aligned}$$

By using this inequality with $f = \mathbf{1}_A$, we deduce:

Lemma 11. *For any subset A of X , $\mu_4(A \times A \times A \times A) \geq \mu(A)^4$.*

For $f \in L^\infty(\mu)$ the integral (6) is non-negative and we can define

$$(7) \quad \|f\|_4 := \left(\int_{X^4} f \otimes f \otimes f \otimes f d\mu_4 \right)^{1/4}.$$

By using formula (4) for μ_4 and the definition of K_4 , we have:

- For $f \in L^\infty(\mu)$, $\|f\|_4$ is the ℓ^4 -norm of the Fourier transform of the function $\mathbb{E}(f | K)$:

$$\|f\|_4 = \left(\sum_{\gamma \in \widehat{K}} |\widehat{\mathbb{E}(f | K)}(\gamma)|^4 \right)^{1/4}.$$

In particular $\|\cdot\|_4$ is a semi-norm on $L^\infty(\mu)$.

Furthermore, for $g \in L^\infty(\mu)$ we have $\|g\|_4 = 0$ if and only if $\mathbb{E}(g | K) = 0$ and we deduce another characterization of the Kronecker factor:

- A function $f \in L^2(\mu)$ is measurable with respect to \mathcal{K} if and only if it is orthogonal in $L^2(\mu)$ to every $g \in L^\infty(\mu)$ with $\|g\|_4 = 0$.

Let \mathcal{I}_4 be the invariant σ -algebra of (X^4, μ_4, T_4) . This σ -algebra plays a central role in the definition of the Conze-Lesigne algebra and is also used in the proof of convergence. We use two operators of conditional expectation on $L^2(\mu_4)$: the conditional expectation on the σ -algebra \mathcal{I}_4 and the conditional expectation on the σ -algebra $\mathcal{K}^4 := \mathcal{K} \otimes \mathcal{K} \otimes \mathcal{K} \otimes \mathcal{K}$. These two operators commute.

Lemma 12. *Let $f_\eta \in L^\infty(\mu)$ for $\eta \in D$.*

- If f_γ is measurable with respect to \mathcal{K} for some $\gamma \in D$, then $\mathbb{E}\left(\bigotimes_{\eta \in D} f_\eta \mid \mathcal{I}_4\right)$ is measurable with respect to \mathcal{K}^4 .
- If $\mathbb{E}(f_\delta | \mathcal{K}) = 0$ for some $\delta \in D$, then $\mathbb{E}\left(\mathbb{E}\left(\bigotimes_{\eta \in D} f_\eta \mid \mathcal{I}_4\right) \mid \mathcal{K}^4\right) = 0$.
- If the conditions in both (a) and (b) are satisfied, then $\mathbb{E}\left(\bigotimes_{\eta \in D} f_\eta \mid \mathcal{I}_4\right) = 0$.

Proof. (a) Since each of the four coordinates plays the same role in this proof, we can assume without loss of generality that $\gamma = (0, 0)$. Since f_{00} is measurable with respect to \mathcal{K} we have noticed in Section 3 that there exists a function F on X^3 , measurable with respect to \mathcal{K}^3 , so that

$$f_{00}(x_{00}) = F(x_{01}, x_{10}, x_{11}) \mu_4\text{-a.e.}$$

and we have that

$$\prod_{\eta \in D} f_\eta(x_\eta) = F(x_{01}, x_{10}, x_{11}) \prod_{\eta \in D^*} f_\eta(x_\eta) \mu_4\text{-a.e.}$$

As noted, the 3 dimensional marginal of μ_4 is $\mu \times \mu \times \mu$ and it is classical that the T_3 -invariant sets of $(X^3, \mu \times \mu \times \mu)$ are measurable with respect to \mathcal{K}^3 . Thus $\mathbb{E}\left(\bigotimes_{\eta \in D} f_\eta \mid \mathcal{I}_4\right)$ is measurable with respect to \mathcal{K}^4 .

(b) As the measure μ_4 is relatively independent with respect to m_4 , the hypothesis implies that $\mathbb{E}\left(\bigotimes_{\eta \in D} f_\eta \mid \mathcal{K}^4\right) = 0$. Since the conditional expectations commute, the result follows immediately.

(c) This follows immediately from parts (a) and (b). \square

4. THE CONZE-LESIGNE ALGEBRA

The *Conze-Lesigne algebra* (or *Conze-Lesigne factor*) of an ergodic system (X, μ, T) was introduced by Conze and Lesigne in a series of papers ([CL84], [CL87], [CL88]). Equivalent constructions were given later by Rudolph in [R95], where the term “Conze-Lesigne algebra” was first used, and by the authors in [HK01]. In [HK02b], this σ -algebra appears as the second level in an increasing sequence of factors of X . For differing reasons, none of these constructions is easy to read. The definition in the papers by Conze and Lesigne is not explicit and is spread over several papers (the first ones contained a mistake that is corrected in the later ones); the paper by Rudolph is very technical and [HK02b] deals with a more general setting. We recall here the definitions, omitting some proofs and sketching some others. Unfortunately, a completely self-contained paper would be too lengthy and technical.

4.1. A measure on X^8 . As for the Kronecker factor, the Conze-Lesigne algebra can be defined in several different ways. We give the construction of [HK02b], explaining along the way the relation between this point of view and the presentations of Conze and Lesigne ([CL84], [CL87], [CL88]) and Rudolph [R95].

Recall that \mathcal{I}_4 is the invariant σ -algebra of (X^4, μ_4, T_4) . Define the measure μ_8 on X^8 to be the relatively independent self-joining of μ_4 over \mathcal{I}_4 . This means that μ_8 is the probability measure on X^8 so that if f_ϵ , $\epsilon \in E$, are eight bounded functions on X , then

$$(8) \quad \int_{X^8} \bigotimes_{\epsilon \in E} f_\epsilon d\mu_8 = \int_{X^4} \mathbb{E}\left(\bigotimes_{\substack{\epsilon \in E \\ \epsilon_1=0}} f_\epsilon \mid \mathcal{I}_4\right) \mathbb{E}\left(\bigotimes_{\substack{\epsilon \in E \\ \epsilon_1=1}} f_\epsilon \mid \mathcal{I}_4\right) d\mu_4$$

for $f_{00}, f_{01}, f_{10}, f_{11} \in L^\infty(\mu)$. Note that these conditional expectations are relative to the measure μ_4 on X^4 .

The measure μ_8 is invariant under the transformation $T_8 = T \times \cdots \times T$ (8 times).

We use two properties of μ_8 established in [HK02b]:

- *The measure μ_8 is invariant under the transformations*

$$T_{8,1} = \mathbf{Id} \times T \times \mathbf{Id} \times T \times \mathbf{Id} \times T \times \mathbf{Id} \times T ; \quad T_{8,2} = \mathbf{Id} \times \mathbf{Id} \times T \times T \times \mathbf{Id} \times \mathbf{Id} \times T \times T$$

$$\text{and } T_{8,3} = \mathbf{Id} \times \mathbf{Id} \times \mathbf{Id} \times \mathbf{Id} \times T \times T \times T \times T .$$

- *The measure μ_8 is invariant under the group of isometries of the Euclidean cube, acting on X^8 by permutation of the coordinates.*

4.2. Definition of the Conze-Lesigne algebra. We identify X^8 with $X \times X^7$. For a subset A of X , the subset $A \times X^7$ of X^8 is invariant under the transformations $T_{8,1}, T_{8,2}$ and $T_{8,3}$. One can show that the converse also holds. We have:

- *The σ -algebra of subsets of X^8 invariant under the transformations $T_{8,1}, T_{8,2}$ and $T_{8,3}$ coincides up to μ_8 -null sets with the σ -algebra of sets of the form $A \times X^7$ for $A \subset X$.*

It follows immediately that

- μ_8 is ergodic under the joint action spanned by $T_8, T_{8,1}, T_{8,2}$ and $T_{8,3}$.

We need a bit more notation. Write a point \mathbf{x} of X^8 as $\mathbf{x} = (x_{000}, \tilde{\mathbf{x}})$ with $x_{000} \in X$ and $\tilde{\mathbf{x}} = (x_\epsilon : \epsilon \in E^*) \in X^7$. Let X^7 be endowed with the measure μ_7 , the projection of μ_8 on X^7 and with the transformations $T_{7,1}, T_{7,2}$ and $T_{7,3}$, induced on X^7 by the transformations $T_{8,1}, T_{8,2}$ and $T_{8,3}$, respectively. Let \mathcal{J}_7 be the σ -algebra of subsets of X^7 invariant under these three transformations.

We are now ready to define the Conze-Lesigne algebra. Let B be a subset of X^7 . If B is measurable with respect to \mathcal{J}_7 , then $X \times B$ is invariant under the transformations $T_{8,1}, T_{8,2}$ and $T_{8,3}$ and there exists a subset A of X with $X \times B = A \times X^7$ up to a set of μ_8 -measure zero. Conversely, this property clearly implies that B is measurable with respect to \mathcal{J}_7 . We have:

Lemma 13. *Let X^7 be endowed with the measure μ_7 . A subset B of X^7 is measurable with respect to \mathcal{J}_7 if and only if there exists a subset A of X so that $X \times B = A \times X^7$ up to a set of μ_8 -measure zero.*

This property can be rewritten as:

$$(9) \quad \mathbf{1}_A(x_{000}) = \mathbf{1}_B(\tilde{\mathbf{x}}) \quad \text{for } \mu_8\text{-almost every } \mathbf{x} \in X^8 .$$

This motivates the following definition.

Definition 1. The *Conze-Lesigne algebra* $\mathcal{C}\mathcal{L}$ of (X, μ, T) is the σ -algebra of subsets A of X such that there exists a subset B of X^7 for which $X \times B = A \times X^7$ up to a set of μ_8 -measure zero. (Equivalently, this is the σ -algebra of subsets of X such that relation (9) is satisfied.)

$\mathcal{C}\mathcal{L}$ is clearly invariant under T and thus is a factor of X . By the preceding remarks and the symmetries of the measure μ_8 , we have:

Lemma 14. *For a bounded function f on X , the following are equivalent:*

- (a) $\mathbb{E}(f \mid \mathcal{C}\mathcal{L}) = 0$.
- (b) $\int_{X^8} f(x_{000})g(\tilde{\mathbf{x}}) = 0$ for every bounded function g on X^7 .

Furthermore, the measure μ_8 is relatively independent with respect to its projection on $\mathcal{C}\mathcal{L}^8$, meaning that the integral (8) is unchanged when each function is replaced by its conditional expectation on $\mathcal{C}\mathcal{L}$. Moreover, $\mathcal{C}\mathcal{L}$ is the smallest factor of X with this property.

Similar to the Kronecker factor, the Conze-Lesigne algebra has good behavior with respect to factors. More precisely, if $q: X \rightarrow Y$ is a factor map, then $q^{-1}(\mathcal{C}\mathcal{L}(Y)) = \mathcal{C}\mathcal{L}(X) \cap q^{-1}(\mathcal{Y})$, where by $\mathcal{C}\mathcal{L}(X)$ (respectively, $\mathcal{C}\mathcal{L}(Y)$), we mean the the Conze-Lesigne algebra of X (respectively, of Y).

Definition 2. We write $p: X \rightarrow \text{CL}$ for the corresponding factor map and we call CL the *Conze-Lesigne factor* of X . An ergodic system is a *Conze-Lesigne system* if it is equal to its Conze-Lesigne factor.

In particular, this means that the Conze-Lesigne factor of any ergodic system is a Conze-Lesigne system.

4.3. Invariant functions on $X \times X$ and the Conze-Lesigne equation. By Lemma 14 and the functorial property of Conze-Lesigne algebras, it can be shown that:

Lemma 15. *The σ -algebra \mathcal{I}_4 is measurable with respect to $\mathcal{C}\mathcal{L} \otimes \mathcal{C}\mathcal{L} \otimes \mathcal{C}\mathcal{L} \otimes \mathcal{C}\mathcal{L}$ and $\mathcal{C}\mathcal{L}$ is the smallest factor of X with this property.*

This characterization of the Conze-Lesigne algebra can be viewed as a reformulation in the vocabulary of [HK02b] of the initial construction of Conze and Lesigne. For the remainder of this section, we essentially follow in their footsteps.

Since the measure μ_4 is defined as a relatively independent joining, Lemma 15 allows us to use the machinery of isometric extensions developed by Furstenberg in [F63]. Careful analysis leads to the following description of a Conze-Lesigne system and thus of the Conze-Lesigne factor for any ergodic system:

Lemma 16. *Let (X, μ, T) be a Conze-Lesigne system and let (K, m, R) be its Kronecker factor. Then X is an extension of K by a compact connected abelian group U . Furthermore, for the cocycle $\rho: K \rightarrow U$ defining this extension, there exists a map $F: K^4 \rightarrow U$ so that*

$$(10) \quad \rho(z_{00})\rho(z_{01})^{-1}\rho(z_{10})^{-1}\rho(z_{11}) = F(Rz_{00}, Rz_{01}, Rz_{10}, Rz_{11}) F(z_{00}, z_{01}, z_{10}, z_{11})^{-1}$$

for m_4 -a.e. $(z_{00}, z_{01}, z_{10}, z_{11}) \in K^4$.

The first part of the Lemma means that X can be identified with $K \times U$, endowed with the product of m and the Haar measure of U , and the transformation T is given by $T(z, u) = (Rz, u\rho(z))$. Another way to rephrase the second part of the Lemma is by saying that the mapping on the left side of Equation (10) is a coboundary of the system (K^4, m_4, R_4) .

The second part of this Lemma has an important consequence when U is a finite-dimensional torus:

Proposition 17. *Let (X, μ, T) be a Conze-Lesigne system and assume that it is an extension of its Kronecker factor (K, m, R) by a finite-dimensional torus U . Then the cocycle $\rho: K \rightarrow U$ defining this extension satisfies the Conze-Lesigne equation:*

For every $s \in K$, there exists a map $f: K \rightarrow U$ and a constant $c \in U$ so that

$$(CL) \quad \rho(sz)\rho(z)^{-1} = f(Rz)f(z)^{-1}c.$$

The Equation (CL) is known as the Conze-Lesigne equation.

4.4. Rudolph's approach. The construction of the Conze-Lesigne algebra in [R95] is closely related to the preceding one. It uses the *eigenfunctions* of the Cartesian square of X . As this system is in general not ergodic, this notion needs an explicit definition.

Definition 3. An eigenfunction of $(X \times X, \mu \times \mu, T \times T)$ is a non-zero complex-valued function f on $X \times X$ such that there exists a complex-valued $T \times T$ -invariant function ϕ with

$$(11) \quad f(Tx_0, Tx_1) = f(x_0, x_1)\phi(x_0, x_1) \mu \times \mu\text{-a.e.}$$

Let f and ϕ be as in the definition. The set $E = \{(x_0, x_1) \in X \times X : f(x_0, x_1) \neq 0\}$ is invariant under $T \times T$. By the Poincaré Recurrence Theorem, $|\phi| = 1$ almost everywhere on E . We can obviously assume that $\phi = 0$ on the complement of E . The function $f \otimes \bar{f}$, defined on X^4 by

$$f \otimes \bar{f}(x_{00}, x_{01}, x_{10}, x_{11}) = f(x_{00}, x_{01}) \overline{f(x_{10}, x_{11})}$$

is invariant under T_4 . Therefore it is measurable with respect to $\mathcal{CL} \otimes \mathcal{CL} \otimes \mathcal{CL} \otimes \mathcal{CL}$, for the measure μ_4 . It follows that f is measurable with respect to \mathcal{CL} .

Conversely, Rudolph shows (translated into our vocabulary) that the T_4 -invariant subspace of $L^2(\mu_4)$ is spanned by finite sums of functions of the type $f \otimes \bar{f}$, where f is a bounded eigenfunction. This leads to another characterization of the Conze-Lesigne algebra:

Every eigenfunction of $(X \times X, \mu \times \mu, T \times T)$ is measurable with respect to $\mathcal{CL} \otimes \mathcal{CL}$ and \mathcal{CL} is the smallest factor of X with this property.

4.5. Conze-Lesigne systems and 2-step nilsystems. It can be shown (and it can be deduced from the discussion in Section 8) that any ergodic 2-step nilsystem is a Conze-Lesigne system. However, this result is not needed in this paper. We use a sort of converse in a fundamental way:

Theorem 18. *Every Conze-Lesigne system is the inverse limit of a sequence of 2-step nilsystems.*

This result was originally proven by Conze and Lesigne ([CL84], [CL87], [CL88]); other proofs can be found in Rudolph [R95] and in [HK02b]. We do not give a complete proof, but we outline the main steps, starting with the Conze-Lesigne equation.

Let (X, μ, T) be a Conze-Lesigne system and let K, U and ρ be as in Lemma 16. Since U is a connected compact abelian group, it can be represented as an inverse limit of a sequence of finite-dimensional tori. This means that there exists a decreasing sequence $\{V_i\}$ of closed subgroups of U , with trivial intersection, such that $U_i = U/V_i$ is a finite-dimensional torus for every i .

For each i , let $\rho_i: K \rightarrow U_i$ be the reduction of ρ modulo V_i and let Y_i be the extension of K by U_i associated to the cocycle ρ_i . For each i , Y_i is a factor of X and thus is a Conze-Lesigne system. Furthermore, X is the inverse limit of the sequence $\{Y_i\}$. Therefore it suffices to prove Theorem 18 under the additional hypothesis that U is a finite-dimensional torus. We henceforth assume that this hypothesis holds.

The compact abelian group K can be represented as an inverse limit of a sequence $\{K_j\}$ of compact abelian Lie groups. By using Proposition 17, it can be shown that ρ is cohomologous to some cocycle ρ' which factorizes through K_{j_0} for some j_0 . Therefore, for $j \geq j_0$, ρ' factorizes through K_j and we can define X_j as the extension of K_j associated to this cocycle. Clearly X is the inverse limit of the sequence $\{X_j: j \geq j_0\}$. We conclude that it suffices to restrict to the case that K is a compact abelian Lie group. Henceforth we assume that this hypothesis also holds.

Let G be the family of transformations g of $X = K \times U$ of the form

$$g(z, u) = (sz, uf(z))$$

where s and f satisfy Equation (CL) for some constant $c \in U$.

It is easy to check that G is a group under composition and that it 2-step nilpotent. Note that β and ρ satisfy the Equation (CL) with constant 1 and so the transformation T of X belongs to the group G .

Let this group of transformations of X be endowed with the topology of convergence in probability. This makes it a locally compact group. By using the classical characterization [MZ55] of Lie groups, we have that G is actually a Lie group.

The group G acts on X transitively; this is just a reformulation of Proposition 17. As is easily checked, the stabilizer of an arbitrary point x_0 of X is a discrete cocompact subgroup Λ of G and thus we can identify G with the nilmanifold G/Λ . Moreover, μ is the Haar measure of X and T is translation by the element (β, ρ) of G , where β is the element of K defining the rotation $R: K \rightarrow K$. Thus, we have an identification of X with a 2-step nilsystem. Moreover, μ is the Haar measure of X and T acts on X by translation.

4.6. More about the measure μ_8 . In the sequel we need a few more properties of the measures μ_8 and μ_7 introduced in Section 4 but which were not needed for the construction of the Conze-Lesigne algebra. These properties are used in Sections 6 and 7.1.

4.7. A seminorm on $L^\infty(\mu)$. By the definition (8) of μ_8 , for $f \in L^\infty(\mu)$ we have

$$(12) \quad \int_{X^8} f \otimes \cdots \otimes f d\mu_8 = \int_{X^4} (\mathbb{E}(f \otimes f \otimes f \otimes f \mid \mathcal{I}_4))^2 d\mu_4 \\ \geq \left(\int_{X^4} f \otimes f \otimes f \otimes f d\mu_4 \right)^2 = \|f\|_4^8.$$

Applying this inequality with $f = \mathbf{1}_A$ and using Lemma 11, we have

Lemma 19. *For any subset A of X ,*

$$\mu_8(A \times A \times A \times A \times A \times A \times A \times A) \geq \mu(A)^8.$$

Furthermore, the integral in (12) is non-negative and we can define:

$$(13) \quad \|f\|_8 := \left(\int_{X^8} f \otimes \cdots \otimes f d\mu_8 \right)^{1/8}.$$

By using the Cauchy-Schwarz inequality three times and the symmetries of the measure μ_8 , we have that for eight bounded functions $f_\epsilon, \epsilon \in E$, on X ,

$$(14) \quad \left| \int_{X^8} \bigotimes_{\epsilon \in E} f_\epsilon d\mu_8 \right| \leq \prod_{\epsilon \in E} \|f_\epsilon\|_8.$$

From these inequalities it follows that

- $\|\cdot\|_8$ is subadditive and thus is a seminorm on $L^\infty(\mu)$.

Let f be a bounded function on X . If $\mathbb{E}(f \mid \mathcal{CL}) = 0$ we have $\|f\|_8 = 0$ by definition of the seminorm and Lemma 14. Conversely, if $\|f\|_8 = 0$, by the inequality (14) we have that $\int_{X^8} f(x_{000})g(\tilde{\mathbf{x}}) d\mu_8(\mathbf{x}) = 0$ for any bounded function g on X^7 and it follows from Lemma 14 that $\mathbb{E}(f \mid \mathcal{CL}) = 0$. This gives another characterization of the Conze-Lesigne algebra:

Lemma 20. *For $f \in L^\infty(\mu)$ we have $\mathbb{E}(f \mid \mathcal{CL}) = 0$ if and only if $\|f\|_8 = 0$.*

4.8. **The measure μ_7 on X^7 .** Recall that μ_7 is the projection of μ_8 on X^7 .

Lemma 21. *The measure μ_7 is relatively independent with respect to \mathcal{K}^7 .*

This means that when $f_\epsilon, \epsilon \in E^*$, are seven bounded functions on X , the integral

$$(15) \quad \int_{X^7} \bigotimes_{\epsilon \in E^*} f_\epsilon d\mu_7$$

is unchanged when each of the functions f_ϵ is replaced by its conditional expectation with respect to \mathcal{K} .

Proof. Let $f_\epsilon \in L^\infty(\mu)$ for $\epsilon \in E^*$ and assume that there exists $\eta \in E^*$ with $\mathbb{E}(f_\eta | K) = 0$. We have to show that integral (15) is 0. Define $f_{000} = 1$. By definition of μ_7 and μ_8 , this integral equals

$$\int \mathbb{E}\left(\bigotimes_{\substack{\epsilon \in E \\ \epsilon_1=0}} f_\epsilon | \mathcal{I}_2\right) E\left(\bigotimes_{\substack{\epsilon \in E \\ \epsilon_1=1}} f_\epsilon | \mathcal{I}_2\right) d\mu_4 = \int F_1 F_2 d\mu_4 .$$

Assume first that $\eta_1 = 0$ and consider the first conditional expectation F_1 . Since f_{000} is measurable with respect to K and $\mathbb{E}(f_\eta | \mathcal{K}) = 0$, by Lemma 20, $F_1 = 0$ and the integral above equals 0.

Now assume that $\eta_1 = 1$. Since f_{000} is measurable with respect to K , by part (a) of Lemma 12, F_1 is measurable with respect to K^4 . Since $\mathbb{E}(f_\eta | \mathcal{K}) = 0$, by part (b) of the same Lemma we have that $\mathbb{E}(F_2 | K^4) = 0$ and so the integral is also 0. \square

Recall that \mathcal{J}_7 is the σ -algebra on X^7 consisting of sets that are invariant under $T_{7,1}, T_{7,2}$ and $T_{7,3}$.

Corollary 22. *Let X^7 be endowed with the measure μ_7 . Then \mathcal{J}_7 is measurable with respect to $\mathcal{C}\mathcal{L}^7$.*

Proof. Since $\mathcal{C}\mathcal{L}^7$ is invariant under $T_{7,1}, T_{7,2}$ and $T_{7,3}$, the conditional expectations on $\mathcal{C}\mathcal{L}^7$ and on \mathcal{J}_7 commute. Therefore it suffices to prove that the unique $f \in L^2(\mu_7)$ measurable with respect to \mathcal{J}_7 with $\mathbb{E}(f | \mathcal{C}\mathcal{L}^7) = 0$ is the zero function.

Let f be a function with this property. By Lemma 13 there exists a function g on X such that $f(x_{000}) = g(\tilde{\mathbf{x}})$ for μ_8 -almost every $\mathbf{x} = (x_{000}, \tilde{\mathbf{x}}) \in X^8$. This gives:

$$\int_{X^7} f^2 d\mu_7 = \int_{X^8} g(x_{000}) f(\tilde{\mathbf{x}}) d\mu_8(\mathbf{x}) .$$

By Lemma 14, the measure μ_8 is relatively independent with respect to $\mathcal{C}\mathcal{L}^8$. Since $\mathbb{E}(f | \mathcal{C}\mathcal{L}^7) = 0$, the conditional expectation on $\mathcal{C}\mathcal{L}^8$ of the function in the last integral is 0 and thus the integral is also equal to 0. \square

Corollary 23. *Let X^7 be endowed with the measure μ_7 . Let $f_\epsilon, \epsilon \in E^*$, be seven bounded functions on X and assume that for some $\eta \in E^*$ we have $\mathbb{E}(f_\eta | \mathcal{C}\mathcal{L}) = 0$. Then $\mathbb{E}(\bigotimes_{\epsilon \in E^*} f_\epsilon | \mathcal{J}_7) = 0$.*

Proof. Let $g_\epsilon, \epsilon \in E^*$, be seven bounded functions on X , measurable with respect to $\mathcal{C}\mathcal{L}$. By Lemma 21,

$$\int_{X^7} \bigotimes_{\epsilon \in E^*} f_\epsilon \cdot \bigotimes_{\epsilon \in E^*} g_\epsilon d\mu_7 = \int_{X^7} \bigotimes_{\epsilon \in E^*} \mathbb{E}(f_\epsilon g_\epsilon | \mathcal{K}) d\mu_7 .$$

We have $\mathbb{E}(f_\eta g_\eta \mid \mathcal{CL}) = 0$ for some $\eta \in E^*$. Thus $\mathbb{E}(f_\eta g_\eta \mid \mathcal{K}) = 0$ and the last integral is 0.

This show that

$$\mathbb{E}\left(\bigotimes_{\epsilon \in E^*} f_\epsilon \mid \mathcal{CL}^7\right) = 0$$

and the statement follows from Corollary 22. \square

5. BERGELSON'S THEOREM

We now prove Bergelson's two-dimensional generalization of Khintchine's Theorem [B00]. We consider sequences of averages of the form

$$(16) \quad \frac{1}{(N_1 - M_1)(N_2 - M_2)} \sum_{n_1=M_1}^{N_1} \sum_{n_2=M_2}^{N_2} f_{01}(T^{n_1}x) f_{10}(T^{n_2}x) f_{11}(T^{n_1+n_2}x),$$

where $f_{01}, f_{10}, f_{11} \in L^\infty(X)$ for some ergodic probability measure-preserving system (X, \mathcal{B}, μ, T) .

Proposition 24. *When $N_1 - M_1$ and $N_2 - M_2$ tend to $+\infty$, the average (16) converges to 0 in $L^2(X)$ whenever $\mathbb{E}(f_\eta \mid K) = 0$ for at least one $\eta \in D^*$.*

In other words, the Kronecker factor is characteristic for the convergence in $L^2(X)$ of the average (16).

Proof. Write

$$u_{n_1, n_2}(x) = f_{01}(T^{n_1}x) f_{10}(T^{n_2}x) f_{11}(T^{n_1+n_2}x).$$

For $k, \ell \in \mathbb{Z}$ we have

$$\begin{aligned} & \int u_{n_1+k, n_2+\ell}(x) u_{n_1, n_2}(x) d\mu(x) \\ &= \int (f_{01} \circ T^k \cdot f_{01})(T^{-n_1}x) (f_{10} \circ T^\ell \cdot f_{10})(T^{-n_2}x) (f_{11} \circ T^{k+\ell} \cdot f_{11})(x) d\mu(x). \end{aligned}$$

Taking the average over n_1 and n_2 , we have that as $N_1 - M_1$ and $N_2 - M_2$ tend to $+\infty$, this converges to

$$\gamma_{k, \ell} = \int f_{01} \circ T^k \cdot f_{01} d\mu \int f_{10} \circ T^\ell \cdot f_{10} d\mu \int f_{11} \circ T^{k+\ell} \cdot f_{11} d\mu.$$

If $\mathbb{E}(f_\eta \mid K) = 0$ for some $\eta \in D^*$, then

$$\frac{1}{L^2} \sum_{k, \ell=0}^{L-1} |\gamma_{k, \ell}| \rightarrow 0$$

as $L \rightarrow \infty$, and the result follows from the two-dimensional van der Corput Lemma (see the Appendix for the statement). \square

Proposition 25 (Bergelson). *The averages (16) converge in $L^2(\mu)$ to*

$$\iint_{K \times K} \mathbb{E}(f_{01} \mid K)(\pi(x)) \mathbb{E}(f_{10} \mid K)(s\pi(x)) \mathbb{E}(f_{11} \mid K)(st\pi(x)) dm(s) dm(t).$$

Proof. By Proposition 24, it suffices to prove the result when the functions f_i are measurable with respect to the Kronecker factor. By density and linearity, we can restrict ourselves to the case that these functions are characters of K , and the result is obvious in this case. \square

By Proposition 25 we have

Theorem 26. *If $f_{00}, f_{01}, f_{10}, f_{11} \in L^\infty(\mu)$, then the average for $n_1 \in [M_1, N_1]$ and $n_2 \in [M_2, N_2]$ of*

$$(17) \quad \int f_{00}(x) f_{01}(T^{n_1}x) f_{10}(T^{n_2}x) f_{11}(T^{n_1+n_2}x) d\mu(x)$$

converges to

$$\int_{X^4} f_{00} \otimes f_{01} \otimes f_{10} \otimes f_{11} d\mu_4$$

as $N_1 - M_1$ and $N_2 - M_2$ tend to $+\infty$.

Combining Theorem 26 applied to the characteristic function of a set E and Lemma 11, we deduce Bergelson's two-dimensional generalization of Khintchine recurrence.

6. CONVERGENCE FOR THE INTEGRAL OF SEVEN TERMS

Given seven bounded functions f_ϵ , $\epsilon \in E^*$, we consider averages over $n_1 \in [M_1, N_1]$, $n_2 \in [M_2, N_2]$ and $n_3 \in [M_3, N_3]$ of

$$(18) \quad \int \prod_{\epsilon \in E^*} f_\epsilon \circ T^{n_\epsilon} d\mu$$

and take the limit when $N_1 - M_1$, $N_2 - M_2$ and $N_3 - M_3$ tend to $+\infty$.

6.1. A characteristic factor for the integral of seven terms.

Lemma 27. *Let $f_\eta \in L^\infty(\mu)$ for $\eta \in D$. Then the lim sup as $N_1 - M_1 \rightarrow \infty$, $N_2 - M_2 \rightarrow \infty$, $N_3 - M_3 \rightarrow \infty$ of*

$$\frac{1}{(N_1 - M_1)(N_2 - M_2)} \sum_{n_1=M_1}^{N_1} \sum_{n_2=M_2}^{N_2} \int \left| \frac{1}{N_3 - M_3} \sum_{n_3=M_3}^{N_3} \prod_{\eta \in D} f_\eta \circ T^{n_\eta - n_3} \right|^2 d\mu$$

is less than or equal to

$$\int |\mathbb{E}(\bigotimes_{\eta \in D} f_\eta \mid \mathcal{I}_4)|^2 d\mu_4.$$

Proof. Without loss of generality, we can assume that each $|f_\eta|$ is bounded by 1. Fix an integer $L > 0$. By the finite version of the van der Corput Lemma (see Appendix), for each $n = (n_1, n_2)$ the integral in the statement is bounded by

$$\frac{4L}{N_3 - M_3} + \sum_{\ell=-L}^L \frac{L - |\ell|}{L^2} \int \prod_{\eta \in D} (f_\eta \circ T^{n_\eta + \ell} \cdot f_\eta \circ T^{n_\eta}) d\mu.$$

Thus the integral is bounded by

$$\sum_{\ell=-L}^L \frac{L - |\ell|}{L^2} \limsup_{\substack{N_1 - M_1 \rightarrow \infty \\ N_2 - M_2 \rightarrow \infty}} \frac{1}{(N_1 - M_1)(N_2 - M_2)} \sum_{n_1=M_1}^{N_1} \sum_{n_2=M_2}^{N_2} \int \prod_{\eta \in D} (f_\eta \cdot f_\eta \circ T^\ell) \circ T^{n_\eta} d\mu$$

and by Theorem 26, this is

$$\sum_{\ell=-L}^L \frac{L - |\ell|}{L^2} \int \bigotimes_{\eta \in D} (f_\eta \cdot f_\eta \circ T^\ell) d\mu_4.$$

Taking the limit as $L \rightarrow \infty$, we obtain the bound. \square

Lemma 28. *The Kronecker factor is characteristic for the average of the integral (18). In other words, this average converges to 0 whenever $\mathbb{E}(f_\eta | K) = 0$ for at least one $\eta \in E^*$.*

Proof. Assume first that $\mathbb{E}(f_\epsilon | K) = 0$ for some $\epsilon \in \{(0, 0, 1), (0, 1, 0), (0, 1, 1)\}$. Set $g_\eta = f_{0\eta_1\eta_2}$ for $\eta \in D^*$ and set $g_{11} = 1$. Then g_{11} is measurable with respect to K and at least one of the other three g_η has 0 conditional expectation on K . By Lemma 20, $\mathbb{E}(\otimes_{\eta \in D} g_\eta | \mathcal{I}_4) = 0$. By Lemma 27,

$$\frac{1}{(N_1 - M_2)(N_2 - M_2)} \sum_{n_1=M_1}^{N_1} \sum_{n_2=M_2}^{N_2} \int \left| \frac{1}{N_3 - M_3} \sum_{n_3=M_3}^{N_3} \prod_{\eta \in D} g_\eta \circ T^{n \cdot \eta - n_3} \right|^2 d\mu \rightarrow 0,$$

as $N_1 - M_1 \rightarrow \infty$, $N_2 - M_2 \rightarrow \infty$ and $N_3 - M_3 \rightarrow \infty$. We can rewrite the average of the integral (18) as the average over $n_1 \in [M_1, N_1]$ and $n_2 \in [M_2, N_2]$ of

$$\int \left(\frac{1}{N_3 - M_3} \sum_{n_3=M_3}^{N_3} \prod_{\eta \in D} g_\eta \circ T^{n_1\eta_1 + n_2\eta_2 - n_3} \right) \left(\prod_{\epsilon \in E, \epsilon_0=1} f_\epsilon \circ T^{n_1\epsilon_1 + n_2\epsilon_2} \right) d\mu.$$

Applying Cauchy-Schwartz to this average and using the limit above, this average converges to 0.

By obvious modifications, the same result holds when $\mathbb{E}(f_\epsilon | K) = 0$ for some $\epsilon \in \{(1, 0, 0), (1, 0, 1), (1, 1, 0)\}$.

Finally, we consider the case that $\mathbb{E}(f_{111} | K) = 0$. By the preceding steps, we can assume that each of the other functions is measurable with respect to K . Define $h_\eta = f_{1\eta_1\eta_2}$ for $\eta \in D$. Then h_{00} is measurable with respect to K and $\mathbb{E}(h_{11} | K) = 0$. By again using Lemma 20, we have $\mathbb{E}(\otimes_{\eta \in D} h_\eta | \mathcal{I}_4) = 0$. We conclude, as above, by using Lemma 27 and the Cauchy-Schwarz inequality. \square

6.2. The limit for seven terms. Let m_7 be the projection of the measure μ_7 on K^7 . It can easily be checked that m_7 is the Haar measure of the closed subgroup

$$K_7 = \{(za, zb, zab, zc, zac, zbc, zabc) : z, a, b, c \in K\}$$

of K^7 . Thus for seven bounded functions g_ϵ , $\epsilon \in E^*$, on K we have

$$\int \bigotimes_{\epsilon \in E^*} g_\epsilon dm_7 = \int_{K^4} \prod_{\epsilon \in E^*} g_\epsilon(zs_1^{\epsilon_1} s_2^{\epsilon_2} s_3^{\epsilon_3}) dm(z) dm(s_1) dm(s_2) dm(s_3).$$

Proposition 29. *Let $f_\epsilon \in L^\infty(\mu)$. The average of the integral (18) converges to*

$$\int_{X^7} \bigotimes_{\epsilon \in E^*} f_\epsilon d\mu_7.$$

Proof. Using Lemma 21, this last integral can be rewritten as

$$\int_{K^7} \bigotimes_{\epsilon \in E^*} \mathbb{E}(f_\epsilon | K) dm_7,$$

By Lemma 28, it suffices to prove the result when all f_ϵ are measurable with respect to K . By density we can restrict to the case that each f_ϵ is a character of K , and the result is obvious in this case. \square

7. REDUCTION FOR THE L^2 CONVERGENCE WITH SEVEN TERMS

We now study averages over $n_1 \in [M_1, N_1]$, $n_2 \in [M_2, N_2]$ and $n_3 \in [M_3, N_3]$ of

$$(19) \quad \prod_{\epsilon \in E^*} f_\epsilon(T^{n \cdot \epsilon} x)$$

and take the limit as $N_1 - M_1, N_2 - M_2$ and $N_3 - M_3$ tend to $+\infty$.

7.1. A characteristic factor for the three-dimensional cube.

Lemma 30. *The factor \mathcal{CL} of X is characteristic for the convergence in $L^2(\mu)$ of the average of the product (19). This means that this average converges to 0 in $L^2(\mu)$ whenever $\mathbb{E}(f_\epsilon | \mathcal{CL}) = 0$ for some $\epsilon \in E^*$.*

Proof. For $n = (n_1, n_2, n_3) \in \mathbb{Z}^3$, set

$$u_n = \prod_{\epsilon \in E^*} f_\epsilon \circ T^{n \cdot \epsilon}$$

and apply the 3-dimensional van der Corput Lemma (see Appendix). For $k \in \mathbb{Z}^3$, we have

$$\int u_{n+k} u_n d\mu = \int \prod_{\epsilon \in E^*} (f_\epsilon \circ T^{k \cdot \epsilon} \cdot f_\epsilon) \circ T^{n \cdot \epsilon} d\mu .$$

By Proposition 29, for $k = (k_1, k_2, k_3) \in \mathbb{Z}^3$ and $n = (n_1, n_2, n_3)$, the averages for $n_1 \in [M_1, N_1]$, $n_2 \in [M_2, N_2]$ and $n_3 \in [M_3, N_3]$ of

$$\int u_{n+k} u_n d\mu$$

converge to γ_k , where

$$\gamma_k = \int_{X^7} \prod_{\epsilon \in E^*} f_\epsilon(T^{k \cdot \epsilon} x_\epsilon) f_\epsilon(x_\epsilon) d\mu_7 = \int_{X^7} F(T_{7,1}^{k_1} T_{7,2}^{k_2} T_{7,3}^{k_3} x) F(x) d\mu_7(x)$$

and

$$F(x) = \prod_{\epsilon \in E^*} f_\epsilon(x_\epsilon)$$

for $x = (x_\epsilon; \epsilon \in E^*) \in X^7$. Thus the average of γ_k with $k_1, k_2, k_3 \in [0, K]$ converges to the square of the $L^2(\mu_7)$ -norm of $\mathbb{E}(F | \mathcal{J}_7)$. The result follows from Corollary 23. \square

7.2. Reduction to nilsystems for Theorem 3. Let $f_\epsilon, \epsilon \in E$, be seven bounded functions on X . By Lemma 30, the difference between the average of the product (19) and the same average with $\mathbb{E}(f_\epsilon | \mathcal{CL})$ substituted for each f_ϵ , converges to 0 in $L^2(\mu)$. Therefore, in order to prove Theorem 3 it suffices to prove it when all the functions are measurable with respect to \mathcal{CL} . Thus we can restrict to the case that the system itself is a \mathcal{CL} -system, meaning that the system is equal to its Conze-Lesigne algebra.

Furthermore, a \mathcal{CL} -system is an inverse limit of 2-step nilsystems. By density, it suffices to prove Theorem 3 for such systems. This is done in Section 8 (Proposition 10).

7.3. The integral with eight terms.

Theorem 31. *Let f_ϵ , $\epsilon \in E$, be eight bounded functions on the ergodic system (X, μ, T) . Then the average for $n_1 \in [M_1, N_1]$, $n_2 \in [M_2, N_2]$ and $n_3 \in [M_3, N_3]$ of*

$$(20) \quad \int_X \prod_{\epsilon \in E} f_\epsilon(T^{n_\epsilon} x) d\mu(x)$$

converges to

$$\int_{X^8} \bigotimes_{\epsilon \in E} f_\epsilon d\mu_8$$

when $N_1 - M_1$, $N_2 - M_2$, and $N_3 - M_3$ tend to $+\infty$.

This Theorem is proven in Section 8 when the system is a 2-step nilsystem. We now explain how to deduce the general case.

By Lemma 30, the difference between the average in (20) and the same average with each function f_ϵ replaced by $\mathbb{E}(f_\epsilon | \mathcal{CL})$ converges to zero. By Lemma 14, the above integral remains unchanged under this modification of the functions. Thus it suffices to prove the Theorem with CL substituted for X and so we can assume without loss that X is a Conze-Lesigne system. In this case the system is the inverse limit of a sequence of 2-step nilsystems and the result follows by density. \square

Combining Lemma 19 and the above theorem, we have the statement of Theorem 4.

8. CONVERGENCE FOR A NILSYSTEM

We are left with showing that Theorem 31 holds for ergodic 2-step nilsystems.

The proof uses a precise description of the measures μ_4 and μ_8 introduced in Section 4 in the particular case of a 2-step nilsystem. This discussion allows us also to sketch a proof of Theorem 9 in the particular case we consider.

Throughout this section, (X, μ, T) is an ergodic 2-step nilsystem and we maintain the notation of Section 2. We assume that the hypotheses (H1) and (H2) are satisfied.

8.1. The nilmanifold X_4 .

$$G_4 = \{ \mathbf{g} = (g_{00}, g_{01}, g_{10}, g_{11}) \in G^4 : g_{00}g_{01}^{-1}g_{10}^{-1}g_{11} \in U \} ;$$

$$\Lambda_4 = \{ \boldsymbol{\lambda} = (\lambda_{00}, \lambda_{01}, \lambda_{10}, \lambda_{11}) \in \Lambda^4 : \lambda_{00}\lambda_{01}^{-1}\lambda_{10}^{-1}\lambda_{11} = 1 \} = \Lambda^4 \cap G_4 .$$

Using the nilpotency of G , it is easy to check that G_4 is a subgroup of G^4 and it is clearly 2-step nilpotent. It is also easy to verify that it is a closed Lie subgroup. The commutator subgroup of G_4 is clearly included in U^4 and we claim that it equals U^4 . For $g \in G$, the three elements (g, g, g, g) , $(1, g, 1, g)$ and $(1, 1, g, g)$ of G^4 belong to G_4 . Taking the commutators of elements of this kind, we get that for $u \in U$, the four elements (u, u, u, u) , $(1, u, 1, u)$, $(1, 1, u, u)$ and $(1, 1, 1, u)$ of U^4 belong to the commutator subgroup of G_4 and the claim follows.

Λ_4 is a discrete cocompact subgroup of G_4 , with $\Lambda_4 \cap U^4 = \{1\}$. There is a natural embedding of the nilmanifold G_4/Λ_4 in X^4 . We identify this nilmanifold with its image

$$(21) \quad X_4 = \{ \mathbf{x} = (x_{00}, x_{01}, x_{10}, x_{11}) \in X^4 : \pi(x_{00})\pi(x_{01})^{-1}\pi(x_{10})^{-1}\pi(x_{11}) = 1 \}$$

in X^4 .

Lemma 32. μ_4 is the Haar measure of X_4 .

Proof. Let μ'_4 be the Haar measure of X_4 . The transformations T_4 , $T_{4,1}$ and $T_{4,2}$ of X^4 are the translations by the elements

$$\alpha_4 = (\alpha, \alpha, \alpha, \alpha) ; \alpha_{4,1} = (1, \alpha, 1, \alpha) \text{ and } \alpha_{4,2} = (1, 1, \alpha, \alpha)$$

of G^4 , respectively, and these three elements belong to G_4 . Therefore, these transformations are translations on the nilmanifold $X_4 = G_4/\Lambda_4$ and in particular leave the measure μ'_4 invariant.

We claim that μ'_4 is ergodic for the action spanned by these transformations. We use Proposition 8. First we observe that the hypothesis (H3) of this Proposition is satisfied because G and α satisfy hypothesis (H1). Let $q_4: G_4 \rightarrow K^4$ be the map

$$q_4(g_{00}, g_{01}, g_{10}, g_{11}) = (\pi(g_{00}), \pi(g_{01}), \pi(g_{10}), \pi(g_{11})) .$$

Then q_4 is a continuous group homomorphism, its range is the subgroup K_4 of K^4 defined in Section 3 and its kernel is $(\Lambda^4 U^4) \cap G_4 = \Lambda_4 U^4$. Thus we can identify $G_4/U^4 \Lambda_4$ with K_4 .

Under this identification, the transformations induced by T_4 , $T_{4,1}$ and $T_{4,2}$ on K_4 are the rotations by $\beta_4 = (\beta, \beta, \beta, \beta)$, $\beta_{4,1} = (1, \beta, 1, \beta)$ and $\beta_{4,2} = (1, 1, \beta, \beta)$, respectively, where as usual, $\beta = q(\alpha)$. The rotation R by β on K is ergodic and thus the subgroup spanned by β is dense in K . It follows that the subgroup spanned by β_4 , $\beta_{4,1}$ and $\beta_{4,2}$ is dense in K_4 and so the joint action of these rotations on K_4 is ergodic.

The hypotheses of Proposition 8 are satisfied and our claim is proven.

By Proposition 6, μ'_4 is the unique measure on X_4 invariant under T_4 , $T_{4,1}$ and $T_{4,2}$. The measure μ_4 also is invariant under these transformations and it follows immediately from its definition (4) and the definition (21) of X_4 that it is concentrated on X_4 . Therefore $\mu_4 = \mu'_4$. \square

8.2. The nilmanifold X_8 . Define

$$H = \{(h, hu, hv, huv) : h \in G; u, v \in U\} .$$

It is easy to verify that H is a normal subgroup of G_4 . Furthermore, it is a closed Lie subgroup of G^4 . Define

$$G_8 = \left\{ (g_{00}, g_{01}, g_{10}, g_{11}, h_{00}g_{00}, h_{01}g_{01}, h_{10}g_{10}, h_{011}g_{11}) : \right. \\ \left. (g_{00}, g_{01}, g_{10}, g_{11}) \in G_4 ; (h_{00}, h_{01}, h_{10}, h_{11}) \in H \right\} .$$

Then G_8 is a closed Lie subgroup of G^8 . It is clearly 2-step nilpotent and it can be checked that G_8 is the set of $\mathbf{g} = (g_\epsilon; \epsilon \in E) \in G^8$ that satisfy the four properties:

$$g_{000} g_{001}^{-1} g_{011} g_{010}^{-1} g_{110} g_{111}^{-1} g_{101} g_{100}^{-1} = 1 ; (g_{000}, g_{001}, g_{010}, g_{011}) \in G_4 ; \\ (g_{000}, g_{001}, g_{100}, g_{101}) \in G_4 ; (g_{000}, g_{010}, g_{100}, g_{110}) \in G_4 .$$

Define

$$\Lambda_8 = \Lambda^8 \cap G^8 ; \\ U_8 = U^8 \cap G_8 \\ = \{ \mathbf{u} = (u_\epsilon; \epsilon \in E) : u_{000} u_{001}^{-1} u_{011} u_{010}^{-1} u_{110} u_{111}^{-1} u_{101} u_{100}^{-1} = 1 \} .$$

Λ_8 is a discrete cocompact subgroup of G_8 . Write X_8 for the nilmanifold G_8/Λ_8 . There is a natural embedding of X_8 in X^8 and we identify X_8 with its image in X^8 .

Proceeding as in the preceding section, it can be checked that the commutator subgroup of G_8 is U_8 . Note that $U_8 \cap \Lambda_8 = \{1\}$.

Lemma 33. *μ_8 is the Haar measure of the nilmanifold X_8 .*

Proof. We proceed as in the proof of Lemma 32. Let μ'_8 be the Haar measure of X_8 .

The four elements

$$\begin{aligned} \alpha_8 &= (\alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha, \alpha) ; \alpha_{8,1} = (1, 1, 1, 1, \alpha, \alpha, \alpha, \alpha) \\ \alpha_{8,2} &= (1, 1, \alpha, \alpha, 1, 1, \alpha, \alpha) ; \alpha_{8,3} = (1, \alpha, 1, \alpha, 1, \alpha, 1, \alpha) \end{aligned}$$

of G^8 belong to G_8 and thus the corresponding transformations $T_8, T_{8,1}, T_{8,2}$ and $T_{8,3}$ are translations on the nilmanifold $X_8 = G_8/\Lambda_8$.

We show that μ'_8 is ergodic for the joint action of these transformations. Note that hypothesis (H3) of Proposition 8 is satisfied. Let $q_8 = G_8 \rightarrow K^8$ be the map

$$q_8(g\epsilon : \epsilon \in E) = (q(g\epsilon) : \epsilon \in E) .$$

Then q_8 is a group homomorphism. Its kernel is $(\Lambda^8 U^8) \cap G_8$ and it is not difficult to check that this group is equal to $\Lambda_8 U_8$. The range of q_8 is the closed subgroup

$$\begin{aligned} K_8 &= \{(z_{000}, z_{001}, z_{010}, z_{011}, cz_{000}, cz_{001}, cz_{010}, cz_{011}) : \\ &\quad (z_{000}, z_{001}, z_{010}, z_{011}) \in K_4, c \in K\} \\ &= \{(z, za, zb, zab, zc, zac, zbc, abc) : z, a, b, c \in K\} \end{aligned}$$

of K^8 . Thus we can identify $G_8/\Lambda_8 U_8$ with K_8 . Under this identification, the transformations induced by $T_8, T_{8,1}, T_{8,2}$ and $T_{8,3}$ on K_8 are the rotations by the elements $\beta_8, \beta_{8,1}, \beta_{8,2}$ and $\beta_{8,3}$, the images of $\alpha_8, \alpha_{8,1}, \alpha_{8,2}$ under q . The joint action of these transformations on K_8 is clearly ergodic.

By Proposition 8, the joint action of $T_8, T_{8,1}, T_{8,2}$ and $T_{8,3}$ on X_8 is ergodic. By Proposition 6, μ'_4 is the unique measure on X_8 invariant under these transformations. Since the measure μ_4 is also invariant under these transformations, we are left with proving that it is concentrated on X_8 .

It can be checked that $H\Lambda_4$ is a closed subgroup of G_4 . Let Y denote the compact space $G_4/H\Lambda_4$ and let θ be the natural projection of $X_4 = G_4/\Lambda_4$ onto Y . Since $\alpha_4 \in H$, the mapping θ is invariant under T_4 .

Recall that the measure μ_8 is the relatively independent self-joining of μ_4 over the T_4 -invariant σ -algebra \mathcal{I}_4 . Since the measure μ_4 is concentrated on X_4 , the measure μ_8 is concentrated on $X_4 \times X_4$. Furthermore, the mapping θ is invariant and so we have that $\theta(\mathbf{x}') = \theta(\mathbf{x}'')$ for μ_8 -almost every $\mathbf{x} = (\mathbf{x}', \mathbf{x}'') \in X_4 \times X_4$. Therefore the measure μ_8 is concentrated on the set

$$\{\mathbf{x} = (\mathbf{x}', \mathbf{x}'') \in X_4 \times X_4 : \theta(\mathbf{x}') = \theta(\mathbf{x}'')\} .$$

However, this set is actually equal to X_8 and so we have the statement of the Lemma. \square

8.3. Proof of Theorem 31 for a 2-step nilsystem. The measure μ_8 is ergodic with respect to the joint action of the translations $T_{8,1}, T_{8,2}, T_{8,3}$ and T_8 (see Section 4.2). By Proposition 8, X_8 is uniquely ergodic for this action. Let $f_\epsilon, \epsilon \in E$, be eight continuous functions on X . Since the diagonal $\Delta_8 = \{(x, x, x, x, x, x, x, x) : x \in X\}$ is included in X_8 , the average over $n_1 \in [M_1, N_1], n_2 \in [M_2, N_2], n_3 \in [M_3, N_3]$ and $p \in [A, B]$ of

$$\prod_{\epsilon \in E} f_\epsilon(T^{p+n \cdot \epsilon} x)$$

converges uniformly to $\int \otimes_{\epsilon \in E} f_\epsilon d\mu_8$ when $N_1 - M_1, N_2 - M_2, N_3 - M_3$ and $B - A$ tend to $+\infty$.

Taking the integral, the announced result holds for continuous functions. The general case follows by density. \square

8.4. Sketch of a proof of Theorem 9 in a particular case. We sketch a proof of Theorem 9 for translations $T_{8,1}, T_{8,2}$ and $T_{8,3}$ on the nilmanifold X^8 .

For $x \in X$, write

$$X_{8,x} = \{\mathbf{x} = (x_\epsilon; \epsilon \in E) \in X_8 : x_{000} = x\} .$$

These sets form a closed partition of X_8 . We give each of them the structure of a nilmanifold. Define

$$G'_8 = \{\mathbf{g} = (g_\epsilon; \epsilon \in E) \in G_8 : g_{000} = 1\} .$$

It is a closed Lie subgroup of G_8 and we note that for every $x \in X$, left translations by elements of this group leave $X_{8,x}$ invariant and act transitively on this space. It is now easy to identify $X_{8,x}$ with a nilmanifold G'_8/Λ'_x for a discrete cocompact subgroup Λ'_x of G'_8 . Let $\mu_{8,x}$ be the Haar measure of $X_{8,x}$. This measure is invariant under G'_8 and so invariant under the transformations $T_{8,1}, T_{8,2}$ and $T_{8,3}$, which are translations by elements of G'_8 . We have

$$(22) \quad \mu_8 = \int_X \mu_{8,x} d\mu(x) ,$$

since the right hand side of this equation is a measure on X_8 and it is invariant under the action of G_8 .

We give a second interpretation of this formula. Let $\pi_{000}: X^8 \rightarrow X$ be the first projection. The family of measures $(\mu_{8,x}; x \in X)$ is a regular version of the conditional probability measure given the σ -algebra $\pi_{000}^{-1}(\mathcal{X})$. Recall that this σ -algebra coincides with \mathcal{J}_8 up to μ_8 -null sets (see Section 4.2). Thus Equation (22) can be viewed as the ergodic disintegration of μ_8 for the action spanned by $T_{8,1}, T_{8,2}$ and $T_{8,3}$.

It follows that for μ -almost every x , the measure $\mu_{8,x}$ is ergodic for this action. By Proposition 8 applied to the nilmanifold $X_{8,x}$, we have that for μ -almost every x , $X_{8,x}$ is uniquely ergodic for the action spanned by the three transformations $T_{8,1}, T_{8,2}$ and $T_{8,3}$. Therefore, for any continuous function F on $X_{8,x}$, for μ -almost every $x \in X$, the average over $n_1 \in [M_1, N_1], n_2 \in [M_2, N_2]$ and $n_3 \in [M_3, N_3]$ of

$$F(T_{8,1}^{n_1} T_{8,2}^{n_2} T_{8,3}^{n_3} x)$$

converges to $\int F d\mu_{8,x}$. The result follows. \square

APPENDIX A. VAN DER CORPUT LEMMA

See Bergelson [B86] for details about the van der Corput Lemma. We use a finite version of the Hilbert space version of this Lemma.

Proposition 34. *Assume that $\{x_n\}$ is a sequence in a Hilbert space with norm $\|x_n\| \leq 1$ for all $n \in \mathbb{Z}$. Let L, M and N be integers with $L > 0$ and $N > M$. Then*

$$\left\| \frac{1}{N-M} \sum_{n=M}^{N-1} x_n \right\|^2 \leq \frac{4L}{N-M} + \sum_{\ell=-L}^L \frac{L-|\ell|}{L^2} \frac{1}{N-M} \sum_{n=M}^{N-1} \langle x_n, x_{n+\ell} \rangle .$$

We also use an infinite version of the van der Corput Lemma for dimensions 1, 2 and 3. We state it only for three dimensions, as the other cases are similar and simpler.

Proposition 35. *Assume that $\{u_{n_1, n_2, n_3} : n_1, n_2, n_3 \in \mathbb{Z}\}$ is a bounded triple sequence of vectors in a Hilbert space. If*

$$\lim_{K \rightarrow \infty} \frac{1}{K^3} \sum_{k_1, k_2, k_3=0}^{K-1} \left| \lim_{N-M \rightarrow \infty} \frac{1}{(N-M)^3} \sum_{n_1, n_2, n_3=M}^{N-1} \langle x_{n_1, n_2, n_3}, x_{n_1+k_1, n_2+k_2, n_3+k_3} \rangle \right|$$

is 0, then

$$\lim_{N \rightarrow \infty} \left\| \frac{1}{(N-M)^3} \sum_{n_1, n_2, n_3=M}^{N-1} x_{n_1, n_2, n_3} \right\| = 0 .$$

REFERENCES

- [AGH63] L. Auslander, L. Green and F. Hahn. Flows on homogeneous spaces. *Ann. Math. Studies* **53**, Princeton Univ. Press (1963).
- [B86] V. Bergelson. A density statement about generalizing Schur's Theorem. *J. Combinatorial Theory (A)* **43** (1986), 338-343.
- [B00] V. Bergelson. The multifarious Poincaré Recurrence Theorem *Descriptive Set Theory and Dynamical Systems*, Eds. M. Foreman, A.S. KeCHRIS, A. Louveau, B. Weiss. Cambridge University Press, New York (2000), 31-57.
- [CL84] J.-P. Conze and E. Lesigne. Théorèmes ergodiques pour des mesures diagonales. *Bull. Soc. Math. France*, **112** (1984), 143-175.
- [CL87] J.-P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *Publications de l'Institut de Recherche de Mathématiques de Rennes, Probabilités*, 1987.
- [CL88] 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.
- [F61] H. Furstenberg. Strict ergodicity and transformations of the torus. *Amer. J. Math.*, **83** (1961), 573-601.
- [F63] H. Furstenberg. The structure of distal flows. *Amer. J. Math.*, **85** (1963), 477-515.
- [F77] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.*, **31** (1977), 204-256.
- [FW96] 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*, Eds.: Bergelson, March, Rosenblatt, Walter de Gruyter & Co, Berlin, New York (1996), 193-227
- [HK01] B. Host and B. Kra. Convergence of Conze-Lesigne averages. *Erg. Th. & Dyn. Sys.*, **21** (2001), 493-509.
- [HK02] B. Host and B. Kra. An Odd Furstenberg-Szemerédi Theorem and quasi-affine systems *J. Anal. Math.* **86** (2002), 183-220.
- [HK02b] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. Preprint available at <http://www.math.psu.edu/kra/publications.html>, 2002.
- [K34] A. Y. Khintchine. Eine Verschärfung des Poincaréschen "Wiederkehrrsatzes" *Comp. Math.*, **1** (1934), 177-179.

- [Le02] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of rotations of a nilmanifold. Preprint available at <http://www.math.ohio-state.edu/leibman/preprints> (2002).
- [L84] E. Lesigne. Résolution d'une équation fonctionnelle. *Bull. Soc. Math. France*, **112** (1984), 177–19.
- [L89] E. Lesigne. Théorèmes ergodiques pour une translation sur une nilvariété. *Erg. Th. & Dyn. Sys.*, **9-1** (1989), 115–126.
- [L91] E. Lesigne. Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques. *Erg. Th. & Dyn. Sys.*, **11-2** (1991), 379–391.
- [Ma] A. Malcev. On a class of homogeneous spaces. *Amer. Math. Soc. Transl.* **39** (1951)
- [MZ55] D. Montgomery and L. Zippin. *Topological Transformation Groups*. Interscience Publishers, New York – London (1955).
- [P69] W. Parry. Ergodic properties of affine transformations and flows on nilmanifolds. *Amer. J. Math.*, **91**, 3 (1969), 757–771.
- [P70] W. Parry. Dynamical systems on nilmanifolds. *Bull. London Math. Soc.* **2** (1970), 37–40.
- [R95] D.J. Rudolph. Eigenfunctions of $T \times S$ and the Conze-Lesigne algebra. *Ergodic Theory and its Connections with Harmonic Analysis*, Eds.: Petersen, Salama, Cambridge University Press, New York (1995), 369–432.

EQUIPE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES, UNIVERSITÉ DE MARNE LA VALLÉE,
77454 MARNE LA VALLÉE CEDEX, FRANCE
E-mail address: `host@math.univ-mlv.fr`

DEPARTMENT OF MATHEMATICS, MCALLISTER BUILDING, THE PENNSYLVANIA STATE UNIVERSITY,
UNIVERSITY PARK, PA 16802
E-mail address: `kra@math.psu.edu`